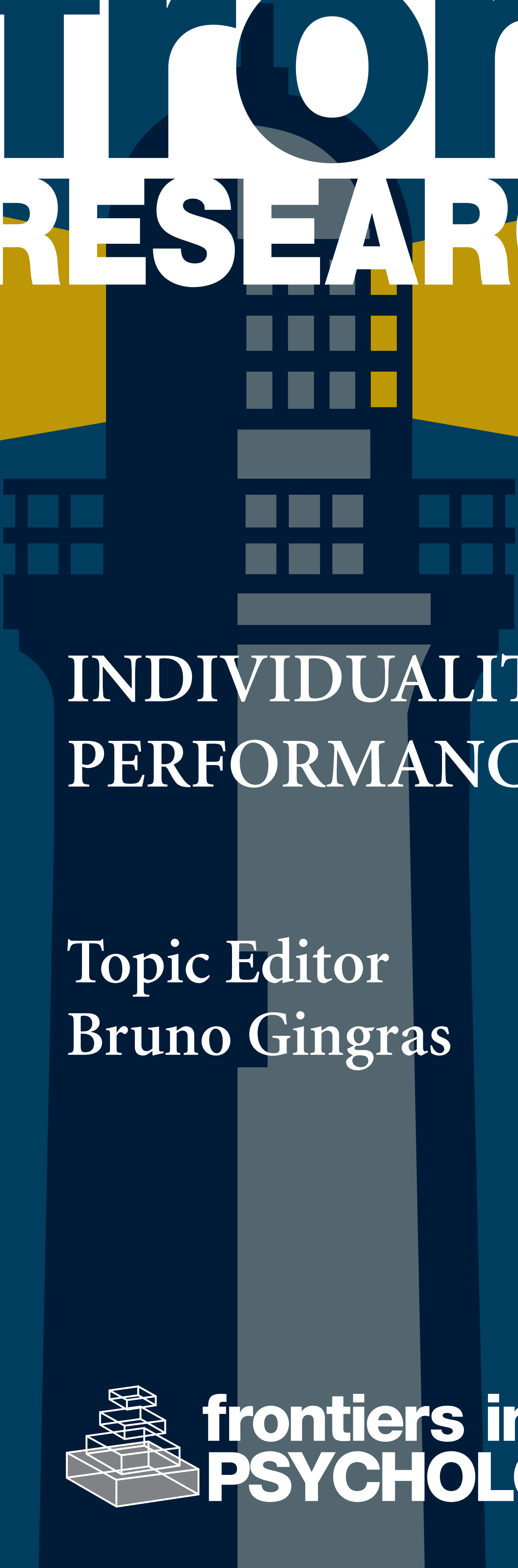
frontiers
RESEARCH TOPICS



INDIVIDUALITY IN MUSIC PERFORMANCE

Topic Editor
Bruno Gingras

frontiers in
PSYCHOLOGY

frontiers

FRONTIERS COPYRIGHT STATEMENT

Cover image provided by Ibbl sarl, Lausanne CH

ISSN 1664-8714
ISBN 978-2-88919-307-3
DOI 10.3389/978-2-88919-307-3

ABOUT FRONTIERS

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

FRONTIERS JOURNAL SERIES

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing.
All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

DEDICATION TO QUALITY

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.
Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view.
By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

WHAT ARE FRONTIERS RESEARCH TOPICS?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area!
Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: researchtopics@frontiersin.org

INDIVIDUALITY IN MUSIC PERFORMANCE

Topic Editor:
Bruno Gingras, University of Vienna, Austria

Humans are remarkably adept at identifying individuals on the basis of their facial features, or other traits such as gait or vocal timbre. Besides voice, another auditory medium capable of carrying identity information is music. Indeed, certain famous musicians, such as John Coltrane or Sonny Rollins, need only to play a few notes to be unequivocally recognized. Along with emotion and structural cues, artistic individuality seems to be a key element communicated in music performance. Yet, the means by which individuality is expressed in performance, as well as the cognitive processes employed by listeners to perceive identity cues, remain poorly elucidated. Other pertinent issues, including the connection between a performer's technical competence and ability to convey a specific musical identity, as well as potential links between individuality and career-defining outcomes such as critical recognition and aesthetic appraisal, warrant further exploration.

Quantitative approaches to the study of music performance have benefited greatly from MIDI technology and the application of computational methods, leading to the flourishing of empirical music performance research over the last few decades. More recently, neuroimaging techniques have provided valuable insights into the neural mechanisms involved in the cognitive processes of performing music. Nevertheless, this field continues to benefit greatly from qualitative approaches, given that the communication of affect and identity cues in music performance leads to a rich subjectivity of impressions that must be accounted for in order to lead to a greater understanding of this multifaceted phenomenon.

The aim of this Research Topic is to provide a forum for interdisciplinary research broadly related to the expression and perception of individuality in music performance. Research methodology includes behavioral, psychophysiological, and neuroimaging techniques. Both quantitative and qualitative approaches are presented The scope of this Research Topic includes laboratory studies as well as studies in real-life performance settings and longitudinal studies on performers.

Table of Contents

frontiers in
PSYCHOLOGY

EDITORIAL
published: 25 June 2014
doi: 10.3389/fpsyg.2014.00661

Individuality in music performance: introduction to the research topic

Bruno Gingras**

Department of Cognitive Biology, Faculty of Life Sciences, University of Vienna, Vienna, Austria
**Correspondence: brunogingras@gmail.com*

Edited and reviewed by:
Eddy J. Davelaar, Birkbeck College, UK

Keywords: music performance, individuality, individual differences, piano performance, vocal emotions

The ability to discriminate among individuals is crucial in species, such as humans, that place a premium on kin recognition (Tang-Martinez, 2001). Identity cues used by humans comprise not only visual cues, including relatively static cues such as facial features (Carey, 1992) or dynamic displays such as gait and walking (Blake and Shiffrar, 2007), but also auditory cues such as voices (Belin et al., 2004), clapping patterns (Repp, 1987), or even tones which follow temporal patterns similar to clapping (Flach et al., 2004).

Cues to individuality can also be communicated efficiently through music. Indeed, along with emotion and structural cues, artistic individuality seems to be a key element conveyed in music performance. Over the last few decades, a growing body of research has examined issues related to individuality in musical performance (e.g., Repp, 1992; see Sloboda, 2000 for a review). Yet, the means by which individuality is musically expressed and perceived have remained poorly elucidated until recently. Hence, the aim of this Research Topic is to provide a forum for interdisciplinary research broadly centered on individuality and individual differences in music performance. This goal was successfully achieved, and the 14 contributed articles illustrate the depth and breadth of the topic, with themes ranging from personality correlates of flow proneness among pianists to unique “fingerprints” in the singing voice.

Setting the tone for the Research Topic, Wöllner (2013) emphasized in an opinion piece the importance of using averaged features, representing the mean of a large sample of performances by different performers, rather than computer-generated “deadpan” reproductions as the baseline for quantifying individuality in music performance. On a related issue, Farbood and Upham (2013) compared listener judgments of musical tension obtained for a recording of a Schubert song and its computer-generated harmonic reduction, showing that differences in perceived tension changes between the two excerpts highlighted interpretive choices in performance.

Historically, a substantial body of music performance research has focused on piano performance (see Gabrielsson, 2003 for a review), and this trend was maintained here. Van Vugt et al. (2013) explored the individuality associated with small but systematic temporal deviations in musical scales played by pianists, showing that although human listeners were not able to distinguish these “temporal fingerprints” by ear, high accuracy rates were obtained by classifiers. Bernays and Traube (2014) investigated individuality in pianists’ performance of timbral nuances, and their analysis revealed that pianists exhibited unique profiles associated with different sonorities, while at the same time displaying common patterns of dynamics and articulation for each timbral color. Marin and Bhattacharya (2013) identified emotional intelligence and amount of daily practice as predictors of individual differences in proneness for flow among pianists, but did not observe a correlation between flow and high achievement in piano performance. Their study was the object of a commentary by Srinivasan and Gingras (2014) exploring the putative role of control and attention in flow states in music performance.

Two articles focused on the harpsichord, another keyboard instrument that, unlike the piano, has been relatively neglected so far in music performance research. Gingras et al. (2013) invited harpsichordists to record three different pieces and identified global markers of individuality, such as performers consistently using a more detached articulation across all three pieces, as well as associations between the note-by-note expressive profiles of different performers that subsisted across pieces or expressive parameters. In a follow-up to an earlier study on organ performance (Gingras et al., 2011), Koren and Gingras (2014) investigated whether listeners could reliably identify harpsichordists playing short excerpts from two different pieces. They found that musicians were more accurate than non-musicians, and only musicians performed above chance when matching the two different pieces to the same performer.

Voice production and perception was a major area of interest, with five contributions. Hutchins and Moreno (2013) proposed a new model to account for the variability between vocal perception and performance abilities in the general population. Their Linked Dual Representation model, which posits that vocal information can be encoded either as a symbolic or as a motoric representation, leads to a series of intriguing predictions about speech imitation, singing, and response timing. In a similar vein, Yang et al. (2013) investigated the coexistence of perceptual pitch deficits with pitch production deficits in music and in Mandarin speech in both amusics and tone agnosics, and their results suggest that the perception-production relationship for pitch among individuals may be domain-dependent. Trehub et al. (2013) confirmed the presence of individual cross-modal signatures in maternal speech and singing which can be discerned by both adults and infants, enabling listeners to successfully link recordings of unfamiliar speaking or singing voices to silent videos of the talkers or singers. Two articles focused more specifically on emotional singing: Quinto et al. (2014) examined the use of facial movements to communicate emotion, confirming the central role of facial expressions in vocal emotional communication while at the same time highlighting individual differences between singers, while Livingstone et al. (2014) analyzed the influence of vocal training and acting experience on the perception of vocal quality and emotional genuineness. They reported that acting experience was associated both with a decrease in voice quality and with an increase in perceived genuineness.

Finally, two studies addressed applied research topics related to individual differences in music performance. Williamon et al. (2014) designed and tested two simulated performance environments to help performers cope with issues related to performance anxiety, and discussed potential implications for performance training. Fritz et al. (2013) showed that participants' mood during exercise machine workout is enhanced more strongly with individualized musical feedback modulated by the participants' movements than with passive music listening.

In summary, this Research Topic both confirms and extends earlier findings, while at the same time opening up new avenues of research, especially in keyboard and voice performance. More generally, it highlights the cross-fertilizing potential of applying a multidisciplinary approach to the study of individuality in music performance, emphasizing the importance of fostering collaborations among musicologists, computer scientists, psychologists, neuroscientists, and the performers themselves.

REFERENCES

Belin, P., Fecteau, S., and Bédard, C. (2004). Thinking the voice: neural correlates of voice perception. *Trends Cogn. Sci.* 8, 129–135. doi: 10.1016/j.tics.2004.01.008

Bernays, M., and Traube, C. (2014). Investigating pianists' individuality in the performance of five timbral nuances through patterns of articulation, touch, dynamics, and pedaling. *Front. Psychol.* 5:157. doi: 10.3389/fpsyg.2014.00157

Blake, R., and Shiffrar, M. (2007). Perception of human motion. *Annu. Rev. Psychol.* 58, 47–73. doi: 10.1146/annurev.psych.57.102904.190152

Carey, S. (1992). Becoming a face expert. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 335, 95–103. doi: 10.1098/rstb.1992.0012

Farbood, M. M., and Upham, F. (2013). Interpreting expressive performance through listener judgments of musical tension. *Front. Psychol.* 4:998. doi: 10.3389/fpsyg.2013.00998

Flach, R., Knoblich, G., and Prinz, W. (2004). Recognising one's own clapping: the role of temporal cues. *Psychol. Res.* 69, 147–156. doi: 10.1007/s00426-003-0165-2

Fritz, T. H., Halfpaap, J., Grahl, S., Kirkland, A., and Villringer, A. (2013). Musical feedback during exercise machine workout enhances mood. *Front. Psychol.* 4:921. doi: 10.3389/fpsyg.2013.00921

Gabrielsson, A. (2003). Music performance research at the millenium. *Psychol. Music* 31, 221–272. doi: 10.1177/03057356030313002

Gingras, B., Asselin, P.-Y., and McAdams, S. (2013). Individuality in harpsichord performance: disentangling performer- and piece-specific influences on interpretive choices. *Front. Psychol.* 4:895. doi: 10.3389/fpsyg.2013.00895

Gingras, B., Lagrandeur-Ponce, T., Giordano, B. L., and McAdams, S. (2011). Perceiving musical individuality: performer identification is dependent on performer expertise and expressiveness, but not on listener expertise. *Perception* 40, 1206–1220. doi: 10.1068/p6891

Hutchins, S., and Moreno, S. (2013). The linked dual representation model of vocal perception and production. *Front. Psychol.* 4:825. doi: 10.3389/fpsyg.2013.00825

Koren, R., and Gingras, B. (2014). Perceiving individuality in harpsichord performance. *Front. Psychol.* 5:141. doi: 10.3389/fpsyg.2014.00141

Livingstone, S. R., Choi, D. H., and Russo, F. A. (2014). The influence of vocal training and acting experience on measures of voice quality and emotional genuineness. *Front. Psychol.* 5:156. doi: 10.3389/fpsyg.2014.00156

Marin, M. M., and Bhattacharya, J. (2013). Getting into the musical zone: trait emotional intelligence and amount of practice predict flow in pianists. *Front. Psychol.* 4:853. doi: 10.3389/fpsyg.2013.00853

Quinto, L. R., Thompson, W. F., Kroos, C., and Palmer, C. (2014). Singing emotionally: a study of pre-production, production, and post-production facial expressions. *Front. Psychol.* 5:262. doi: 10.3389/fpsyg.2014.00262

Repp, B. H. (1987). The sound of two hands clapping: an exploratory study. *J. Acoust. Soc. Am.* 81, 1100–1109. doi: 10.1121/1.394630

Repp, B. H. (1992). Diversity and commonality in music performance—an analysis of timing microstructure in Schumann's "Träumerei." *J. Acoust. Soc. Am.* 92, 2546–2568. doi: 10.1121/1.404425

Sloboda, J. A. (2000). Individual differences in music performance. *Trends Cogn. Sci.* 4, 397–403. doi: 10.1016/S1364-6613(00)01531-X

Srinivasan, N., and Gingras, B. (2014). Emotional intelligence predicts individual differences in proneness for flow among musicians: the role of control and distributed attention. *Front. Psychol.* 5:608. doi: 10.3389/fpsyg.2014.00608

Tang-Martinez, Z. (2001). The mechanisms of kin discrimination and the evolution of kin recognition in vertebrates: a critical re-evaluation. *Behav. Process.* 53, 21–40. doi: 10.1016/S0376-6357(00)00148-0

Trehub, S. E., Plantinga, J., Brcic, J., and Nowicki, M. (2013). Cross-modal signatures in maternal speech and singing. *Front. Psychol.* 4:811. doi: 10.3389/fpsyg.2013.00811

Van Vugt, F. T., Jabusch, H.-C., and Altenmüller, E. (2013). Individuality that is unheard of: systematic temporal deviations in scale playing leave an inaudible pianistic fingerprint. *Front. Psychol.* 4:134. doi: 10.3389/fpsyg.2013.00134

Williamon, A., Aufegger, L., and Eiholzer, H. (2014). Simulating and stimulating performance: introducing distributed simulation to enhance musical learning and performance. *Front. Psychol.* 5:25. doi: 10.3389/fpsyg.2014.00025

Wöllner, C. (2013). How to quantify individuality in music performance? Studying artistic expression with averaging procedures. *Front. Psychol.* 4:361. doi: 10.3389/fpsyg.2013.00361

Yang, W.-X., Feng, J., Huang, W.-T., Zhang, C.-X., and Nan, Y. (2013). Perceptual pitch deficits coexist with pitch production difficulties in music but not mandarin speech. *Front. Psychol.* 4:1024. doi: 10.3389/fpsyg.2013.01024

Conflict of Interest Statement: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 27 May 2014; accepted: 09 June 2014; published online: 25 June 2014.

Citation: Gingras B (2014) Individuality in music performance: introduction to the research topic. Front. Psychol. ***5****:661. doi: 10.3389/fpsyg.2014.00661*

This article was submitted to Cognitive Science, a section of the journal Frontiers in Psychology.

frontiers in PSYCHOLOGY

OPINION ARTICLE
published: 19 June 2013
doi: 10.3389/fpsyg.2013.00361

How to quantify individuality in music performance? Studying artistic expression with averaging procedures

Clemens Wöllner *

Institute of Musicology and Music Education, University of Bremen, Bremen, Germany
*Correspondence: woellner@uni-bremen.de

Edited by:
Bruno Gingras, University of Vienna, Austria

In artistic fields such as western music performance of the past 200 years, individuality is highly valued as a performer's expression of his or her aesthetic concepts. Yet characterizations of individual performance qualities have largely remained on a descriptive level. In this opinion article, it is argued that if researchers aim at quantifying individuality, then the only feasible approach is to determine the baseline from which individual performances diverge. Rather than using a computer-generated "deadpan performance" with no expressive features, this baseline should refer to average features comprising a number of different human performances in a given cultural context.

AVERAGES AS PROTOTYPES

Research in cognitive psychology has shown that people prefer average features in visual and auditory modalities. Averaged human faces (Galton, 1878; Langlois and Roggman, 1990) and voice utterances (Bruckert et al., 2010) were rated as more attractive. Averaging procedures typically result in even, smooth visual displays or sounds that show no extremes in any feature. Psychological theories suggest that people construct mental prototypes based on a large number of individual objects or people they encounter in their lives—an idea already expressed by Kant (1790/1995) in *Critique of Judgment*. Averaged individual features can thus approximate mental prototypes within an epoch or culture. Displays with prototypical characteristics conform to people's expectations and enable more fluent processing, which in turn may cause a cognitive bias for averages (Rubenstein et al., 1999). Objects that are easily processed are thus often perceived to be more attractive. This bias resembles the enhanced ease of processing for repeatedly presented stimuli as demonstrated in the well-known mere exposure effect.

EXPRESSIVENESS INDICATES INDIVIDUALITY

These advantages for averaged features stand in remarkable contrast to notions of artistic and musical individuality. Listeners typically do expect more from a concert or a recording than a smooth and even performance. In popular music genres, the sound of a singer's voice, of the instruments and the mix of audio tracks are often aimed at conveying a distinct, individual character (cf. Frith, 1998). In classical genres, emphasis is laid on subtle timing perturbations and fluctuations in dynamic intensity. Sudden delays or changes in intensity that do not conform to prototypical expectations may cause surprise and other emotional reactions (Huron, 2006) and reveal the individual performer's musical intentions. Research along these lines has for a long time studied expressive timing deviations from a non-expressive metronomic version. These timing deviations constitute an individual expressive microstructure (for an overview, see Clarke, 1995). Although it can be revealing to analyze the lengthening of note values in a final ritard for a number of different performers or for historical recordings during the course of the 20th century, no statements can be drawn about the *degree of individuality* in these performances. In other words, an expressive microstructure of a performance does not reveal *per se* whether the performance will be perceived as being individual. In earlier decades of the 20th century, for instance, musicians typically employed large *rubati* (Timmers, 2007) in accordance with listeners' expectations of that time, while nowadays these variations would not conform to prototypical listening expectations. Furthermore, timing deviations are to some degree also caused by human physiological constraints (Loehr and Palmer, 2009); performers are thus not able to render a perfect mechanical, metronomically exact performance. For these reasons, deviations from so-called deadpan renditions are no valid indicator of individuality.

PERCEPTIONS OF INDIVIDUALITY MAY DIVERGE FROM QUALITY JUDGMENTS AND BEHAVIORAL ADVANTAGES

In contrast to many other forms of art, musical performances can be averaged according to the main quantifiable dimensions of duration, dynamic intensity, and pitch. MIDI technology allows for both relatively simple analysis of these parameters from a given set of individual performances as well as synthesis, which results in an averaged performance approximating mental prototypes. In a seminal study by Repp (1997), experienced listeners ranked the quality of artificial piano performances with averaged timing patterns higher as compared to actual performances with individual timing. This outcome was obtained both for student and professional pianists, some of their performances showing large deviations from the average timing pattern. At the same time, individuality of averaged performances was ranked lower. These results suggest that averaged musical performances are preferred in one dimension (quality) as prototypes, while on the other hand, they may be perceived as somewhat "dull" in comparison with some highly individual performances.

In a recent study, we asked whether quantitatively averaged point-light displays of orchestral conductors are perceived as prototypes and lead to advantages in behavioral and evaluative experimental tasks (Wöllner et al., 2012). While conductors shape musical performances according to their individual expressive intentions, they also need to organize the balance of the orchestral sound and synchronize the timing by means of gestures. In order to be recognizable to a large number of different musicians, there are thus constraints and limits to individuality. In our study, twelve orchestral conductors were recorded with a 3D motion capture system while they conducted typical four-beat measures with metronome-controlled timing. Based on the horizontal and vertical dimensions, averages were created and presented as point-light displays to participants in an experiment. Their task was to tap to the beat and to evaluate the conductors in terms of quality, clarity of beat, conventionality, and expressiveness. Our analyses revealed advantages for prototypes in action responses, which adds to previous research using perceptual judgments of attractiveness or quality. Participants' synchronization with averaged conducting displays was more consistent (reduced tapping variability) and more synchronous (smaller asynchronies) compared with displays of individual conductors. Kinematic analyses revealed reduced normalized jerk in averaged conducting, indicating smoother movements than for individual displays. Averages were also judged to be more conventional, which demonstrates that participants indeed perceived them as prototypes. Beat clarity of conducting gestures and quality, in contrast to Repp's (1997) findings, were not significantly higher for averaged compared with individual movements. Yet individual conductors were perceived to be more expressive. As a consequence, the predictability and smoothness of prototypical movements enhanced action responses, given that they were easier to perceive and process, while individual expressiveness was reduced. For fields with transitive gestures such as orchestral conducting, then, experienced individuals need to balance the functionality of their profession with the demand of conveying their distinct expressive intentions.

QUANTIFYING INDIVIDUALITY

Apart from the above mentioned investigations of expressive timing deviations using MIDI technology, studies of musical individuality may focus more on musical timbre, pitch and intensity. Methods used by Bruckert et al. (2010) for acoustical morphing of the human voice could be employed to investigate individual musical timbres. Musicians are able to shape the timbre of certain instruments to some extent. Similarly, for singers as well as for instruments without fixed pitch such as many wind and string instruments, averaged deviations from notated pitch in equidistant temperament could be analyzed. The sharpening or flattening of tones may reveal certain expressive intentions of individual performers. Intensity can be measured with MIDI technology or by acoustical analyses. Studies have shown some dependencies between timing and intensity fluctuations (cf. Parncutt, 2003), and it would furthermore be revealing to analyze relationships with timbre and pitch in a systematic way.

When measuring the four musical dimensions of timing, intensity, timbre, and pitch to capture an individual musician's "fingerprint" of his or her performance, researchers may consider three relatively novel approaches in the field. First, as argued above, performances can be averaged according to these dimensions. In comparison to the analysis-by-synthesis approach that has been employed primarily for the study of the human voice (Sundberg, 2006), averaged performance dimensions are not artificially generated to produce a naturally sounding voice or instrument. Rather, distances to a synthesized performance based on actual renditions are used to estimate individuality. Averaging and synthesizing music may, as a caveat, result in the loss of some musical detail present in individual performances, and the averaged timbre may even sound unnatural, since the averaging procedure smoothes out extremes. It is also worth considering whether one dimension in question should be investigated while keeping the others constant. Repp (1997) only averaged timing patterns and compared them to individual performances while using the same timbre, intensity and pitches for all examples he presented to participants. Researchers can thus measure continuous deviations from an averaged performance in any dimension or combination of dimensions. Second, rather than only employing perceptual judgments, action-specific effects should be investigated. Synchronization studies offer a particularly valid research paradigm in the field of music. People may behave differently to stimuli that conform to their mental prototypes, even if they are not aware of these effects. Therefore, it is intriguing to combine perceptual and behavioral tasks for analyzing the dimensions that distinguish individual performances from others. Third, musical performers may take part in research studies both as musicians and listeners/observers of their own performances to investigate sense of agency for individual musical characteristics. In a study using motion capture of a Mendelssohn string symphony (Wöllner, 2012), orchestral conductors were able to identify point-light displays of their own conducting movements, while the corresponding short musical excerpts or point-light displays of gait did not contain sufficient cues for distinguishing their individual performances from those of other conductors.

CONCLUSION

Individual artistic expression should be considered in the boundaries of a given cultural context. It can be defined as deviation from a prototypical exemplar within this context. Research suggests that averages composed of a number of individual objects or performances approximate people's mental prototypes. These prototypes have no universal validity, since cultural norms vary and change considerably across time even for fairly specific questions such as what is considered to be an appropriate performance of a musical piece. The curious research finding that prototypes are only preferred in some dimensions such as quality—while they are rated lower in important dimensions such as expressiveness—should be given more attention. It may well be that even for performing arts there are limits to individuality as soon as overall quality and mastery of a technical skill come into question. Finally, the development of individual performance manners should be addressed. A great deal of learning occurs implicitly by imitating influential others or by trying to reach the standard of a given prototype, and individual intentions need to be well balanced with cultural norms.

REFERENCES

Bruckert, L., Bestelmeyer, P., Latinus, M., Rouger, J., Charest, I., Rousselet, G. A., et al. (2010). Vocal attractiveness increases by averaging. *Curr. Biol.* 26, 116–120. doi: 10.1016/j.cub.2009.11.034

Clarke, E. F. (1995). "Expression in performance: generativity, perception and semiosis," in *The Practice of Performance,* ed J. Rink (Cambridge: Cambridge University Press), 21–54

Frith, S. (1998). *Performing Rites: on the Value of Popular Music.* Boston, MA: Harvard Universisty Press.

Galton, F. (1878). Composite portraits. *Nature* 18, 97–100. doi: 10.1038/018097a0

Huron, D. (2006). *Sweet Anticipation: Music and the Psychology of Expectation.* Cambridge, MA: MIT Press.

Kant, I. (1790/1995). *Kritik der Urteilskraft (Critique of Judgment).* Cologne: Könemann.

Langlois, J. H., and Roggman, L. A. (1990). Attractive faces are only average. *Psychol. Sci.* 1, 115–121. doi: 10.1111/j.1467-9280.1990.tb00079.x

Loehr, J. D., and Palmer, C. (2009). Sequential and biomechanical factors constrain timing and motion in tapping. *J. Mot. Behav.* 41, 128–136. doi: 10.3200/JMBR.41.2.128-136

Parncutt, R. (2003). "Accents and expression in piano performance," in *Perspektiven und Methoden einer Systemischen Musikwissenschaft,* ed K. W. Niemöller (Frankfurt: Peter Lang), 163–185

Repp, B. (1997). The aesthetic quality of a quantitatively average music performance: two preliminary experiments. *Music Percept.* 14, 419–444. doi: 10.2307/40285732

Rubenstein, A. J., Kalakanis, L., and Langlois, J. H. (1999). Infant preferences for attractive faces: a cognitive explanation. *Dev. Psychol.* 35, 848–855. doi: 10.1037/0012-1649.35.3.848

Sundberg, J. (2006). The KTH synthesis of singing. *Adv. Cogn. Psychol.* 2, 131–143. doi: 10.2478/v10053-008-0051-y

Timmers, R. (2007). Vocal expression in recorded performances of Schubert songs. *Musicae Scientiae* 11, 237–268.

Wöllner, C. (2012). Self-recognition of highly skilled actions: a study of orchestral conductors. *Conscious. Cogn.* 21, 1311–1321. doi: 10.1016/j.concog.2012.06.006

Wöllner, C., Deconinck, F. J. A., Parkinson, J., Hove, M. J., and Keller, P. E. (2012). The perception of prototypical motion: synchronization is enhanced with quantitatively morphed gestures of musical conductors. *J. Exp. Psychol.* 38, 1390–1403. doi: 10.1037/a0028130

Received: 28 May 2013; accepted: 02 June 2013; published online: 19 June 2013.

Citation: Wöllner C (2013) How to quantify individuality in music performance? Studying artistic expression with averaging procedures. Front. Psychol. ***4****:361. doi:* ***10.3389/fpsyg.2013.00361***

This article was submitted to Frontiers in Cognitive Science, a specialty of Frontiers in Psychology.

ORIGINAL RESEARCH ARTICLE
published: 30 December 2013
doi: 10.3389/fpsyg.2013.00998

Interpreting expressive performance through listener judgments of musical tension

Morwaread M. Farbood* and Finn Upham

Department of Music and Performing Arts Professions, New York University, New York, NY, USA

Edited by:
Bruno Gingras, University of Vienna, Austria

Reviewed by:
Freya Bailes, University of Western Sydney, Australia
Hauke Egermann, Technische Universität Berlin, Germany

****Correspondence:***
Morwaread M. Farbood, Department of Music and Performing Arts Professions, New York University, 35 W. 4th St. Rm. 1077, New York, NY 10012, USA
e-mail: mfarbood@nyu.edu

This study examines listener judgments of musical tension for a recording of a Schubert song and its harmonic reduction. Continuous tension ratings collected in an experiment and quantitative descriptions of the piece's musical features, include dynamics, pitch height, harmony, onset frequency, and tempo, were analyzed from two different angles. In the first part of the analysis, the different processing timescales for disparate features contributing to tension were explored through the optimization of a predictive tension model. The results revealed the optimal time windows for harmony were considerably longer (∼22 s) than for any other feature (∼1–4 s). In the second part of the analysis, tension ratings for the individual verses of the song and its harmonic reduction were examined and compared. The results showed that although the average tension ratings between verses were very similar, differences in how and when participants reported tension changes highlighted performance decisions made in the interpretation of the score, ambiguity in tension implications of the music, and the potential importance of contrast between verses and phrases. Analysis of the tension ratings for the harmonic reduction also provided a new perspective for better understanding how complex musical features inform listener tension judgments.

Keywords: tension, continuous response analysis, activity analysis, expressive performance, trend salience, harmony

INTRODUCTION

The percept of musical tension provides a window into the disparate components that comprise an expressive performance. Performers' interpretations of a composed piece highlight structural features of the music, shaping the listener's perception of both apparent and unusual aspects of the score (Palmer, 1996). In addition to explicit tempo and dynamics markings, the intrinsic structural features of the score—the harmony, melody, and hierarchical grouping structures—feed into and are reinforced by the expressive interpretation of the performer. Musical tension is a function of both the structural features inherent in the score and the expressive components contributed by the performer.

Tension has long been central topic of interest in music theory (Lerdahl and Krumhansl, 2007), and since the 1980s, it has also been the focus of numerous empirical studies. These studies have examined various aspects of musical tension. However, most of them either focus on specific features or do not attempt to explain how disparate features interact and integrate from a global perspective. The current study offers an explanatory model for how listeners perceive global tension that builds and improves upon an earlier cognitive, computational model (Farbood, 2012) alongside a detailed analysis of tension-rating differences between multiple interpretations of the same musical material. The musical and auditory features that are incorporated into this model include loudness (dynamics), tempo, onset frequency of note events, harmonic tension, and melodic pitch height.

One of the most frequently discussed features with respect to tension is loudness. There have been a number of studies that have identified loudness as a significant contributing factor to tension (Nielsen, 1983, 1987; Krumhansl, 1996; Ilie and Thompson, 2006; Granot and Eitan, 2011). Another commonly discussed feature is tempo, particularly with respect to the effect of rhythm and timing on tension (Krumhansl, 1996; Ilie and Thompson, 2006). Despite the observed contributions of tempo, the effect of rubato on tension perception (as defined by highly local changes in tempo) is unclear (Fredrickson and Johnson, 1996). These local changes might be better quantified in terms of onset frequency of note events, which has also been observed as a textural feature contributing to tension (Farbood, 2012).

A significant number of studies have examined harmonic tension (Nielsen, 1983, 1987; Bigand et al., 1996; Krumhansl, 1996; Bigand and Parncutt, 1999; Toiviainen and Krumhansl, 2003; Lerdahl and Krumhansl, 2007). Most of these studies have explored the psychological validity of Lerdahl's tonal tension model (1996; 2001), and the results have generally indicated that the model accurately predicts harmonic tension. Furthermore, the hierarchical component of the model—a reflection of the tonal context of a given chord—is an essential element for quantifying harmonic tension. Lerdahl's model also has a melodic "attraction" component; however, this aspect of the model has not been supported by empirical evidence. It appears that melodic contour, in terms of pitch height, contributes to global tension as a factor distinct from harmonic or tonal context (Nielsen, 1983, 1987; Bigand et al., 1996; Krumhansl, 1996; Granot and Eitan, 2011; Farbood, 2012).

Tension has also been linked to expectation by music theorists. Margulis's model of melodic expectation (2005), which combines elements of Narmour's (1990, 1992) implication–realization model of melodic expectation and Lerdahl's (2001) tonal pitch space and melodic attraction models, outlines three possible tension responses that listeners may experience: surprise-tension, which correlates inversely with expectancy ratings (i.e., something that is predictable generates little tension), denial-tension, the result of the difference between what is most expected and what actually occurs, and expectancy tension, which is related to the strength of the expectancy that has been generated about future events. Huron (2006) proposes a more general model of expectation that has a tension component related to arousal, corresponding to the physiological response generated when a listener is preparing for an upcoming event.

In addition to Huron's work, there have been other researchers who have suggested links between tension and affective arousal (Krumhansl, 1997). In some empirical work, it appears researchers assume that overlaps between tension and arousal exist, or they use terms that seem equate the two concepts (Rozin et al., 2004; Eerola and Vuoskoski, 2010; Olsen et al., 2010; Lehne et al., 2013). However, this connection has not been explicitly addressed or explored anywhere. The term "tension" has been, in general, used rather broadly in an under-defined manner; for further discussion about this, as well as a more extensive review of the tension literature, see Farbood (2012).

Tension as a measure is particularly useful from an empirical perspective due to the fact that listeners evaluate it with consistency, as indicated in previous studies by high within-subject and between-subject agreement for discrete and continuous tension judgments (Bigand et al., 1996; Farbood, 2012; Upham and Farbood, 2013). Average continuous tension judgments also appear not to be influenced by the musical preferences of listeners (Lychner, 1998) or to change with familiarity to musical stimuli (Fredrickson, 1999).

The reliability of tension ratings and its function as an emergent phenomenon arising from the interaction and integration of multiple, disparate musical parameters make it an effective high-level abstraction with which to examine the psychology of expressive performance and listener interpretation. However, the exploration of continuous tension ratings, in particular the examination of differences in ratings between related musical works or excerpts, has been severely limited by the methodological challenges of time-series analysis. New methods, in particular Activity Analysis (Upham and McAdams, unpublished manuscript), make it possible to investigate these responses in greater temporal detail than has previously been statistically defensible, providing new insights into the time course of tension ratings and agreement between responses.

In this paper, we focus on two previously unexplored aspects of musical tension: (1) the individual timescales at which disparate parameters contributing to tension are processed, and (2) detailed differences in how participants rate similar excerpts. We examine tension from two very different angles—from the perspective of the average tension response and from the ratings identifying tension-related differences between the stimuli. On both accounts, novel approaches to analyzing continuous tension data are explored, providing new insight into how tension is perceived and processed. From a broader cognitive point of view, timescales for feature processing can be viewed from a "levels-of-processing" perspective (Craik and Lockhart, 1972; Craik, 2002), where depth of memory-encoding operations are related to retrieval times. Through our modeling approach, we test the hypothesis that musical features requiring higher levels of cognitive processing, such as harmony, contribute to tension on longer timescales than low-level auditory features such as loudness and onset frequency.

MATERIALS AND METHODS

PARTICIPANTS

A total of 29 subjects took part in the experiment, of which four were excluded from this analysis. The remaining 25 participants were primarily undergraduate and graduate students at New York University, mean age 25.16 years ($SD = 6.50$), 11 female, 14 male. Subjects had an average of 9.78 years of formal training on a primary musical instrument ($SD = 5.18$) and an average self-rank in instrumental skill level of 3.90 ($SD = 0.94$) on a scale of 1–5. Average number of semesters of college-level music theory training was 3.32 ($SD = 2.43$) and the mean overall self-ranked musical training level was 3.52 ($SD = 0.92$), where $0 =$ no training and $5 =$ professional-level training. The four outlier participants were excluded for not completing the tension-rating task. In the first case, the participant became bored and started hitting random keys, resulting in the data collection interface exiting prematurely; in the other cases, the subjects were unresponsive and did not indicate any tension changes during some or all presented stimuli.

STIMULI

The stimuli for the experiment consisted of six musical excerpts, two of which are the focus of this paper. Although the task was the same throughout the experimental session, responses to the other four stimuli were collected for a different purpose (as a follow-up to an fMRI study). These other stimuli were an original 4′15″ excerpt from a Brahms piano concerto and three scrambled versions of it. The two stimuli used for the current study are a recording of "Morgengruss" from Schubert's *Die schöne Müllerin* performed by Peter Pears, tenor, and Benjamin Britten, piano, and a harmonic reduction of the piece by Fred Lerdahl (**Figure 1**). The recording, which includes a piano introduction and four repeated verses, has a duration of 3′55″. The harmonic reduction consists of the chord progression for a single verse and is 40 s long; it was rendered in QuickTime MIDI grand piano timbre for the experiment.

PROCEDURE

Participants were seated in front of a computer and presented the stimuli over Sennheiser HD 650 headphones in a hemi-anechoic chamber. They were asked to indicate changes in musical tension while listening to the stimuli by moving a horizontal slider on a MATLAB GUI that used Psychtoolbox extensions (Brainard, 1997; Pelli, 1997; Kleiner et al., 2007) for audio playback. The slider position was sampled at 10 Hz. After one practice trial with a one-minute-long musical stimulus that was not part of the test set, each audio excerpt was presented twice in a

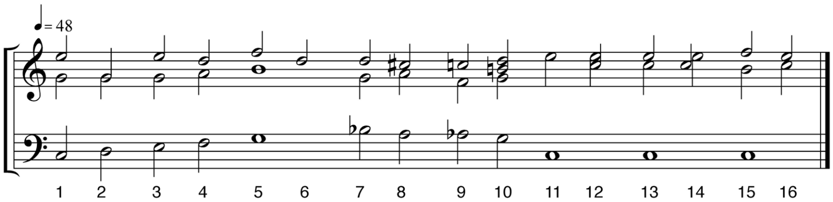


FIGURE 1 | Harmonic reduction of Schubert's Morgengruss by Lerdahl (2013). Chord numbers (not measure numbers) are indicated below the staves.

pseudo-randomized order where no stimulus was repeated before all stimuli had been presented at least once, and no stimulus was presented twice in a row (this avoided the situation where the last stimulus in the first set was the same as the first stimulus in the second set). Repeated presentations were utilized because it has been shown in previous work that within-subject consistency is very high for repeated continuous tension judgments of the same stimuli; habituation does not appear to be a problem in tension-rating tasks even after four repeated presentations of the same stimulus (Farbood, 2012).

DATA PREPROCESSING AND ANALYSIS METHODS

Activity analysis and rating change coordination

While individual raters may vary in whether or not they report changes in tension and how quickly they report these changes, the temporal dynamics of average tension-rating time series should be fairly representative of future listeners' judgments if the responses show a tendency to change at the same moments in time. In the terminology of activity analysis (Upham, 2011), for a given type of activity event (say an increase in tension ratings), the activity level of a given time frame is the proportion of responses showing indicating the event in question. Ratings-increase activity-level time series, which indicate the frequency of reported increases in tension over successive time frames, and ratings-decrease activity-level time series, which indicate reported decreases, describe whether responses actively agree on changes in tension over time. **Figure 6**, for example, shows tension-rating-change activity-levels for increases and decreases in each verse and the harmonic reduction of the Schubert excerpt; these time series report the proportion of responses that are active in the half-second following the onset of each 16th note, or 1/12th of the measure.

Like the average, calculation of these summary time series is not sufficient to claim that the resultant pattern is stimulus-driven and robust. Rather than validate the average directly, we test the coordination of tension-rating changes by looking at the distribution of activity levels measured for each stimulus compared to the null hypothesis of independent activity events. If responses do not show coordinated change with each other, we do not have sufficient reason to expect another set of responses to the same stimulus would yield the same temporal dynamics in the average of the activity-level time series. Coordination of rating-change activity can then be evaluated via the distribution of activity levels recorded in successive time frames spanning the piece. A simple test against the null hypothesis of independently timed tension-rating changes compares a model of the activity-level distribution for a collection of uncoordinated responses of equally frequent activity events to that of the experimental data. If the ratings are showing concurrent changes in tension, the experimental activity-level distribution will have more time frames of exceptionally low and relatively high activity levels (around 0.5 for rating data).

Built on this process of evaluating coordination of activity in a collection of time series is the Coordination Score. It is helpful to consider which collections of responses, e.g., responses to different stimuli, show more or less coordinated activity. The coordination score is calculated from the averaged negative log of the p-value from goodness-of-fit tests on multiple variants of the time series segmented into frames, with data-driven algorithmic controls on the independent model and distribution testing (Upham and McAdams, unpublished manuscript). The coordination test is evaluated using non-overlapping time frames of size greater than the sampling interval. To reduce the impact of arbitrarily cutting up the time series in even time frames and the risk of splitting apart changes related to the same event, the coordination score is calculated from all distinct slicings of the time series into adjacent non-overlapping time frames of fixed duration. The details of the calculation are accessible via the activity analysis MATLAB toolbox (Upham, 2013). The coordination score is a number from 0 to 16 which estimates the degree of coordination of a given activity event in a collection of stimulus-synchronous responses. Scores of less than 2 are equivalent to $p > 0.01$ for the measured activity-level distributions to occur by chance out of independently changing tension ratings (Upham and McAdams, unpublished manuscript). Very high coordination scores indicate that across the responses, activity is very well coordinated and strongly suggests repeatable summary time series, however, this does not mean that all ratings are in complete agreement on how and when tension changes.

The principle of the coordination score for single-activity events can also be applied to evaluate the independence of different forms of activity in the same collection of responses, for example, both increases and decreases in tension ratings (Upham and McAdams, unpublished manuscript). If a collection is coordinated in their activity, increases and decreases should alternate, but if the responses are independent, both types of activity are likely to happen in the same time frames. The coordination between two types of activity are similarly quantified and scored. If a collection of ratings is coordinated in each direction of tension change but fails to reject independence of increases and decreases, this undermines the robustness of the average time series. Its temporal profile, most often the object of analysis in relating continuous-response data to the time course of the stimulating music, is not likely to be robust since contributions of individual responses cancel out rather than strengthen the common progress of tension.

In the analyses below, activity-level time series are used with average time series to describe the four collections of tension ratings, while the coordination scores, rating change increases, decreases, and alternation between the two, provide arguments for which response collections are employed for modeling and comparisons between verses.

Feature quantification

As a first step in the analysis procedure, seven musical and expressive features were quantified for the Morgengruss performance: dynamics, tempo, onset frequency, harmonic tension, and pitch height of the melody, inner voice, and bass line. Although this is not an exhaustive list of features, it was deemed sufficient to account for most of the variation in the tension responses.

The note onsets of the vocal line and accompaniment were determined using marker references in Amadeus Pro (HairerSoft, V 1.5.4) in conjunction with the audio-editing program Audacity (V 2.0.0) for locating precise onset times. With some musical discretion, onsets were generally marked at vowel onsets in consonant lead syllables and piano onsets were also included whenever they did not coincide with the vocal line. Onset frequency was calculated directly from the onset times and was quantified as the difference between the maximum event duration in the performance (3.51 s) and the current note-event duration. Thus the shorter the length of the note event was, the higher the onset frequency. Tempo was determined by using beat onsets. In cases where there wasn't an onset available on a beat, the onset time was linearly interpolated from the positions of the last onset and the next available onset.

Pitch height of the melody, inner voice, and bass lines were initially encoded as MIDI values and aligned to onset times. The inner voice is only active in the third phrase where the accompaniment echoes the vocal line; however, it was deemed prominent enough to merit addition to the feature set. A harmonic tension graph was created by using the mean tension response to the harmonic reduction stimulus (second rating only; see section Coordination in Tension Ratings for explanation). Tension values sampled 2 s after the onset of each chord were used to quantify harmonic tension for the entire piece (see **Figure 2**). This sampling lag was employed to compensate for response delay to the event onsets and is in line with lag times proposed by Schubert (2004), who showed that continuous arousal ratings reflect response lags of 1–3 s to musical features other than loudness. Although the harmonic reduction was shortened, it covered all the chord progressions in the piece, including the piano introduction (which is, harmonically, a compressed version of the first phrase of each verse). The harmonic tension graph was created by mapping the sampled tension values to the appropriate onset times corresponding to the performance.

Loudness was evaluated directly from the audio file. The analysis was done using Glasberg and Moore's (2002) psychoacoustic loudness model for time-varying sound available as a function in the Loudness Toolbox for MATLAB (short-term loudness output for omnidirectional sound recording) (Genesis, 2009). The output of the model is quantified in terms of sones, a standard unit of perceived loudness; one sone is equivalent the loudness level of a 1 kHz tone at 40 dB SPL (Stevens, 1936).

The mean tension response and feature graphs for loudness, tempo, onset frequency, and harmony were then normalized to zero-mean and unit standard deviation (Z-score). For pitch height, the mean and standard deviation for all three melodic lines combined were used to normalize the graphs. The normalized mean tension response and features are shown in **Figure 3**.

Parametric tension model

The first analysis approach explored the timescales at which performers and listeners process individual musical features and structures. The vehicle for this analysis was a modified version of Farbood's (2012) trend salience model. The main components of the model consist of (1) an *attentional window* that represents a perceptual moving window in time and extracts a current tension trend; (2) a *memory window*, which immediately precedes the attentional window and represents an abstraction of a previous tension trend; (3) differing weights for each of the musical parameters.

The concept of tension trends is the theoretical core of the model, describing how individual musical and auditory features integrate and interact to produce a global feeling of changing

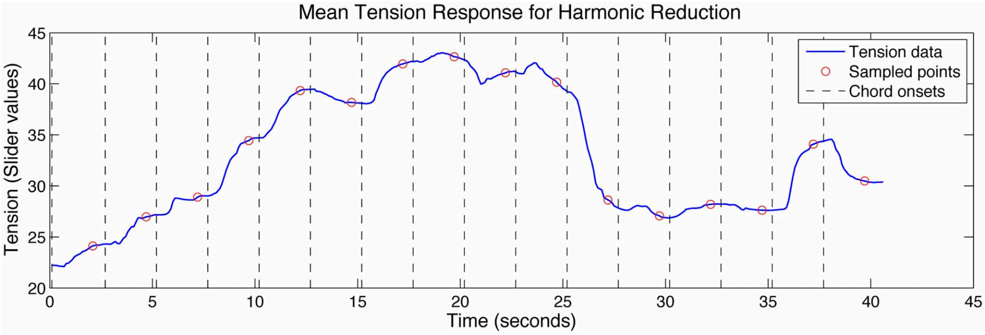


FIGURE 2 | Mean tension response to the harmonic reduction (second trial only) showing chord onsets and the corresponding sampled points used to describe harmonic tension for each chord change.

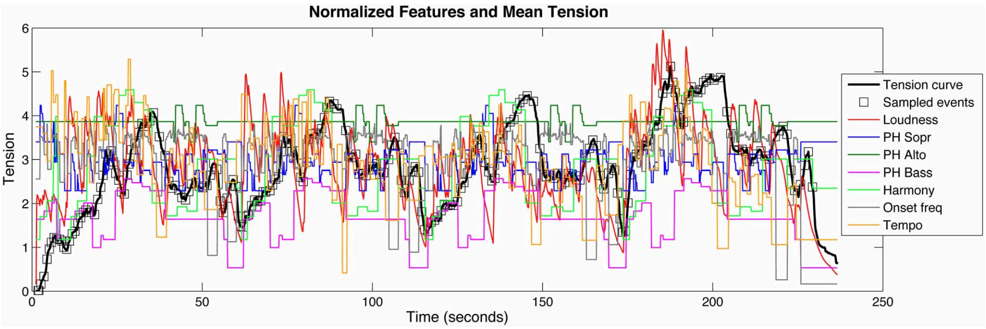


FIGURE 3 | Mean tension response (second trial only) with sampled points at each beat overlaid with all of the feature descriptions.

tension. The idea is that the *trend salience* of musical features is the key determiner of tension judgments. The model predicts tension as a function of individual musical features by taking into account the combined directional change of all of the features in the attentional window. This combined directional change, or tension trend, is weighted by what immediately precedes it in the memory window—if the direction of the trend in the memory window matches the current attentional window trend, the magnitude of the cumulative tension trend is additionally increased. These trends are integrated over time to generate a final tension prediction.

The tension trends are essentially the sum of the slopes of all the features for a particular time window. If all features have concurrent negative slopes, the sum of those slopes would indicate a clear decrease in tension for that time window. However, if the slope directions conflict, they might cancel each other out to some degree. The slope of a tension trend at time t is defined as

$$s'(t) = \beta \sum_f w_f s_f(t), \tag{1}$$

where $s_f(t)$ is the slope of best linear fit of feature f at time t; $\beta = 1$ if the sign of $s(t-d)$ does not equal the sign of $s(t)$; β is some positive value, empirically determined, if the sign of $s(t-d)$, where d represents the duration of the memory window, is the same as the sign of $s(t)$; w_f is the weight of feature f with $\sum_f w_f = 1$. See Farbood (2012) for a discussion on the ideal value of β. The optimal value obtained in that prior study ($\beta = 5$) was used in the current analysis. The β component describes the relationship between the memory and attentional windows and adds a nonlinearity to the model. The model becomes linear only when the memory window duration is set to 0 s.

Moment-to-moment integration of tension trends are simulated by overlapping moving windows; in this case, the increment is 250 ms, a step-size deemed sufficient for sub-beat time resolution. Each attentional window trend is merged with overlapping previous trends, resulting in recent windows weighted more strongly to simulate memory decay. This result in the slope of the tension curve at time t is defined as

$$S(t) = \sum_{\tau=0}^{\frac{d}{h}-1} s'(t-\tau h)k_\tau, \tag{2}$$

where h is the step size of the moving window increment, d is the attentional window duration, and k is a decay constant for a moving average filter with $\sum_i k_i = 1$. From this equation, a final tension value F_{ten} at time t can be derived:

$$F_{\text{ten}}(t) = h \sum_{i=0}^{T-1} S(i) \tag{3}$$

where $T = \lfloor \frac{t}{h} \rfloor$, and (in this simplified case) t is a multiple of h.

This model was modified for the current study in one significant respect: instead of employing fixed attentional and memory windows, different time windows were used for each feature in order to optimize the model predictions. Farbood's (2012) evaluation of the model indicated that the ideal values for d, in both the attentional and memory window cases, was 3 s. Here, trend salience model predictions were generated *separately* for each feature at different time windows before these individual predictions were integrated into a final tension prediction as described in Equations (2, 3). Furthermore, the feature weights were not used—differences in feature contributions were solely determined by the individual timescales.

Alignment

To translate between the recorded stimuli and metrical time, timings for every 1/36th of a measure were interpolated from the onset times in the performance, accommodating quadruple and triple subdivisions of quarter notes. Linear interpolation had to be used in absence of a better model for perceived tempo between onsets, since the resulting inaccuracies would be too small for the timescales employed in the analyses below. This mapping translated participants' tension ratings and musical features, initially sampled at 10 Hz, to metrical time with nearest-neighbor interpolation. Counting measures from the beginning of the verse, the harmonic reduction omitted mm. 7, 18, and 19, since they were repetitions of preceding chords. Matching these ratings to the verses, gaps were either interpolated or left blank, depending on the analysis.

Local coordination analysis

While global measures of coordination are useful for confirming shared behavior between ratings of tension to the same stimulus, they are insufficient for comparing ratings to related stimuli with any temporal sensitivity. For that, we employed a nonparametric method for generating distributions of activity levels for each time frame against which actual activity level could be assessed. First, each response time series was reduced to a point process, a sequence of 0 s and 1 s, where 1 indicates the activity in question, such as an increase in tension rating from one sample to the next. Shift values for each response were randomly selected from a time interval (in this case [−5, 5] s), yielding an alternate alignment of these point processes by shifting the complete series in time by the corresponding time value (with the ends looped for continuity), and the activity levels for each time frame were calculated as for the original alignment (Pipa et al., 2008). Repeating this shuffle 1000 times provides a distribution of potential activity levels for each frame for these same responses were they not aligned precisely to the stimulus. This process of shuffling preserves the complex characteristics of these responses, including serial behaviors and longer-term temporal structure (Pipa et al., 2008), while providing a reasonable comparison for assessing the potency of the particular coordination present in the stimulus-aligned response collection. The rank of the experimental activity levels against the alternatives essentially gives a p value for that time frame, thus providing a local estimate of likelihood for the experimental activity at each moment. Those moments with activity levels exceeding 95% of the alternatives have been marked as having significant coordination. By this method, it is possible to assess overlapping time frames (non-overlapping time frames are used in the collection coordination score). The basis of this approach is to assume that rating changes are not related to the stimulus—that they are noisy signals rather than clean—and that the points selected are those that have strong evidence to the contrary. This conservative perspective is necessary so long as we do not have a model for forecasting activity levels and the noise in rating changes with more precision. For now, moments of high activity coordination can be considered likely driven by the common stimulus, but moments of less notable activity levels should not be dismissed as noisy and unrelated to the music.

RESULTS

COORDINATION IN TENSION RATINGS

The coordination of rating-change activity in these responses is significant, allowing for more detailed temporal analysis of the summary reports. **Table 1** describes the rating-change activity for each collection of responses (stimulus and presentation order) in terms of average activity rates and rating-change coordination scores. Activity ratings of the second presentation were higher for both increases (Inc) and decreases (Dec). Coordination of rating changes was also higher for the second ratings of the simpler, more sparse harmonic reduction. The combination of higher activity rates and better coordination suggest that the subjects were more confident in their judgments in these later ratings. This is consistent with the hypothesis that subject reports of continuous response are cleaner and faster with increased familiarity with the task and the stimulus, and that agreement between participants for judgment tasks improve with a common context (the full stimulus set), at least for simple stimuli. Tension ratings for the recorded performance manifested very high coordination during the second trial as well, although not necessarily the highest in comparison to the harmonic reduction. With the density of information in the performed stimulus, we should expect that some disagreement in tension ratings will remain despite repetition of the task. In summary, these collections of continuous ratings of tension are strongly coordinated in their rating changes, and we can assume their shared temporal variation to be driven by the common temporal experience of the stimuli, justifying the use of average tension ratings in the modeling analysis described in section Model Optimization and Feature Timescales.

CORRELATION ANALYSIS

Correlation of features and mean tension

To get a general idea of how all of the feature descriptions and the mean tension response were related to each other, correlations were performed between all pairs. The time series were sampled at every beat instead of the original 10 Hz rate. Beat sampling was used in all subsequent correlation analyses as well. Spearman's ρ was calculated because the values of several features were not normally distributed. The mean tension response used in all of the following analyses included only the second ratings of the subjects; see section Coordination in Tension Ratings for a detailed explanation for this. All correlations between the mean tension response and features, as well as any other time series in subsequent analyses, incorporate a response lag of 2 s (Schubert, 2004).

The results of the feature correlations, shown in **Table 2**, generally indicate a weak to moderate positive correlation between features. This includes a weak to moderate correlation between

Table 1 | Tension-rating change activity and coordination for the first and second presentations of two stimuli, as measured in 1 s time frames.

Stimulus	Activity rate (Inc)	Coordination score (Inc)	Activity rate (Dec)	Coordination score (Dec)	Coordination score (Alt)
Harmony 1 (40 1 s fr.)	0.21	7.0	0.12	1.8	N/A
Harmony 2 (40 1 s fr.)	0.22	9.9	0.16	7.5	N/A
Performance 1 (235 1 s fr.)	0.17	15	0.15	11	8.9
Performance 2 (235 1 s fr.)	0.19	14	0.17	16	9.7

Tension-rating increases (Inc) are described by their activity rate, the average likelihood of any response reporting an increase in tension during any 1 s time frame, and their coordination score for each response collection. Decreases in ratings (Dec) are described in the same terms, and the last column reports the coordination score for the alternation between increases and decreases in ratings (Alt). Note that the shorter stimulus has no alternating coordination score because it is too brief in duration to evaluate.

Table 2 | Spearman ρ values for correlations between features and mean tension (second rating only).

	Loudness	Pitch height: melody	Pitch height: inner	Pitch height: bass	Harmony	Onset frequency	Tempo	Tension
Loudness	–	−0.23	−0.09	0.20	0.30	0.33	0.37	0.52
Pitch height: melody	−0.23	–	−0.15	−0.18	0.19	−0.30	−0.005	−0.17
Pitch height: alto	−0.09	−0.15	–	0.04	0.11	0.26	0.25	0.03
Pitch height: bass	0.20	−0.18	0.04	–	0.34	0.25	0.17	0.43
Harmony	0.30	0.19	0.11	0.34	–	0.06	0.25	0.58
Onset frequency	0.33	−0.30	0.26	0.25	0.06	–	0.33	0.21
Tempo	0.37	−0.005	0.25	0.17	0.25	0.33	–	−0.07
Tension	0.52	−0.17	0.03	0.43	0.58	0.21	−0.07	–

Data points are sampled every beat.

harmony and all features; between dynamics and all other features except pitch height of the melody and inner voice; and between the remaining features (pitch height of the bass, onset frequency, and tempo) and all features except pitch height of the melody. In short, pitch height of the melody was the most negatively correlated with other features. Loudness, pitch height of the bass, harmony, and onset frequency had weak to moderately strong positive correlations with the mean tension response; pitch height of the melody had a weak negative correlation; and all other features had no apparent correlation with mean tension.

Table 3 | Optimal memory and attentional window durations for each feature.

	Features				
Window type	Loudness (s)	Pitch height (s)	Harmony (s)	Onset frequency (s)	Tempo (s)
Memory	0	2	13	1	1.5
Attentional	2	2	8.5	1	1

Note on reporting of correlations

Correlations are the most common statistic for comparing stimulus features and continuous response data in the existing literature, however, the interpretation of the significance of these calculations has been identified by many researchers as problematic (Schubert, 2002; Upham, 2011, 2012; Alluri et al., 2011). Throughout this paper, we include the correlation values for comparison with numbers published in previous work, but exclude estimates of significance for these calculations since the popular estimation method is inapplicable and no obvious alternatives particularly appropriate for the time series under analysis. The Spearman ρ is still informative for the reader as a relative measure of fit.

MODEL OPTIMIZATION AND FEATURE TIMESCALES

As described above, a trend salience model that predicts tension given a set of continuous feature descriptions was used as a basis for exploring the timescales at which disparate features contribute to global tension. Instead of integrating all features across identical time windows, the original model was altered so that separate attentional and memory window durations were utilized for each feature instead of using fixed weights for each feature. The tension predictions for the individual features at different timescales became the input feature vectors for the global tension prediction. This final integration step did not use a memory window (equivalent to a memory window duration of 0 s) and was integrated in 1 s attentional window frames. The goal was to optimize the output of the model by adjusting the window durations for each feature and correlating the prediction results with the empirical data.

The model was trained on the first half of the data (consisting of the piano introduction and first two verses) and then tested on the final two verses. From a theoretical perspective, it was assumed that the timescales for the three pitch-height descriptions should be identical. That resulted in five memory and attentional window durations to optimize, totaling 10 variables, each with 42 possible durations (0–20 s in 500 ms increments). This high-dimensional space is too large for an exhaustive search, so a step-wise testing procedure was implemented to explore the state space. Starting with all variables at the minimum duration of 0 s, one feature at a time was incremented until the correlation between the model output and mean tension response reached a peak. After optimal values were found for all features, the model output was tested again by exploratory deviations from the fixed optimal values in order to provide further confirmation of the result.

Given this heuristic approach, it is possible that there exists a better solution than what was found—i.e., that the resulting values represent a local maximum. However, the contributions of each feature appear to be predictable around individual maxima for each variable, indicating that a better solution is unlikely. The optimal values found, listed in **Table 3**, produced a strong correlation with the mean tension response, ρ (Spearman) $= 0.86$. The results indicate there is a significant difference in the way harmony is processed compared to other features. The memory and attentional windows for harmony were 13 s and 8.5 s, respectively, far longer than any other feature. Dynamics appeared to be the most instantaneously processed feature, having an optimal memory window of 0 s and an attentional window of 2 s. Short windows of $\sim$1 s were optimal for both onset frequency and tempo, while pitch height had slightly longer windows of 2 s.

These values, optimized only for the training data, were then used to produce a tension prediction for the test data. This also resulted in a strong correlation with the mean tension response, $\rho = 0.79$, providing some evidence that the model was not overfitting the data. This comes with the caveat that it cannot be determined with certainty that overfitting did not occur since the training and test data are very similar both in terms of the tension responses and feature descriptions (the only substantial difference being the inclusion of the piano introduction in the training data). See **Figure 4** for a comparison between the optimized model predictions and the training and test data.

COMPARISON BETWEEN VERSES AND HARMONIC REDUCTION

The next part of the analysis entailed an in-depth comparison of tension ratings between different presentations of related musical material, namely the individual verses and the harmonic reduction. Given the differences between the stimulus excerpts, we first considered statistics summarizing the cohesion of the collections of tension ratings per excerpt in their second presentations, including rating-change activity coordination. Second, we explored the summary times series for these tension ratings, as aligned to the common musical time course, examining differences between mean tension ratings and moments of extreme coordination. These analyses were primarily descriptive and speculative, demonstrating new approaches to explore and compare continuous ratings of tension. The results provided insight into continuous ratings of tension, potential features to explore in modeling efforts, and interesting examples of musical moments with counterintuitive rating-change behavior.

Differences between stimuli

As stimuli, the harmonic reduction and the performed piece are substantially different in character. The harmonic reduction is limited in timbre, with a steady tempo, uniform loudness of notes, very sparse onsets of notes with little impression of meter. Its relationship to the performed verse is restricted to the harmonic sequence, the piano as a sound source, and the approximate duration. As such, it is remarkable how similar the tension ratings were to those of the performed piece, at least as captured by the average.

The verses differ in lyrics, expression, structural order, and other performance parameters. As the participants in the study were US residents, we presume that most if not all were unable to understand the German text. The impact of the lyrics on the differences in tension ratings between verses were therefore likely related to the non-verbal aspects of the musicians' performance. The analysis of tension-rating differences between verses were thus focused on features that were accessible to listeners regardless of language comprehension: articulation, tempo deviation, timbre, and other aspects of affective expression.

Summaries of tension ratings per excerpt

Ignoring the temporal sequence for the moment, **Table 4** shows summary descriptions of the ratings for each verse and the harmonic reduction. In terms of absolute tension rating values, the average rating values for each stimulus and excerpt were not very different, ranging from 33 to 37.5 on a scale of 0–100, with the greatest difference being between the harmonic reduction and the fourth verse. Consistent with the idea that ratings of simple stimuli have less opportunity for disagreement, the coordination in rating-change activity was higher for the harmonic reduction, as was the inter-response agreement as captured by the standard deviation ratio. The standard-deviation ratio is a rough measure of agreement between ratings, complementary to the coordination score, and is calculated by dividing the standard deviation of the average rating time series by the average standard deviation of the ratings over time. Values closer to 0 reflect greater noise or contradictory behavior between

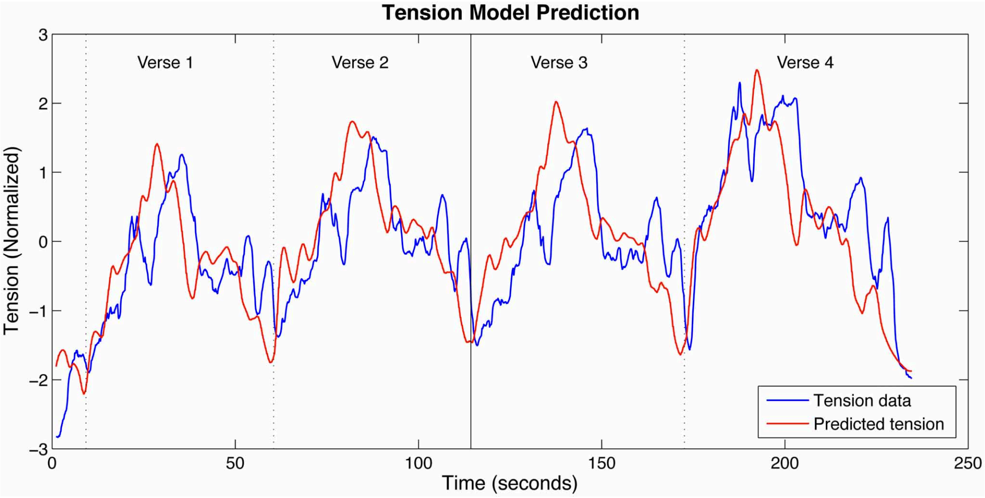


FIGURE 4 | Trend salience model predictions and the mean tension response. The solid vertical line indicates the boundary between Verses 2 and 3, the division between the training data (left side) and the test data (right side). The vertical dotted lines mark the other verse boundaries.

responses (Upham, 2012). This ratio for the harmonic reduction may, however, be inflated because these ratings include an orienting period, in which participants adjust the slider from the standard initial values programmed in the rating interface to their comfortable rating range during the first 10 s of a rating task (Schubert, 2012). This interval was not a factor for the verses because they were excerpts from the longer rating task consisting of the entire song, which includes an 8 s introduction. Between verses, the coordination scores varied, but they stayed above the significance threshold of 2. The tension ratings ranged more widely (as per the average standard deviation) for the last verse than all previous verses—nearly as much as the case for the harmonic reduction. It is also notable that the activity rates for increases in tension were consistently higher than that of decreases, reflecting the general shape of the average tension time series, with slow increases and quick falls visible in **Figure 5**.

Comparison of average tension time series between excerpts

Looking at the aligned average tension ratings in **Figure 5**, the similarity in contour for each verse was notable: rising through the first phrase (mm. 1–7) and cresting at m. 6, rising and staying high for the second phrase (mm. 8–11), falling in the transition to the third phrase (m. 12), and then remaining fairly stable until the small arcs that mark the closing cadence (mm. 16–17) and its repetition (mm. 18–19). At this level, ratings of tension for the harmonic reduction and the verses were quite similar: the tension implications of the harmony and voicing of this reduction were not strongly contradicted by the myriad of other musical features contributing to the perceived tension in the recorded performance.

Across verses, the loudness and tempo of the first phrase (mm. 1–7) were similar or greater than those of the second phrase, mm. 8–11. Despite this, the second phrase always had a higher tension range than the first. The harmonic reduction offers one

Table 4 | Summary statistics of tension-rating collections per verse and harmonic reduction.

Stimulus excerpt (fr.)	Activity rate (Inc)	Coordination score (Inc)	Activity rate (Dec)	Coordination score (Dec)	Tension mean	Tension STD	STD ratio
Harmony (40)	0.22	9.9	0.16	7.5	33.1	8.64	0.72
Verse 1 (52)	0.20	3.3	0.17	6.4	33.8	5.95	0.59
Verse 2 (54)	0.20	5.7	0.16	11	35.8	5.19	0.64
Verse 3 (58)	0.20	2.2	0.16	5.1	35.4	6.17	0.61
Verse 4 (64)	0.19	8.1	0.17	2.9	37.5	7.87	0.71

The collections of tension ratings for each excerpt are described in terms of the average rates of increases and decreases (1 s frames), rating change coordination, average mean tension per response, average standard deviation of tension responses, and the standard-deviation ratio.

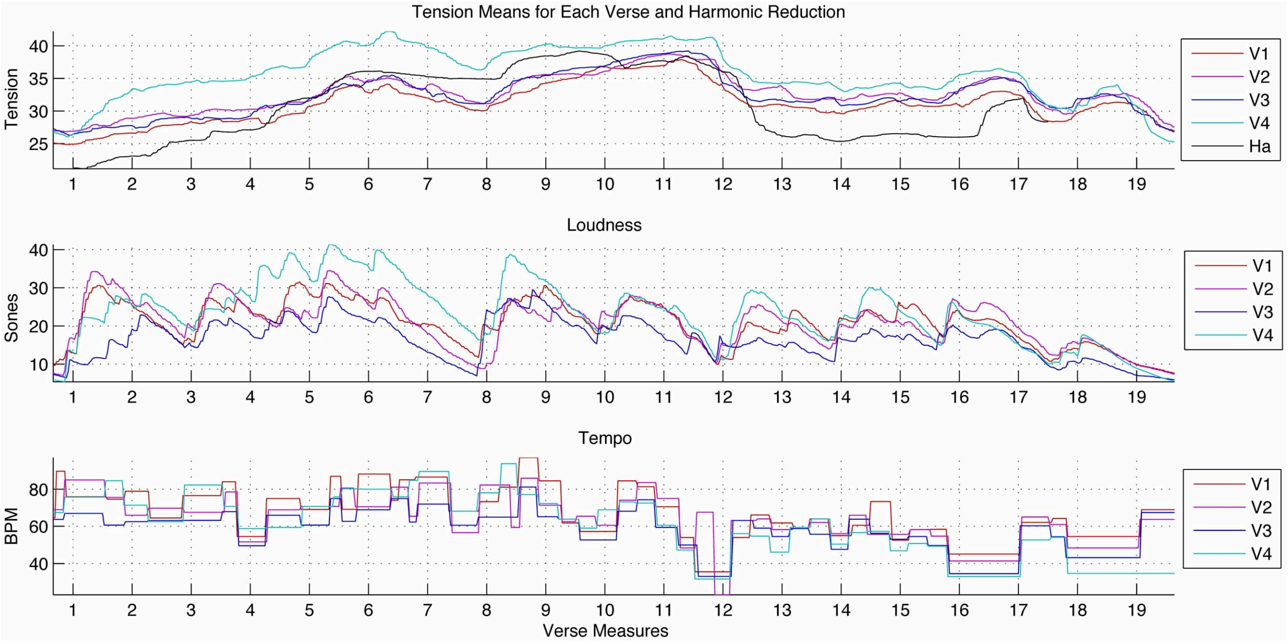


FIGURE 5 | The average musical-tension-rating time series for the harmonic reduction and each performed verse; the acoustic loudness of each verse in the performed rendition aligned to metrical time; and the onset-determined local tempo, also aligned to the verse measures, for the performed verses. These last two feature graphs do not include the harmonic reduction, since tempo and loudness were constant across all onsets.

possible explanation for this pattern, as the tonal character of the second phrase is farther removed from the tonic. Although loudness and tempo have often been noted to have significant influence over the perceived tension and intensity of music, this moment demonstrates that other features can contribute independently.

With regard to absolute tension-rating values, a few peculiarities of the ratings for the harmonic reduction require some explanation. As mentioned before, these values included the beginning of the rating task, the period when participants moved from the initial start position to a stimulus-related value. The relatively steep rise over mm. 1–3, shown in **Figure 5**, reflected participants' gradual transition from the start position of the rating interface. Another distinguishing moment was the dramatic drop in m. 12. In the harmonic reduction, this measure presents an audible leap downwards of a fifth in the bass line (chords 10–11 in **Figure 1**). This (relatively) dramatic tonic landing gives the impression of a final cadence followed by a harmonically conservative coda as opposed to the beginning of a new phrase. The strength of the drop in tension at this moment was shared by the first and second ratings of this excerpt, indicating that the perceptual experience was consistent despite exposure to the performed version, which encouraged a different tension profile over the same sequence of chords.

Comparison of tension-rating change activity between excerpts

Additional interesting contrasts can be seen in the rating-change activity levels in **Figure 6**. Each graph in the figure presents rating-change activity levels for each version over a phrase of music. Each line represents the proportion of responses showing rating increases or decreases in the half-second following that point in the music, calculated for every 16th note onset, or 1/12 of every measure. The lines in the positive range reflect the activity levels for increases in ratings greater than 1% of the rating scale: a line close to zero indicates few reported increases in tension at that point in the music; moments where the line rises above 0.5 in the positive range indicate that over half the participants are reporting tension increases during the corresponding half-second interval. The lines below zero report decreases in ratings; when most participants are reporting decreases in tension, the activity-level line falls below 0.5 in the negative range. Plots of activity-level time series make it easy to see when there is little total rating-change activity (when both lines per stimulus are close to zero), when there are contradictory rating changes between responses, and when ratings alternate between periods of widespread increases and decreases. **Figure 6** reports activity over shorter time frames of 0.5 s. The finer resolution is useful for avoiding accidental overlaps of contrary rating changes. However, each frame is unlikely to capture all participants' responses to salient events since response lags vary across participants and contexts. The activity-level time series for each verse and harmonic reduction in **Figure 6** are annotated with bars and dots to highlight moments in each excerpt with exceptionally high change activity ($p < 0.05$), as described in section Local Coordination Analysis. The following discussion will concentrate on these "significant" moments as a way to contrast ratings for each verse and the harmonic reduction.

Rating-change activity levels for the harmonic reduction are shown in black in **Figure 5**. Activity coordination for this excerpt was shaped by the relatively high activity levels in the first half of each measure in response to each new chord or note on the downbeat. Since the time between note events was long (2.5 s), the rating changes may also have been delayed in comparison to naturalistic stimuli because participants did not have precise beat entrainment to help them anticipate the changes at the next downbeat. Nonetheless, these note events provoked widespread changes in tension ratings with little disagreement as to direction of change. The only measure that showed significantly coordinated rating changes in both directions was m. 10 (corresponding to chord 9 in the reduction shown in **Figure 1**). Around the downbeat, there was a concentration of participants reporting decreases in tension, followed rapidly by a significant number of reported increases. Looking directly at the individual responses, only 10 of 25 subjects indicated increases during the measure, while seven indicated only decreases. Surprisingly, six others—a quarter of the participant group—initially reported decreases in tension in the first moments of the measure before changing directions to indicate an increase 0.5–1 s later; no responses showed the reverse pattern. This change may come from a conflict between participants' expectations and the stimulus, or perhaps these listeners reinterpreted the event after it had sounded. Although the material presented is controlled and sparse, this moment in particular appeared to be controversial and potentially dynamic in the ears of the listeners.

The last verse of the performance had the highest average tension values of all excerpts, as reported in **Table 4** and **Figure 5**. The difference between Verse 4 and the other verses was strongest in the first phrase, with the gain gradually lost over the second phrase. The separation began at the start of m. 1, throughout which increases persisted at significant levels for the full duration of the measure, in contrast to the shorter-lived activity in other verses. Many factors may have contributed to these reported changes in tension. Verse 4 was neither the loudest nor the fastest excerpt, although both of these features were increasing during this interval. Verse 3, the slowest and softest verse, did not yield strikingly lower tension ratings than its predecessors. However, from a qualitative perspective, the beginning of Verse 4 sounds loud and emphatic after such restraint. The sung words at the beginning of Verse 4 come across with great emphasis—the melodic passage is performed for the first time with detached articulation, adding to the contrast. Lastly, structure may also have had an influence; it is possible through these and other cues that listeners were aware that this verse was the last.

Contrasts between repetitions are rarely discussed in models of musical tension, despite their importance in performance practice. Performers generally make contrasts in returning material, particularly when semantic differences such as lyrics support a difference. While the overall tension averages were very similar, there were many other interesting contrasts in rating-change activity between verses that co-occurred with deliberate differences in the performances of the same melodic line and accompaniment.

The second phrase (middle graph, **Figure 6**) features the harmonic extremes of the piece. This phrase consists of two iterations

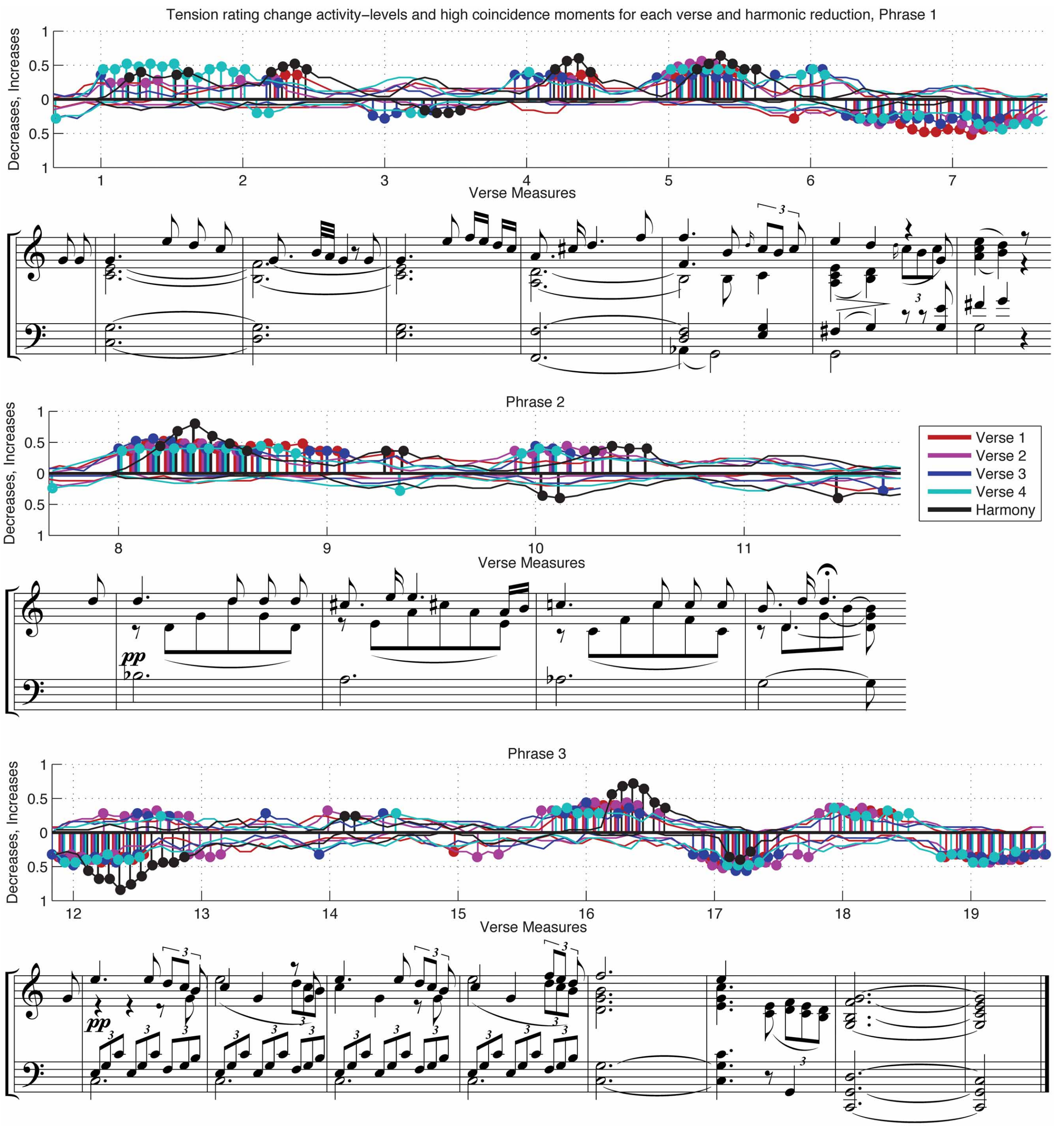


FIGURE 6 | Rating-change activity-level time series of the harmonic reduction and individual verses. Bars and dots mark moments with exceptional activity levels ($p < 0.05$, as assessed locally). The notated score includes the vocal line, denoted by the upward stems and broken beaming, with the piano accompaniment on the top staff; the notation format is transcribed exactly as engraved in Schubert (1894) edition.

of a two-measure motive, the second a step down from the first, over new harmonic material. The average tension ratings for all the verses showed a gradual increase in tension that peaked in m. 11. This smooth ascent seems in direct contradiction to the fact that many musical cues would suggest that the second two measures should be lower in tension than the first (in particular the cues encoded for the predictive model). Loudness, tempo, melody, and bass pitch height, even the harmonic tension ratings do not suggest a higher tension level for mm. 10–11. What else in the music could be driving these ratings? Though hard to quantify, this second phrase was performed without release or resolution. Through legato and upward motion in the melodic line, the singer communicated sustained tension—a feeling of not letting go; the last note of the phrase was even extended by a fermata. Though decreases in loudness and tempo often express a move toward calm and relaxation, in this case they might have added to the perceived tension as a signal of active restraint.

The activity levels revealed another layer of complication for interpreting the tension ratings of this second phrase. Although most participants did show a slight increase in tension ratings in mm. 9 and 11, the most concentrated activity occurred in mm. 8 and 10, as can been seen in **Figure 6**. The "peak" in tension ratings seen in the averages at m. 11 is only the product of diffuse rating changes, including decreases during the fermata, and not a widely shared response to a specific event in the music. It may be that activity levels relate more directly to the feature variations than the average tension ratings, at least for this second phrase.

The specific timing of tension-rating changes to the cadential motion in mm. 16–19 of the performed verses and mm. 16–17 in the harmonic reduction point to the importance of context and clarity of cues. Measures 18–19 in the performed recording were very similar to chords 16–17 in the reduction in terms of harmony, timbre, and voicing. However, the ratings for mm. 18–19 in the performance were never as coherent as those of the harmonic reduction. The ratings for the performed verses anticipated the final event, a tonic arrival, decreasing considerably before its onset. There are two plausible explanations for this: metrical anticipation and contextual richness. As mentioned above, the time between events in the harmonic reduction was too long for participants to predict the exact moment of the next downbeat. This lack of definite pulse would have forced participants to wait for cues, instead of anticipating them. In contrast, the performed version provided listeners with a rich metrical framework. Prepared by temporal expectations, they may have been reporting decreases early in anticipation of a downbeat that had been displaced by the expressive rubato of the performers. Alternately, these decreases may have been the result the stronger cadence in mm. 16–17 preceding this moment, during which participants also indicated tension decreases before the arrival of tonic. Other cues present in mm. 16–17 in the performed version could have encouraged tension-rating decreases before tonic arrival, including dynamic cues in the vocal line, which include timbre and loudness tapering over the last sung syllable of the verse. Like the confusion in m. 10, these moments are reminders that much of perceptual effects of music are not easily deduced from the notated page.

There are many other nuances of performance that could be explored using activity-level time series since they allow for the identification of salient moments of rating change. With more precise representations of tempo variation, articulation, and vocal timbre, it may be possible to quantify the factors behind inter-verse differences in m. 2, the transition to m. 4, the transition to m. 6, and at the end of m. 8. The timing of changes in this collection of responses point to the relevance of context—structural, metrical, and harmonic contrasts—between successive verses and phrases.

DISCUSSION

This study examined listener judgments of musical tension for a recording of the Schubert song Morgengruss and its harmonic reduction. We focused on two previously unexamined aspects of tension: the differences in processing timescales of disparate expressive and structural features contributing to tension and differences in tension ratings between the verses and the harmonic reduction. The methodologies employed in both set of analyses were novel, providing a fresh perspective on how tension reflects listeners' and performers' musical perceptions.

The first part of the analysis examined timescales of musical feature processing by using a modified version of Farbood's (2012) trend salience model of tension. The model describes tension in terms of a moving attentional window in time that represents a current tension trend. Additionally, a memory window has an increased effect on the magnitude of the perceived trend—whether negative or positive—if the current trend is continuing in the same direction as the previous trend in the memory window. The model was modified by using different memory and attentional window sizes for each feature instead of fixed durations across all features. The goal was to better understand how processing timescales differ between features contributing to tension. This was accomplished by finding the optimal window durations for each feature that resulted in a model prediction best correlated with the mean tension response. The features examined included harmony, pitch height, dynamics, onset frequency, and tempo.

The results indicated that harmony was processed across a far longer time span than all other features, having a combined attentional plus memory window duration of 20.5 s. The feature with the next-longest combined window duration was pitch height at 4 s. Onset frequency, tempo, and dynamics all had combined window durations of ∼2 s, however, dynamics had an optimal memory window of 0 s, indicating that loudness induces the most instantaneous tension response. These results align with more general perspectives on working memory such as Craik and Lockhart (1972), who theorized that higher levels of information abstraction are associated with longer persistence in memory. Deutsch and Feroe (1981) suggested that this theoretical framework should apply to music as well.

It is perhaps no surprise that listeners are responsive to highly local changes in loudness. Instinctual response to loudness is a basic, low-level function—anticipating and sensing approaching and retreating objects in the environment is important from an evolutionary perspective; in particular, listeners are highly sensitive to looming sounds (Neuhoff, 1998, 2001; Granot and Eitan, 2011). Tempo perception, predicated on beat induction, requires higher-level abstractions than loudness perception. The optimal combined time windows for tempo in fact encompass a time span (2.5 s) that is just beyond the upper limit for beat induction (2 s; London, 2012). The average tempo for the Morgengruss performance is 60 BPM, meaning that the optimal time windows in this case spanned approximately two and a half beats. The optimal window sizes for onset frequency were slightly shorter than for tempo (a combined duration of 2 s vs. 2.5 s), but they arguably fall under the same general timescale. Onset frequency and tempo are linked; in the simplest case, an increase in tempo of isochronous onsets corresponds directly to an increase in onset frequency.

The optimal window durations for pitch height were approximately twice as long as those for tempo and onset frequency. This reflects the likelihood that melodic contour is processed at a higher level of abstraction than all the other features examined here except harmony. Gestalt perception is a primary factor in melodic contour perception (cf. Meyer, 1956; Tenney and Polansky, 1980; Lerdahl and Jackendoff, 1983; Narmour, 1990), and the results suggest that—at least in the specific case of the Schubert—contour is evaluated over a time window that spans slightly longer than a measure.

The process of tonal induction and harmony perception requires higher-level cognitive abstractions than any other features, and this is evidenced by the long optimal time windows. These findings are in concordance with the result of a preliminary study by Farbood (2010) that examined how the constraints of working memory might affect perception of hierarchical tonal structures in the context of a proposed memory decay component to Lerdahl's (2001) tonal tension model. Farbood analyzed continuous tension responses to a one-minute Bach-Vivaldi excerpt using regression analysis that included harmonic tension, melodic contour, and onset frequency as independent variables. These features were described in terms of change over time spans ranging from 0.25 to 20 s. The results showed that changes in harmony best correlated with the tension data when the time differential was around 10–12 s, while other features best fit the data at a time differential of around 3 s. These results indicated far longer memory effect for harmony compared to other features. It should be noted, however, that harmonic processing is by nature subject to a longer timescale of processing due to the temporal trajectory of harmonic progressions. This potential confound may be a contributing factor in these converging results.

The second part of this study examined the differences in tension ratings for the varying interpretations of the Morgengruss score across the four performed verses and the harmonic reduction. While the average tension ratings for the verses were very similar, differences in how participants reported tension changes support the importance of performance decisions that provide nuance to the interpretation of the score. Performer-controlled features such as loudness and tempo, less-easily-quantified articulation and sustain, and contrasts between successive verses were each highlighted as likely factors for these subtle but substantial differences in the tension ratings.

Analysis of ratings of the reduction underscored the importance of harmony by providing an explanation for the high tension ratings in the second phrase of each verse. It also offered an interesting case study for analyzing activity in tension ratings; having a reduction representing at least one complex feature made it easier to make sense of the complicated information contributing to the dynamics of average tension ratings.

There is some precedent for comparing tension or other continuous ratings between interpretations of a common work (Fredrickson and Johnson, 1996; Goodchild et al., 2010), as well as comparisons between section repetitions within pieces (Livingstone et al., 2012). In the current study, the expressive range and multiple verses in Morgengruss provided particularly fruitful data for exploring the effects of performance parameters. In prior work, most comparisons of performed interpretations of a common score have been limited to noting dramatic contrasts or making vague claims due to insufficient tools for assessing the reliability of differences or the significance of small changes in the average time series. By using rating-change activity levels and a novel approach to assessing the significance of coordination at each moment of the music, it was possible to detect salient differences in tension-rating behavior at a finer temporal scale than employed in previous work. Combining activity analysis with local coordination measures also provided new evidence that contrary rating activity is a common and important phenomenon that justifies using representations of continuous responses that acknowledge the multiplicity of perceptions reported.

While the coordination analyses of the tension-rating changes were data-driven and robust, the comparisons between responses across different stimuli were not performed systematically. Interpretation of contrasts between verses were informed by musical experience and need to be tested, perhaps with other examples of controlled stimuli. The study of Romantic lieder in particular would be greatly improved with the incorporation of a dynamic model of tempo prediction, sensitive to the grouping implications of different degrees of rubato and descriptors of the sung timbre and articulation.

Comparisons of the tension ratings between verses also raises the question of structure and tension judgments, in particular, the importance of contrast between verses. How the same moment in a verse is treated from one iteration to the next warrants more attention since it might be a means to capture the role of form in musical memory and the continuous experience of music. What would it mean for a model to include information about the previous presentation of the same material? While the idea of contrast between repetitions affecting tension is intuitive and familiar to performers, testing this hypothesis would require more instances of specially composed stimuli.

The effectiveness of any analysis of tension is dependent on the inclusion of a complete set of features contributing to tension. Although we examined several key parameters, they do not represent an exhaustive set of features that account for all tension variations perceived by the performers and listeners of the Schubert. Perhaps the most significant feature that requires future examination is timbre. Defining the perceptual dimensions of timbre is a difficult task, and that is perhaps one reason why timbre is underexplored in the tension literature. Prior studies that have explored timbral tension in some capacity have looked at features such as roughness, brightness, spectral flatness, and density (von Helmholtz, 1877; Plomp and Levelt, 1965; Hutchinson and Knopoff, 1978; Nielsen, 1987; Krumhansl, 1996; Pressnitzer et al., 2000; Dean and Bailes, 2010). Further work needs to be done to empirically investigate and confirm the primary components that contribute to timbral tension. Given this knowledge, it might be possible to better understand the nuances in a solo vocal performance such as the Schubert, where tone and articulation are of great expressive importance.

Although additional experiments using more varied musical stimuli would undoubtedly strengthen and solidify these findings, the analyses and observations made in this study present new perspectives on tension. They help illuminate the complex process of how tension is affected by performance decisions and how listeners respond to those differences. Having a deeper understanding of tension perception can help us better grasp the interplay between expressive performance, listener interpretation, and musical structure.

REFERENCES

Alluri, V., Toiviainen, P., Jääskeläinen, I. P., Glerean, E., Sams, M., and Brattico, E. (2011). Large-scale brain networks emerge from dynamic processing of musical timbre, key and rhythm. *Neuroimage* 59, 3677–3689. doi: 10.1016/j.neuroimage.2011.11.019

Bigand, E., and Parncutt, R. (1999). Perceiving musical tension in long chord sequences. *Psychol. Res.* 62, 237–224. doi: 10.1007/s004260050053

Bigand, E., Parncutt, R., and Lerdahl, F. (1996). Perception of musical tension in short chord sequences: the influence of harmonic function, sensory dissonance, horizontal motion, and musical training. *Percept. Psychophys.* 58, 125–141. doi: 10.3758/BF03205482

Brainard, D. H. (1997). The psychophysics toolbox. *Spat. Vis.* 10, 433–436. doi: 10.1163/156856897X00357

Craik, F. I. M. (2002). Levels of processing: past, present... and future? *Memory* 10, 305–318. doi: 10.1080/09658210244000135

Craik, F. I. M., and Lockhart, R. S. (1972). Levels of processing: a framework for memory research. *J. Verb. Learn. Verb. Be.* 11, 671–684. doi: 10.1016/S0022-5371(72)80001-X

Dean, R. T., and Bailes, F. (2010). Time series analysis as a method to examine acoustical influences on real-time perception of music. *Empir. Musicol. Rev.* 5, 152–175.

Deutsch, D., and Feroe, J. (1981). The internal representation of pitch sequences in tonal music. *Psychol. Rev.* 88, 503–522. doi: 10.1037/0033-295X.88.6.503

Eerola, T., and Vuoskoski, J. K. (2010). A comparison of the discrete and dimensional models of emotion in music. *Psychol. Music* 39, 18–49. doi: 10.1177/0305735610362821

Farbood, M. (2010). "Working memory and the perception of hierarchical tonal structures," in *Proceedings of the 11th International Conference of Music Perception and Cognition*, eds S. M. Demorest, S. J. Morrison, and P. S. Campbell (Seattle, USA), 119–222.

Farbood, M. M. (2012). A parametric, temporal model of musical tension. *Music Percept.* 29, 387–428. doi: 10.1525/mp.2012.29.4.387

Fredrickson, W. E. (1999). Effect of musical performance on perception of tension in Gustav Holst's First Suite in E-flat. *J. Res. Music Educ.* 47, 44–52. doi: 10.2307/3345827

Fredrickson, W. E., and Johnson, C. M. (1996). The effect of performer use of rubato on listener perception of tension in Mozart. *Psychomusicology* 15, 78–86. doi: 10.1037/h0094078

Genesis (2009). Loudness toolbox. Available online at: http://www.genesis-acoustics.com/en/index.php?page=32

Glasberg, B., and Moore, B. (2002). A model of loudness applicable to time-varying sounds. *J. Audio Eng. Soc.* 50, 331–342.

Goodchild, M., Gingras, B., Asselin, P., and McAdams, S. (2010). "Construction and perception of formal structure in an unmeasured prelude for harpsichord," in *Proceedings of the 11th International Conference of Music Perception and Cognition*, eds S. M. Demorest, S. J. Morrison, and P. S. Campbell (Seattle, USA).

Granot, R. Y., and Eitan, Z. (2011). Tension and dynamic auditory parameters. *Music Percept.* 28, 219–246. doi: 10.1525/mp.2011.28.3.219

Huron, D. (2006). *Sweet Anticipation: Music and the Psychology of Expectation*. Cambridge, MA: MIT Press.

Hutchinson, W., and Knopoff, L. (1978). The acoustic component of western consonance. *Interface* 7, 1–29. doi: 10.1080/09298217808570246

Ilie, G., and Thompson, W. F. (2006). A comparison of acoustic cues in music and speech for three dimensions of affect. *Music Percept.* 23, 319–329. doi: 10.1525/mp.2006.23.4.319

Kleiner, M., Brainard, D., and Pelli, D. (2007). What's new in psychtoolbox-3? *Perception* 36, ECVP Abstract Supplement. doi:10.1068/v070821

Krumhansl, C. (1996). A perceptual analysis of Mozart's Piano Sonata K. 282: segmentation, tension, and musical ideas. *Music Percept.* 13, 401–432. doi: 10.2307/40286177

Krumhansl, C. L. (1997). An exploratory study of musical emotions and psychophysiology. *Can. J. Exp. Psychol.* 51, 336. doi: 10.1037/1196-1961.51.4.336

Lehne, M., Rohrmeier, M., and Koelsch, S. (2013). Tension-related activity in the orbitofrontal cortex and amygdala: an fMRI study with music. *Soc. Cogn. Affect. Neurosci*. doi: 10.1093/scan/nst141. [Epub ahead of print].

Lerdahl, F. (1996). Calculating tonal tension. *Music Percept.* 13, 319–363. doi: 10.2307/40286174

Lerdahl, F. (2001). *Tonal Pitch Space*. New York, NY: Oxford University Press.

Lerdahl, F. (2013). "Tension and expectation in a Schubert song," in *Musical Implications: Essays in Honor of Eugene Narmour*, eds L. Bernstein and A. Rozin (New York, NY: Pendragon Press), 255–274.

Lerdahl, F., and Jackendoff, R. (1983). *A Generative Theory of Tonal Music*. Cambridge, MA: MIT Press.

Lerdahl, F., and Krumhansl, C. L. (2007). Modeling tonal tension. *Music Percept.* 24, 329–366. doi: 10.1525/mp.2007.24.4.329

Livingstone, S., Palmer, C., and Schubert, E. (2012). Emotional response to musical repetition. *Emotion* 12, 552–567. doi: 10.1037/a0023747

London, J. (2012). *Hearing in Time: Psychological Aspects of Musical Meter*. 2nd Edn. New York, NY: Oxford University Press. doi: 10.1093/acprof:oso/9780199744374.001.0001

Lychner, J. A. (1998). An empirical study concerning terminology relating to aesthetic response to music. *J. Res. Music Educ.* 46, 303–319. doi: 10.2307/3345630

Meyer, L. (1956). *Emotion and Meaning in Music*. Chicago, IL: University of Chicago Press.

Narmour, E. (1990). *The Analysis and Cognition of Basic Melodic Structures*. Chicago, IL: University of Chicago Press.

Narmour, E. (1992). *The Analysis and Cognition of Melodic Complexity: The Implication-Realization Model*. Chicago: University of Chicago Press.

Neuhoff, J. G. (1998). A perceptual bias for rising tones. *Nature* 395, 123–124. doi: 10.1038/25862

Neuhoff, J. G. (2001). An adaptive bias in the perception of looming auditory motion. *Ecol. Psychol.* 13, 87–110. doi: 10.1207/S15326969ECO1302_2

Nielsen, F. (1987). "Musical 'tension' and related concepts," in *The Semiotic Web '86: An International Yearbook*, eds T. A. Sebeok and J. Umiker-Sebeok (Berlin: Mouton de Gruyter), 491–514.

Nielsen, F. V. (1983). *OplevelseafMisikalsk Spending (The experience of musical tension)*. Copenhagen: Akademisk Forlag.

Olsen, K. N., Stevens, C. J., and Tardieu, J. (2010). Loudness change in response to dynamic acoustic intensity. *J. Exp. Psychol. Hum. Percept. Perform.* 36, 1631–1644. doi: 10.1037/a0018389

Palmer, C. (1996). Anatomy of a performance: sources of musical expression. *Music Percept.* 13, 433–453. doi: 10.2307/40286178

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spat. Vis.* 10, 437–442. doi: 10.1163/156856897X00366

Pipa, G., Wheeler, D. W., Singer, W., and Nikoliæ, D. (2008). NeuroXidence: reliable and efficient analysis of an excess or deficiency of joint-spike events. *J. Comput. Neurosci.* 25, 64–88.

Plomp, R., and Levelt, W. J. M. (1965). Tonal Consonance and Critical Bandwidth. *J. Acoust. Soc. Am.* 38, 548–560. doi: 10.1121/1.1909741

Pressnitzer, D., McAdams, S., Winsberg, S., and Fineberg, J. (2000). Perception of music tension for nontonal orchestral timbres and its relation to psychoacoustic roughness. *Percept. Psychophys.* 62, 66–80. doi: 10.3758/BF03212061

Rozin, A., Rozin, P., and Goldberg, E. (2004). The feeling of music past: how listeners remember musical affect. *Music Percept.* 22, 15–39. doi: 10.1525/mp.2004.22.1.15

Schubert, F. (1894). *Sämtliche einstimmige Lieder und Gesänge*. ed E. Mandyczewski. Leipzig: Breitkopf & Härtel.

Schubert, E. (2002). Correlation analysis of continuous emotional response to music: correcting for the effects of serial correlation. *Music Sci.* 5, 213–236.

Schubert, E. (2004). Modeling perceived emotion with continuous musical features. *Music Percept.* 21, 561–585. doi: 10.1525/mp.2004.21.4.561

Schubert, E. (2012). Reliability issues regarding the beginning, middle and end of continuous emotion ratings to music. *Psychol. Music* 41, 350–371. doi: 10.1177/0305735611430079

Stevens, S. S. (1936). A scale for the measurement of a psychological magnitude: loudness. *Psychol. Rev.* 43, 405–416. doi: 10.1037/h0058773

Tenney, J., and Polansky, L. (1980). Temporal gestalt perception in music. *J. Music Theor.* 24, 205–241. doi: 10.2307/843503

Toiviainen, P., and Krumhansl, C. (2003). Measuring and modeling real-time responses to music: the dynamics of tonality induction. *Perception* 32, 1–27. doi: 10.1068/p3312

Upham, F. (2011). *Quantifying the temporal dynamics of music listening: a critical investigation of analysis techniques for collections of continuous responses to music.* Master thesis. McGill University, Montreal, QC.

Upham, F. (2012). "Limits on the application of statistical correlations to continuous response data," in *Proceedings of the 12th International Conference on Music Perception and Cognition*, eds. E. Cambouropoulos, C. Tsougras, P. Mavromatis, and K. Pastiades (Thessaloniki), 1037–1041.

Upham, F. (2013). *Activity analysis toolbox (1.0) [MATLAB toolbox]*. Available online at: https://github.com/finn42/ActivityAnalysisTB.git

Upham, F., and Farbood, M. M. (2013). "Coordination in musical tension and liking ratings of scrambled music," in *Meeting of the Society for Music Perception and Cognition. SMPC*, eds M. Schutz and F. A. Russo (Toronto, ON), 148.

von Helmholtz, H. (1877). *Die Lehre von der Tonempfindungen als physiologische Grundlagefür die Theorie der Musik*. New York, NY: Dover.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 01 September 2013; accepted: 15 December 2013; published online: 30 December 2013.

Citation: Farbood MM and Upham F (2013) Interpreting expressive performance through listener judgments of musical tension. Front. Psychol. ***4****:998. doi: 10.3389/fpsyg.2013.00998*

This article was submitted to Cognitive Science, a section of the journal Frontiers in Psychology.

frontiers in
PSYCHOLOGY

ORIGINAL RESEARCH ARTICLE
published: 21 March 2013
doi: 10.3389/fpsyg.2013.00134

Individuality that is unheard of: systematic temporal deviations in scale playing leave an inaudible pianistic fingerprint

Floris Tijmen Van Vugt[1,2]*, **Hans-Christian Jabusch**[3] **and Eckart Altenmüller**[1]

[1] Institute of Music Physiology and Musicians' Medicine, University of Music, Drama, and Media, Hanover, Germany
[2] Lyon Neuroscience Research Center, CNRS-UMR 5292, INSERM U1028, University Lyon-1, Lyon, France
[3] Institute of Musicians' Medicine, University of Music "Carl Maria von Weber," Dresden, Germany

Edited by:
Bruno Gingras, University of Vienna, Austria

Reviewed by:
Ian E. Holliday, Aston University, UK
Sarah Creel, University of California San Diego, USA

***Correspondence:**
Floris Tijmen Van Vugt, Institute of Music Physiology and Musicians' Medicine, University of Music, Drama, and Media, Emmichplatz 1, 30175 Hanover, Germany.
e-mail: f.t.vanvugt@gmail.com

Whatever we do, we do it in our own way, and we recognize master artists by small samples of their work. This study investigates individuality of temporal deviations in musical scales in pianists in the absence of deliberate expressive intention. Note-by-note timing deviations away from regularity form a remarkably consistent "pianistic fingerprint." First, eight professional pianists played C-major scales in two sessions, separated by 15 min. Euclidian distances between deviation traces originating from different pianists were reliably larger than traces originating from the same pianist. As a result, a simple classifier that matched deviation traces by minimizing their distance was able to recognize each pianist with 100% accuracy. Furthermore, within each pianist, fingerprints produced by the same movements were more similar than fingerprints resulting in the same scale sound. This allowed us to conclude that the fingerprints are mostly neuromuscular rather than intentional or expressive in nature. However, human listeners were not able to distinguish the temporal fingerprints by ear. Next, 18 pianists played C-major scales on a normal or muted piano. Recognition rates ranged from 83 to 100%, further supporting the view that auditory feedback is not implicated in the creation of the temporal signature. Finally, 20 pianists were recognized 20 months later at above chance level, showing signature effects to be long lasting. Our results indicate that even non-expressive playing of scales reveals consistent, partially effector-unspecific, but inaudible inter-individual differences. We suggest that machine learning studies into individuality in performance will need to take into account unintentional but consistent variability below the perceptual threshold.

Keywords: piano scale, individuality, expertise, music, recognition

INTRODUCTION

Our actions are highly individual and we can tell people apart by how they move (Flach et al., 2004; Loula et al., 2005; Prasad and Shiffrar, 2009; Sevdalis and Keller, 2011). People may recognize those close to them by the way they sneeze or walk the stairs. Even when trying to achieve the same aim, the actions that are selected toward this aim and the way in which they are executed vary considerably between individuals. The human observer seems to rely on action simulation to recognize individuals by their movements, since recognition is generally stronger when distinguishing one's own performance from that of others (Jeannerod, 2003).

A first question is how movements from different individuals vary *physically*. Why are certain parameters of our actions remarkably stable between multiple iterations by the same person, and yet strikingly different between individuals? A second question is to what extent movements vary *perceptually*. For example, some movements may differ so subtly that the individual features are not distinguishable to a human observer under normal conditions.

Music is a suitable paradigm to study individuality since actions are directed toward a clearly defined auditory goal: when we play music, the aim is to make a certain sound. Furthermore, differences between performers are sometimes so salient that listeners will often refuse to listen to a musical piece that is a mere "cover" of the original. Music played by different individuals varies *physically*. For example, machine ensemble learning approaches are able to tell musical performers apart based on structural features such as timing and loudness differences (Stamatatos and Widmer, 2005) or kinematics (Dalla Bella and Palmer, 2011). The individuality is also *perceptual*. Indeed, non-musicians and musicians alike were able to recognize performances reliably (Gingras et al., 2011). Again, action simulation in the form of musical imagery appears to play a role in the recognition process. For example, piano players turn out to be capable of recognizing their own playing from a few months previously, even if the sound was switched off at the time of the recording (Repp and Knoblich, 2004).

In music performance recognition the differences in sound that different players produce are often understood as a result of their artistic individuality. However, there is no reason to assume that the individuality in the way we walk serves any particular purpose. Indeed, even task-irrelevant sounds matching a golf swing are recognized significantly better than chance (Murgia et al., 2012).

On the other hand, individuality in music performance is tacitly assumed to define a performer's unique artistic identity. But we have to date no empirical validation of the extent to which individuality in music performance is deliberate. The study coming closest to answering this question requested pianists to play mechanically, and found that recognition was somewhat impaired for these inexpressive recordings (Gingras et al., 2011). However, even metronomic playing has been shown to contain the same timing patterns as expressive playing, but to a lesser extent (Repp, 1999a). To avoid this problem, we instead investigated the playing of musical scales (Wagner, 1971; MacKenzie and Van Eerd, 1990). When participants are instructed to play a scale as regularly as possible and in a legato style, there is a clear auditory target of perceptual evenness and it is understood that the task at hand is not to play scales in one's own particular way. In other words, isolated scales are not thought of as expressive musical materials. There is some objective standard and trying to meet it is a merely technical task.

Yet, it is found that musical scales show systematic temporal deviations (MacKenzie and Van Eerd, 1990; van Vugt et al., 2012). These deviations are thought of as the result of perceptual distortions (Drake, 1993), residual expressive timing (Repp, 1999a), or of some note transitions involving more difficult movements (Engel et al., 1997).

Our question is whether these temporal deviations are individual in the same way that expressive performance is. We restrict our attention to timing of note onsets, discarding information such as differences in loudness and note duration. In Experiment I, we first established timing deviations of individual notes (van Vugt et al., 2012). The resulting timing profile is then used to recognize pianists across two sessions, separated by 15 min. In this way, we aim to establish individuality that is physically present in the timing of musical scales. In Experiment II, we then proceed to assess whether the timing differences can be perceived by musically trained observers. In Experiment III we investigate the role of auditory feedback in the formation of these timing profiles. Finally, in order to investigate to what extent these timing deviation profiles are stable, we follow a group of pianists over 27 months in Experiment IV.

EXPERIMENT I

MATERIALS AND METHODS

The data reported here were collected as part of a validation procedure for a scale unevenness quantification method published elsewhere (Jabusch et al., 2004). Eight pianists (six female) were recruited from the student/teacher pool at the Hanover University of Music and were 24.3 (SD 2.4) years old. All but one were right-handed ($M = 57.2$, SD = 66% right-handed according to the Edinburgh handedness inventory). None of the participants reported any neurological condition. Participants played on a MP 9000 MIDI keyboard (Kawai, Krefeld, Germany). The keyboard's digital music interface (MIDI out) signal was captured on a PC using a commercially available sequencer software (Musicator Win, version 2.12; Music Interactive Technology, Bergen, Norway).

Participants were requested to play two-octave C-major scales beginning with the C (131 Hz) one octave below the middle C and ending with the C (523 Hz) one octave higher than the middle C. Ascending and descending scales were interleaved. The instruction to the participants was to play as evenly as possible, without expression, and in a legato style at mezzo-forte loudness. A metronome gave a beat at 120 BPM and the instructions were to play at four notes per metronome beat, resulting in eight notes per second. Participants performed 10–15 scales with the right hand and with the left hand (*first measurement*). After a 15 min break, the procedure was repeated (*follow-up*).

ANALYSIS OF SCALE TIMING

First, we isolated correctly performed scale runs, discarding those containing errors or surplus notes. We then converted the note values to their rank in the C-major scale (i.e., C has rank 0, D has rank 1, E has rank 2, etc., up to C$''$ with rank 14) and performed a least-square straight line fit to this set of pairs of rank and timing. This allowed us to compute for each note the expected onset time (according to this fit) and then the deviation of the timing of the actually measured onset (in ms) (van Vugt et al., 2012). We performed this fit for all scale runs and then pooled the results by hand (left or right), playing direction (inward or outward) and note, calculating the mean lateness (in ms) for that condition. The result was a 2 (hands) × 2 (directions) × 15 (notes) matrix of timing deviations, which we will refer to as our irregularity trace. As an illustration, **Figure 1A** shows the irregularity trace for right hand ascending scales in one pianist in the two measurement sessions, and **Figure 1B** for two different pianists. It is clear that the irregularity traces originating from the same pianist (**Figure 1A**) are strikingly similar, whereas those originating from different pianists (**Figure 1B**) are qualitatively different. This is the observation that our analysis (described below) aims to capture.

Additionally, we calculated the unevenness of the scale in accordance with a previously established protocol (Jabusch et al., 2004) as follows. For each correct scale run, the intervals between the consecutive note onsets were calculated and then we took the standard deviation of these. For each hand, direction, and recording (first or follow-up) we took the median of the standard deviations of the scale runs (in ms). The higher this unevenness score, the more temporally irregular the scales.

In ANOVAs we report η^2_G as the generalized effect size (Bakeman, 2005). Following musicological notational convention, we will refer to the notes in the scale as 1, 2, 3, 4, 5, 6, 7, 1$'$, 2$'$, 3$'$, 4$'$, 5$'$, 6$'$, 7$'$,1 $''$, in ascending order.

RESULTS

Preliminaries

First, we isolated the correctly played scales, yielding an average total of 11.7 (SD 0.97) scales per person and condition. As a control analysis, we used the number of scales as an outcome measurement in an ANOVA that revealed no significant difference according to hand [$F(1, 7) = 3.43$, $p = 0.11$], direction [$F(1, 7) \approx 0.00$, $p \approx 1.00$], recording session [$F(1, 7) = 1.19$, $p = 0.74$] nor any interaction effect [all $F(1, 7) < 0.11$]. We can conclude that there is no selection bias due to the discarding of scales.

Now we turn to the unevenness measure (the standard deviation of the inter-keystroke-intervals). ANOVA yielded a significant main effect of hand [$F(1,7) = 5.73, p < 0.05, \eta^2_G = 0.04$], showing

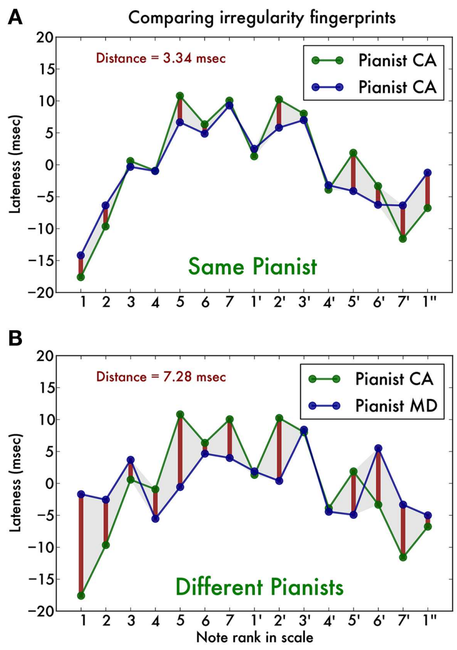


FIGURE 1 | Illustration of the note onset timing traces of two typical pianists, showing only the right hand ascending scale timings. One pianist (CA) was recorded playing two-octave C-major scales. Using a previously established technique, we are able to determine the precise timing of each individual note (for further details see text). **(A)** The note-by-note temporal deviation (in ms) is strikingly similar between the two recordings (blue and green line). The red vertical bars and shaded area indicate the temporal distance between the traces, which is on average around 3 ms. **(B)** Comparison of CA's temporal deviation trace with that of a different pianist (MD). The traces are qualitatively different, which is captured by a higher temporal distance of around 7 ms.

that left hand scales were played more unevenly (mean unevenness 9.19 ms, SD 1.67) than right hand scales (mean unevenness 8.44 ms, SD 1.81). This replicates a previous finding (Kopiez et al., 2011). There was no main effect of playing direction [$F(1,7) = 0.01, p = 0.92$] nor of recording session [$F(1,7) = 1.00$, $p = 0.35$] but there was a two-way interaction between direction and recording [$F(1,7) = 7.00, p = 0.03, \eta^2_G = 0.02$], showing that although outward scales were played equally evenly across the sessions, inward scales were more even in the follow-up session (unevenness 8.43 ms, SD 1.86) than in the first session (unevenness 9.13 ms, SD 2.33), perhaps revealing a habituation effect.

Recognizing individual pianists

A salient feature of the temporal traces is that they are highly individual: traces from the same individual but different sessions vary little, whereas traces from different pianists vary much more (**Figure 1**). To quantify this observation, we define the temporal distance as the Euclidian distance between any pair of vectors representing the irregularity traces. That is, we calculated the sum of squares of the item-by-item distances. Then we divided this by the number of notes in the traces (15 notes for a two-octave scale). Finally, we took the square root to yield a distance value in ms. First we calculate these distances for each of the two hands, two directions separately. We find that irregularity traces originating from the same pianist have a distance of 3.42 ms (SD 0.89), whereas those originating from different pianists have a distance of 7.24 ms (SD 0.54) (**Figure 4**). ANOVA with distance as dependent variable shows a significant main effect of self vs. other [$F(1, 7) = 108.18$, $p < 0.001$, $\eta^2_G = 0.79$] but no effect of hand [$F(1, 7) = 0.55$, $p = 0.48$] nor playing direction [$F(1, 7) = 0.30$, $p = 0.60$] nor any interaction effect [all $F(1, 7) < 1.1$].

As a result, we designed the simplest possible classification algorithm as follows. Our algorithm is given a database of the irregularity traces for the first measurements of each of the eight pianists. Then it is presented each of the follow-up irregularity traces, without the player label, and its task is to match each pianist to one of the traces in its database. Our algorithm simply chooses the irregularity trace that matches most closely.

This procedure is performed separately for the four sets of average irregularity traces from the two hands and two playing directions. Classification was flawless (100%) for all the right hand scales (inward and outward), as well as the left hand outward scales. In the left hand inward scales, six pianists are classified correctly and two incorrectly. Chance is at 0.125 recognition rate, meaning that in all cases classification is significantly better than chance [binomial $p < 0.001$, 95% confidence interval = (0.35, 0.97) for the left hand inward scales and (0.63, 1.0) for the other cases]. When instead of the complete irregularity trace (15 data points per two-octave scale) we used only the unevenness (one data point per two-octave scale) classification rate dropped to between 0.25 and 0.5, which exceeded chance performance only for the right hand inward scales [binomial $p = 0.01$, 95% confidence interval = (0.16, 0.84)].

The Euclidian distance is not necessarily the only or best way to quantify the (dis)similarity between irregularity traces. To illustrate this, we perform the same analysis, but this time we compute the correlation (Pearson r) between pairs of irregularity traces. ANOVA on the Fisher r-to-z transformed correlation coefficients shows a main effect of self vs. other [$F(1, 7) = 63.92$, $p < 0.001$, $\eta^2_G = 0.74$], showing that correlations between irregularity traces from the same pianists are higher [$z(r) = 1.39$, SD 0.42] than irregularity traces from different pianists [$z(r) = 0.40$, SD 0.21]. There is no effect of hand except for a trend [$F(1, 7) = 5.40$, $p = 0.05, \eta^2_G = 0.03$], nor a main effect of direction [$F(1,7) = 2.76$, $p = 0.14$]. Of the interaction effects only that between hand and direction [$F(1, 7) = 11.50$, $p = 0.01$, $\eta^2_G = 0.10$] is significant [all other $F(1, 7) < 1.05$], revealing that whereas left hand traces correlate equally in both playing directions, right hand inward scales correlate higher than outward scales.

We re-ran our recognition algorithm with the only difference that this time, given an irregularity trace to recognize, it chose the irregularity trace that showed the greatest correlation. Recognition rates are identical to those for Euclidian distance: flawless in all but the case of left hand inward scales with six out of two correctly classified (hence still exceeding chance performance).

Comparing irregularity traces of the same pianist

So far, we have only compared the irregularity traces produced by the same hand and in the same playing direction but by different pianists. How do the traces produced by the same pianist but by different hands and different directions compare? We argue that these comparisons may provide crucial insight into what causes the timing deviations (**Figure 2A**). Our reasoning was as follows. If the temporal deviations result from remnants of expressive timing (Repp, 1999a), then we expect irregularity traces that sound similar to be more similar. That is, we expect the left hand inward and right hand outward traces to be closest together (since they have the same auditory result, modulo octave differences), and similarly the right hand outward and left hand inward scales to be close. If, on the other hand, the temporal deviation traces are mostly determined by biomechanical or neuromuscular factors, then we expect traces generated by the same movements to be closer together than those generated by different movements (**Figure 2B**). More specifically, the pairs of inward and pairs of outward scales are expected to be closer together than pairs with an inward and outward scale.

Furthermore, note that in all these comparisons we have aligned the irregularity traces in time (in the order in which they are played) and not in space (the order in which they appear on the keyboard). That means, when we compare left hand inward and right hand outward scales, they are the same movement in time, but mirrored in space.

An ANOVA with distance as dependent measure revealed a main effect of movement [$F(1, 7) = 7.63$, $p = 0.03$, $\eta^2_G = 0.10$], reflecting that distances between irregularity traces produced by the same movement are shorter (6.47 ms, SD 0.85) than those produced by different movements (7.64 ms, SD 1.58) (**Figure 2C**). That is, the results are in line with the hypothesis that the temporal deviations are mostly neuromuscular in nature. No other factor has a main effect [all $F(1, 7) < 1.6$] and there were no interactions [all $F(1, 7) < 2.0$].

Effector-specificity of the individuality

To what extent is the individuality in the traces specific to the effector (i.e., hand)? To answer this question, we repeated the analysis above, but comparing the distances across hands within and between pianists. That is, we computed the distance between left and right hand irregularity traces for the same movement direction (inward or outward) and for either the same pianist or different pianists. We found a main effect of same vs. different pianist [$F(1, 7) = 28.35$, $p = 0.001$, $\eta^2_G = 0.01$], revealing that cross-hand distances are smaller between traces from the same pianist ($M = 6.41$, SD = 0.87 ms) than traces from different pianists ($M = 7.47$, SD = 0.42 ms). There were no main effects of hand, direction or recording, nor any interaction effects [all $F(1, 7) < 2.74$, $p > 0.14$].

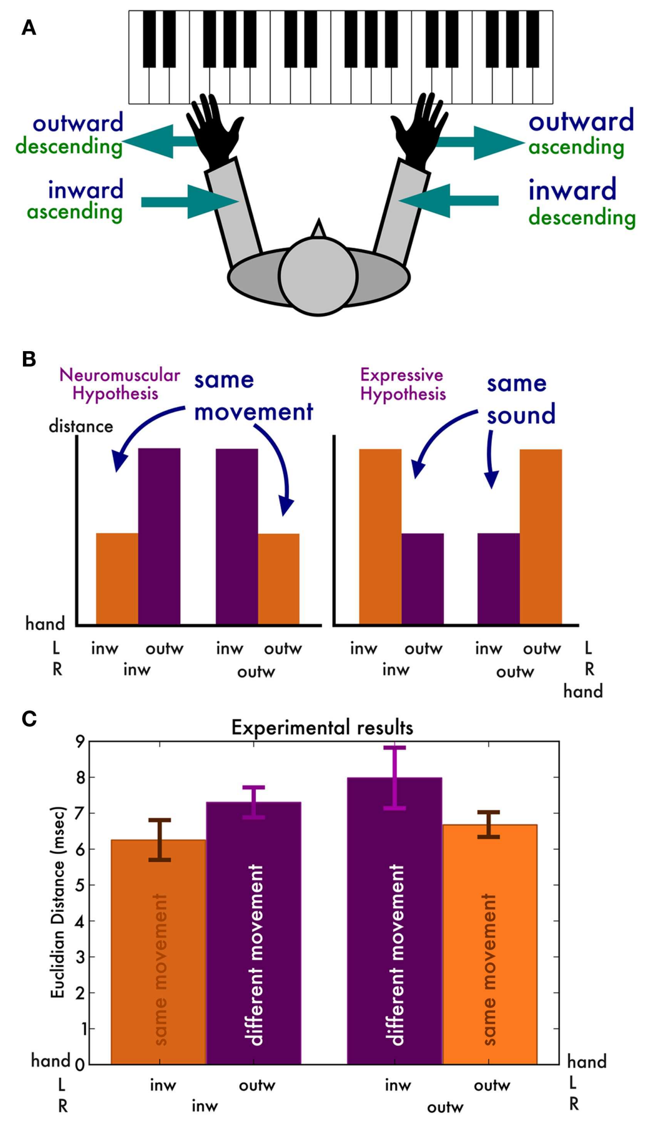


FIGURE 2 | (A) Overview of the body-central directions (inward and outward, in blue) and the keyboard-central directions (ascending and descending, in green). **(B)** Predictions of the two hypotheses. If the irregularity traces mostly stem from neuromuscular constraints, we expect traces originating from the same movements to be similar. If they originate mostly from residual expression, we expect traces producing the same sounds to be similar. **(C)** Experimental results, in line with the neuromuscular hypothesis.

DISCUSSION

Let us pause an instant to take stock. We have shown that pianists do not play scales perfectly regularly. Rather, consistent temporal deviations are present. For the first time we show that these deviations are not mere noise, since they are reliably reproduced across two recording sessions. Furthermore, differences between individuals are so pronounced that a surprisingly simple recognition algorithm is able to recognize pianists nearly flawlessly using the average timing profile of a dozen runs of two-octave scales. The algorithm works equally well when it matches irregularity traces by minimizing distance or by maximizing correlation.

An important observation is that the pianists' temporal irregularities are *qualitatively* different. If the irregularity profiles had been qualitatively the same, that is, the same vector simply multiplied by a coefficient, then recognition on the overall unevenness would perform as well as recognition using the entire irregularity trace. But we find the contrary: recognition using a simple overall unevenness metric (the median of the inter-keystroke-intervals) was barely above chance. We can conclude that it is the qualitative differences in the scale timing that enable us to tell the different pianists apart. Hence we can speak of a *pianistic fingerprint*.

What determines this temporal fingerprint? We showed that temporal irregularity traces generated by the same movement are more similar than those generating the same sound. As a consequence, the contribution of biomechanical constraints to these timing profiles must be stronger than expressive or perceptual influences. Furthermore, we found that the individuality in the traces is to some extent effector-independent: the two hands of the same pianist are less different than hands of different pianists. This suggests that the individuality is represented in cortical areas accessible to both effectors (Rijntjes et al., 1999).

In sum, temporal differences are physically present in the produced timing in musical scales. At this point, it remains unclear whether this individuality is also *perceptually* present: are human observers able to identify performers in the same way our algorithm could?

EXPERIMENT II

MATERIALS AND METHODS

Our perceptual experiment comprised two parts. In the first part (*recognition*), listeners (see details below) were presented with pairs of fingerprint recordings and asked to judge whether they originated from the same or different pianists. Essentially, participants were given the same task that our algorithm in Experiment I performed. In the second part (*irregularity threshold*), we investigated whether participants were able to pick up the temporal irregularities at all by establishing their psychophysical threshold for temporal irregularity. That is, participants were presented a single scale and had to judge whether it was regular (isochronous) or irregular.

Recognition test

We took the irregularity traces for the right hand ascending scales for three pianists (CA, ES, and TY) from the *first* and *follow-up* measurements in Experiment I. For each, we furthermore choose one alternative pianist from the *follow-up* measurements (MD, IM, and VH, respectively). Each stimulus consists of a pair of scales played one after the other. These six scale pairs are listed in **Table 1**. Participants responded by pressing a button whether they felt the two scales were played by the same pianist or different pianists.

The two scales in a pair were played preceded by two high-pitched notes (MIDI note 96), providing a tempo reference at 120 BPM. The scales were then played with four notes per metronome click, that is, at eight notes per second. The second scale always started 3.5 s after the first. All notes had a duration of 137.5 ms to generate legato style and a standardized loudness level. That is, we removed all loudness cues as well as articulation. Furthermore, each scale pair came in two versions: a *veridical* rendition, and a *magnified* rendition where all timing deviations were increased by a factor 5 (for a similar strategy in the context of a recognition experiment, see Hill and Pollick, 2000). In other words, we multiplied the irregularity vector by a scalar, making the differences more salient. The six stimuli (**Table 1**) were rendered twice (veridical and magnified), and presented in the two possible orderings, yielding 24 stimuli. Each of these were presented six times, yielding a total of 144 stimuli. The order was randomized for each participant and divided into 4 blocks of 36 trials.

Table 1 | Stimuli for the recognition experiment.

Pianist (*first*)	Pianist (*follow-up*)	Comparison	Fingerprint distance (ms)	SD-IKI *first* (ms)	SD-IKI *follow-up* (ms)
CA	CA	Self	3.34	7.78	5.08
CA	MD	Other	7.28	7.78	6.65
ES	ES	Self	3.38	8.85	9.35
ES	IM	Other	8.30	8.85	8.39
TY	TY	Self	3.37	7.29	7.42
TY	VH	Other	7.34	7.29	9.22

SD-IKI is the Standard Deviation of the inter-keystroke-intervals (in ms).

For data analysis, we used the R Package for Statistical Computing and the signal detection scripts developed by Prof. Abby Kaplan (http://home.utah.edu/∼u0703432/).

Irregularity threshold test

We extracted the irregularity traces of the right hand ascending scales for three pianists (CA, ES, and MD). The irregularity vector was multiplied by a scalar *factor* (between 0 and 5) and was then written as a MIDI file with eight notes per second, preceded by two metronome clicks at 120 BPM. For example, a factor of 0 means a perfectly regular (i.e., isochronous) scale, a factor of 1 corresponds to the scale as it was played in actuality, and a factor of 5 means that all note timings are five times more early or late than they were in reality whilst keeping the overall tempo intact. Participants were asked to report whether the scale sounded regular or irregular.

We used the maximum likelihood procedure (MLP) (Green, 1993; Gu and Green, 1994) to detect the threshold of the factor variable. Participants performed three thresholding blocks, one for each of the sample fingerprints. At the beginning of each block, we deployed 500 hypothetical psychometric curves with their midpoints linearly spaced over the factor levels from 0 to 5, crossed by the five false alarm rates of 0, 10, 20, 30, and 40%, yielding a total of 2,500 hypothetical psychometric curves maintained online in parallel. The slope parameter of these curves was set to four, since no prior experimental data exists and the slope has been shown not to influence the resulting thresholds all that much (Gu and Green, 1994). This yielded the following equation for the psychometric curves: $p(\text{yes}) = a + (1-a) \times (1/(1+\exp[-k \times (x-m)]))$, where x is the stimulus level (i.e., the factor), a is the false alarm rate, m is the mean of the psychometric curve, k is the slope parameter (4), $p(\text{yes})$ is the probability of responding "irregular."

Each block consisted of 36 trials. On each trial, we calculated online the likelihood of the set of previous participant responses for each of the 2,500 hypothetical psychometric curves. The curve with the maximum likelihood was chosen as the current estimate. The magnification factor for that given trial was determined by the 64%-response point of this current estimate psychometric curve. In this way, the algorithm is shown to converge rapidly to the participant's threshold (Green, 1993). We furthermore inserted two catch trials (with factor level 0 regardless of the current psychometric curve estimate) the first 12 trials at random locations, as well as four more over the remaining 24 trials.

Stimuli were written as MIDI files and then played through Timidity++ on a Windows computer, called by our Python (Pygame) graphical interface that registered the responses. The MLP computation was implemented in Python-MLP (which we have made available open-source online at: https://github.com/florisvanvugt/PythonMLP).

Participants

Ten pianists from the Hanover University of Music student pool participated in this perceptual experiment. Participants (four female) were 24.8 (SD 3.7) years old and studied piano as their primary instrument. Further, they had normal hearing and reported no neurological impairments. The experiment took approximately half an hour and participants received a nominal payment for their participation.

RESULTS

Recognition test

We used signal detection theory to calculate sensitivity (d') for the individual participants, fingerprint pairs, and the factors (veridical or magnified) separately. There was a main effect of factor [$F(1, 9) = 10.84$, $p = 0.001$, $\eta^2_G = 0.25$], reflecting that sensitivity was greater for magnified (mean $d' = 0.70$, SD $= 0.58$) than for veridical (mean $d' = -0.11$, SD $= 0.31$) pairs (**Figure 3A**). There was no main effect of fingerprint pair [$F(2, 18) = 1.44$, $p > 0.2$] but there was an interaction between factor and fingerprint pair [$F(2, 18) = 6.09$, $p < 0.01$, $\eta^2_G = 0.23$]. As a result, we investigated the sensitivity for each extract separately. For the veridical renditions, none of the sensitivities significantly exceeded zero [all $t(9) < 0.7$, $p > 0.25$], indicating that participants were not able to distinguish pairs of recordings from the same pianist from pairs from different pianists. However, for the magnified renditions of the CA-MD and ES-IM pairs, sensitivity was significantly above zero [$t(9) = 2.79$, $p = 0.01$, and $t(9) = 3.85$, $p < 0.01$, respectively]. Only for the magnified TY-VH pair participants' sensitivity was zero [$t(9) = 0.58$, $p = 0.29$].

After completing all blocks in this part of the experiment, participants were asked to subjectively rate the confidence in their answers on a five-point Likert scale from very confident (1) to very unsure (5). For the magnified fingerprint pairs, participants were mildly confident (median 3.5, range 2–4). For the veridical pairs, participants were similarly confident (median 4, range 3–5). The ratings did not differ significantly (Mann–Whitney $U = 10.5$, $p = 0.29$). We can conclude that although participants performed much better in the magnified pairs, they were not aware of this improvement in performance.

Irregularity threshold test

We discarded blocks in which participants' "irregular" response ratio for the catch trials exceeded 30%. This was the case for one block of one participant. The threshold for the remaining blocks was defined as the midpoint of the maximum likelihood estimate psychometric curve. Overall, curve midpoints expressed as factor were around or slightly above one (**Figure 3B**), meaning that the irregularities became audible only when they were slightly increased (factor >1). The thresholds were entered into a one way ANOVA with fingerprint (the three example fingerprints) as a factor. There was a main effect of fingerprint [$F(2, 26) = 4.85$, $p = 0.02$, $\eta^2_G = 0.27$], indicating that the threshold factors were different for the different extracts. However, the fingerprints differed in evenness at the outset (see **Table 1**). As a result, we expressed the threshold not as a factor but as the corresponding unevenness value (SD of the inter-keystroke-intervals). We then re-ran the ANOVA and found no main effect of extract [$F(2, 26) = 1.58$, $p = 0.22$]. The average threshold unevenness threshold value was 10.22 ms (SD 2.51).

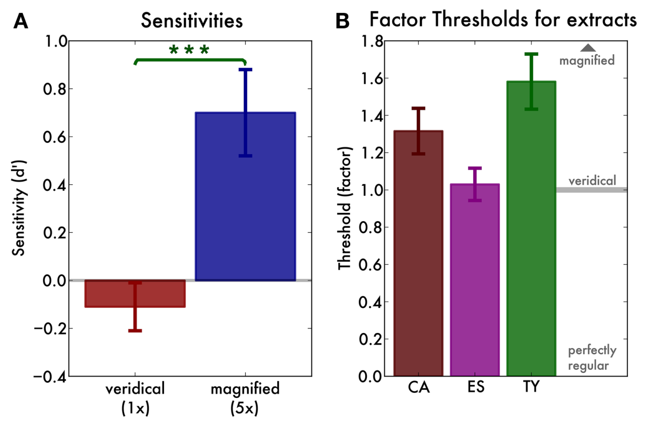


FIGURE 3 | (A) Main effect of factor (veridical or magnified) in the recognition experiment. Sensitivity (d') is not greater than zero for the veridical rendering (factor 1), but is greater than zero for the magnified (factor 5) rendering. The error bars indicate the standard error of the mean. **(B)** Irregularity thresholds for three representative fingerprints. We find that the thresholds for all three extracts are one or above, that is, their irregularity is heard only when we exaggerate it slightly. Error bars indicate the standard error of the mean.

DISCUSSION

From our threshold experiment, we can conclude that the thresholds straddle the boundary of the timings as actually played (i.e., slightly above factor 1). Our interpretation is that pianists train to make their scale playing more regular until the irregularities are no longer audible.

We conclude that participants are not able to tell the difference between a scale as played by a pianist and an isochronous scale. It naturally follows that they will then not be able to differentiate between pianists since both scales sound regular (isochronous) to them. Indeed, in our recognition test participants were unable to distinguish pairs of scales played by the same pianist from pairs played by different pianists. However, when we magnified the timing deviations by a factor of five, the participants performed above chance in the recognition task. This shows that, in principle, the task of distinguishing scale playing of one pianist from another can be done. These two tests, taken together, constitute evidence that participants were not able to hear the differences between the pianist fingerprints and categorize them on the basis of these differences.

Our study is also the first to systematically investigate thresholds for perception of irregularity in piano scales. We find that the irregularities in recorded piano scales are slightly below the perceptual threshold. This in itself is an interesting finding. Our interpretation is that pianists practice to make their scale playing sound regular but do not continue to make it more regular once it is below the perceptual threshold. For one, listeners will not be able to tell the difference, and secondly, if the motor learning of scale regularity is guided by auditory feedback (Jäncke, 2012) only, they will not be able to improve their temporal regularity once they fall below the auditory threshold.

We furthermore found that the differences in threshold between the extracts can be explained by their difference in unevenness: more temporally uneven fingerprints have a lower factor threshold, whereas more temporally even fingerprints have a higher threshold. This suggests that the obtained threshold of 10.22 ms is independent of the particular temporal fingerprint. We conclude that the unevenness captures the auditory percept of unevenness and no more complex auditory gestalt needs to be taken into account to explain the thresholds. The threshold corresponds to some 8.2% of the interval at this tempo, which is in line with the typical 10% threshold of a single late or early note in an otherwise isochronous sequence (Hyde and Peretz, 2004; Ehrlé and Samson, 2005).

Since these individual characteristics of the scale fingerprints are inaudible, it seems that their production is not dependent on auditory feedback. However, this conclusion is not warranted, since it could be that the timing deviations are residuals of expressive timing (Repp, 1999a). To clarify this issue, we investigated whether the pianistic fingerprints were affected by playing on a mute piano.

EXPERIMENT III

MATERIALS AND METHODS

Eighteen piano students (nine female) from the Hanover University of Music were invited to play two-octave C-major scales in two recordings. Participants were 28.2 (SD 5.8) years old. In the first recording, participants heard the sounds they produced (*sound*) but in the second recording the sound was switched off (*mute*). In both recordings, scales were played by one hand and then by the other. Otherwise, the procedure and analysis was identical to before. We report 95% confidence intervals (CI) unless otherwise stated.

RESULTS

We discarded incorrectly played scales leaving a total of 13.4 (SD 1.77) per condition. There was no effect of hand, direction, or recording on the number of correctly produced scales [all $F(1, 17) < 1.8$]. There was a significant but marginally small interaction between hand and direction [$F(1, 17) = 4.71$, $p = 0.04$, $\eta^2_G = 0.001$] and none of the other interactions was significant [all $F(1, 17) < 4.3$].

As before, the distances between fingerprints originating from the same pianist are smaller than those originating from different pianists [$F(1, 17) = 168.2$, $p < 0.001$, $\eta^2_G = 0.55$]. There was a (small) interaction between hand and direction [$F(1, 17) = 7.45$, $p = 0.01$, $\eta^2_G = 0.03$], indicating that for the right hand, inward scales are more similar than outward scales, whereas for the left hand this was the opposite.

Our distance-minimizing algorithm introduced in Experiment I correctly recognized between 8 (44%) and 12 (67%) of the 18 pianists using the fingerprint for only one hand and direction at a time. This exceeds chance performance, which lies at 6%. The correlation-maximizing algorithm correctly recognized between 7 (39%) and 15 (83%) pianists.

When we combined the two hands and two directions (yielding a 2 × 2 × 15 fingerprint matrix for each participant) and perform the same classification, the distance-minimizing algorithm correctly identified 15 out of 18 pianists [83%, binomial $p < 0.001$, confidence interval (0.59, 0.96)]. Crucially, the result is the same whether matching the *mute* fingerprints, one by one, to the set of *sound* fingerprints, or the other way around, indicating that there is no loss of information in the *mute* condition. The correlation-maximizing algorithm also recognizes 15 out of 18 pianists when it finds matching *sound* fingerprints to a given *mute* fingerprint, and the other way around spectacularly recognizes all 18 pianists [100%, binomial $p < 0.001$, CI (0.81, 1.00)].

In order to compare our results with those of Experiment I, we take 10,000 bootstrap samples of eight (unique) pianists and perform the classification with those. The correlation-maximizing algorithm recognizes 95% of pianists [SD 8%, bootstrap CI (75, 100)] whereas the distance-minimizing algorithm recognizes 90% of pianists [SD 8%, bootstrap CI (75, 100)]. That is, they do not perform significantly differently.

DISCUSSION

It is becoming clear that having auditory feedback while playing the scales is not of importance in the formation of the pianistic fingerprint. Indeed, it is a typical finding in performance literature that absence of auditory feedback only marginally affects performance (Repp, 1999b) or not at all (Gates and Bradshaw, 1974). The findings are furthermore in line with our previous result that fingerprints generated by the same movements are more similar than those generating the same sounds (Experiment I).

Finally, we turn to the question of how stable these fingerprints are over time.

EXPERIMENT IV

MATERIALS AND METHODS

We re-analyzed data published previously (Jabusch et al., 2009) in which 20 pianists' (eight female) scale playing was measured twice (first, follow-up) with an interval of 27.8 (SD 8.8) months. At the first measurement, pianists were 27.7 (SD 6.0) years old and had accumulated 21.6 (SD 11.0) thousand hours of lifetime piano practice (not counting one pianist who had not reliably reported this figure). In between the two measurement sessions, pianists accumulated an additional 2.8 (SD 1.8) thousand practice hours, amounting to an average 3.31 (SD 1.79) hours per calendar day (including weekends and holidays). All but two pianists were right-handed according to the Edinburgh handedness inventory (Laterality Quotient: $M = 73\%$, SD 56).

RESULTS

After discarding incorrect scales we were left with 13.5 (SD 0.8) scales of the *first* measurement and 12.8 (SD 1.2) scales at the *follow-up* measurement. This difference was significant [$F(1, 19) = 5.65$ $p = 0.03$ $\eta_G^2 = 0.09$].

As before, distance was smaller between recordings of the same pianist than that of different pianists [$F(1, 19) = 184.90, p < 0.001$, $\eta_G^2 = 0.30$]. Furthermore, distance was generally smaller between fingerprints of the right hand than those of the left hand [$F(1, 19) = 6.33$, $p = 0.02$, $\eta_G^2 = 0.05$], perhaps reflecting the greater training of the right hand (Kopiez et al., 2011). For brevity, we only report recognition results using the fingerprint combining both hands and directions. Recognition based on minimizing distance successfully found first recordings given the follow-up fingerprints in 13 pianists [65%, binomial $p < 0.001$, CI (40, 85)%]. Conversely, seven pianists were recognized based on their follow-up measurement [35%, binomial $p < 0.001$, CI (15, 59)%]. Recognition by maximizing correlation performed similarly with 13 (65%) and 8 [40%, binomial $p < 0.001$, CI (19, 64)%] correct identifications.

Bootstrap analysis was performed (see Experiment III) with 10,000 samples of eight pianists. Correlation recognition identified 73% [SD 15%, bootstrap CI (38, 100)%] of pianists and distance recognition 71% [SD 18%, bootstrap CI (38, 100)%]. Based on the bootstrap CI we can see that across the three experiments, identification was equally successful.

How is the stability of a pianist's fingerprint related to how much he or she practised between the two measurements? We calculated the distance for both hands and playing directions and correlated this to the number of practice hours accumulated between the two measurement points. The distances between the right hand outward scale fingerprints correlated negatively with amount of practice (Pearson $r = -0.71$, $p = 0.001$). That is, those who practised more showed smaller distances between their fingerprints. This does not mean that the fingerprints showed less deviations from regularity, but instead, that the deviations that were present were more consistently reproduced. The right hand inward fingerprints showed a tendency for the same correlation (Pearson $r = -0.46$, $p = 0.05$) but the left hand fingerprints did not ($r > -0.34$, $p > 0.16$).

DISCUSSION

The fingerprints that enabled reliable identification of pianists were sufficiently stable to still allow recognition after 27 months. **Figure 4** compares the distances across the Experiment I, III, and IV and **Figure 5** displays the recognition rates. Although it seems the recognition is worse in Experiment III and IV, the 95% bootstrap CI still include the 100% recognition rate of Experiment I. Therefore we conclude that recognition is not significantly different across the experiments.

GENERAL DISCUSSION

Artists are recognized reliably based on their work (Yamamura et al., 2009). The present study investigated pianist recognition based on non-expressive materials. Taking scale playing as an example, this study brings to light a highly individual temporal signature that enables robust identification of pianists using a simple algorithm. Clearly an individual timing signature is present physically, but perceptual recognition performance by musician listeners was at chance because the deviations were below their perceptual thresholds. Fingerprints appear to stem from neuromuscular factors in the pianists, rather than auditory feedback. This is confirmed in Experiment III that shows fingerprint formation is not affected by absence of sound. The fingerprint is furthermore robust, showing only mild changes in professional pianists over a 27-month interval.

The findings are in line with previous studies showing that pianists can be reliably recognized even when asked to not play expressively (Gingras et al., 2011). Our result strengthens the interpretation that recognition is based on non-expressive clues by employing materials (musical scales) with a clear auditory

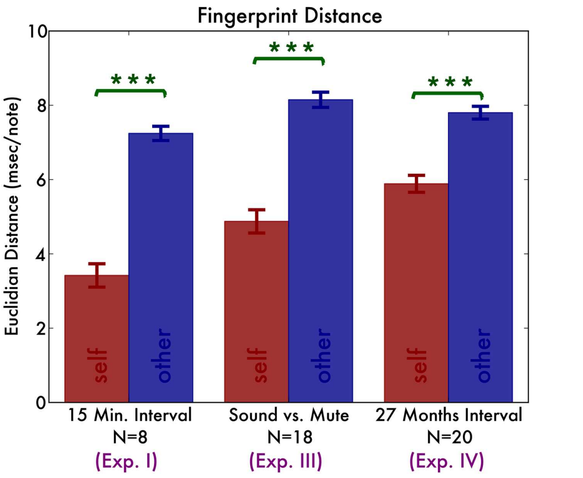


FIGURE 4 | Summary of the distances between fingerprints originating from the same pianist (*self;* the red bars) and fingerprints originating from different pianists (*other*; the blue bars).

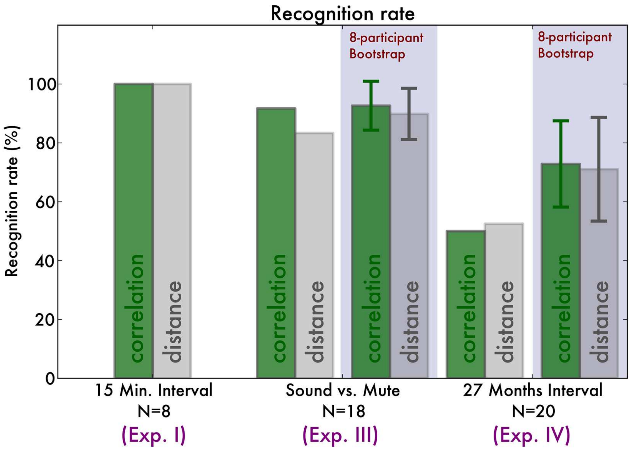


FIGURE 5 | Overview of the recognition rates of our recognition algorithm. The green bars indicate the correct classification rate by maximizing fingerprint correlation, and the gray bars by minimizing fingerprint distance. For comparison, we indicate the bootstrap classification results, indicating for each experiment the average recognition rates across eight-pianist bootstrap samples. Error bars indicate the standard deviation of the recognition rates.

goal of regularity. Moreover, we have at present only used timing information, discarding loudness and articulation markings that could potentially be used to enhance recognition. The recognition algorithm that we present pairs fingerprints with minimum distance or maximum correlation. The proposed similarity metric is transparent and easy to interpret (see **Figure 1**). As such, it is surprisingly simple compared with neural networks typically employed (Stamatatos and Widmer, 2005; Dalla Bella and Palmer, 2011).

The idea that artists can be recognized by a non-artistic feature of their work is not new. For example, painters can be automatically recognized by stroke style (Li et al., 2012). Beyond the realm of art, authorship can be established by relatively irrelevant features of produced work. For example, handwriting is highly individual (Rijntjes et al., 1999) and pattern recognition using word frequencies has been employed to establish Madison as the author of the 12 disputed Federalist papers (Mosteller and Wallace, 1964). Similarly, telegraph operators during the Second World War claimed to be able to identify the sender by the timing of his keystrokes ("Fist of sender"). The emerging field of keystroke dynamics puts this to use to authenticate computer users by their typing rhythm instead of through a password (Bergadano et al., 2002). Typically the problem remains that over time these dynamics change and recognition becomes impaired. In light of this, it is interesting that our recognition was highly stable even in a fairly homogeneous sample of expert pianists (Experiment IV). Recognition in keystroke dynamics as well as in our result may be based to some extent on the subunits that the produced sequences are divided into, i.e., its chunking (Sakai et al., 2003). On the other hand, more low-level neuromuscular properties such as the individual anatomy, especially tendon-ligament anatomy or the strengths of the individual muscles are more likely to be at the root of these individual temporal irregularities, since the sequences under consideration here (the scales) are greatly over-learned. Future studies may decide this issue by investigating recognition of pianists playing at various tempi, since although chunking may vary across speeds, the neuromuscular properties will remain constant.

We propose that studies investigating the individuality of artists, especially those employing machine learning strategies (Stamatatos and Widmer, 2005), may take into account that a large part of this individuality is inaudible and merely neuromuscular in nature. In the future, one could tease apart cues that are uniquely expressive and those that are neuromuscular.

Artistic individuality is typically thought to be deliberate and determined by top-down cognition. Our study opens the road to investigation into the tantalizing question of how biomechanical constraints may determine artistic performance in a bottom-up fashion.

ACKNOWLEDGMENTS

This work was supported by the EBRAMUS, European Brain and Music Ph.D. Grant to Floris Tijmen Van Vugt (ITN MC FP7, GA 238157). The authors are indebted to Laura Lübbe data collection in Experiment II, as well as Melinda Lee for running Experiment III. We also thank Meaghan Fowlie for proofreading the manuscript.

REFERENCES

Bakeman, R. (2005). Recommended effect size statistics for repeated measures designs. *Behav. Res. Methods* 37, 379–384.

Bergadano, F., Gunetti, D., and Picardi, C. (2002). User authentication through keystroke dynamics. *ACM Trans. Info. Syst. Secur.* 5, 367–397.

Dalla Bella, S., and Palmer, C. (2011). Rate effects on timing, key velocity, and finger kinematics in piano performance. *PLoS ONE* 6:e20518. doi:10.1371/journal.pone.0020518

Drake, C. (1993). Perceptual and performed accents in musical sequences. *Bull. Psychon. Soc.* 31, 107–110.

Ehrlé, N., and Samson, S. (2005). Auditory discrimination of anisochrony: influence of the tempo and musical backgrounds of listeners. *Brain Cogn.* 58, 133–147.

Engel, K. C., Flanders, M., and Soechting, J. F. (1997). Anticipatory and sequential motor control in piano playing. *Exp. Brain Res.* 113, 189–199.

Flach, R., Knoblich, G., and Prinz, W. (2004). Recognizing one's own clapping: the role of temporal cues. *Psychol. Res.* 69, 147–156.

Gates, A., and Bradshaw, J. L. (1974). Effects of auditory feedback on a musical performance task. *Percept. Psychophys.* 16, 105–109.

Gingras, B., Lagrandeur-Ponce, T., Giordano, B. L., and McAdams, S. (2011). Perceiving musical individuality: performer identification is dependent on performer expertise and expressiveness, but not on listener expertise. *Perception* 40, 1206–1220.

Green, D. M. (1993). A maximum-likelihood method for estimating thresholds in a yes–no task. *J. Acoust. Soc. Am.* 93, 2096.

Gu, X., and Green, D. M. (1994). Further studies of a maximum-likelihood yes–no procedure. *J. Acoust. Soc. Am.* 96, 93.

Hill, H., and Pollick, F. E. (2000). Exaggerating temporal differences enhances recognition of individuals from point light displays. *Psychol. Sci.* 11, 223–228.

Hyde, K. L., and Peretz, I. (2004). Brains that are out of tune but in time. *Psychol. Sci.* 15, 356–360.

Jabusch, H.-C., Alpers, H., Kopiez, R., Vauth, H., and Altenmüller, E. (2009). The influence of practice on the development of motor skills in pianists: a longitudinal study in a selected motor task. *Hum. Mov. Sci.* 28, 74–84.

Jabusch, H.-C., Vauth, H., and Altenmüller, E. (2004). Quantification of focal dystonia in pianists using scale analysis. *Mov. Disord.* 19, 171–180.

Jäncke, L. (2012). The dynamic audio-motor system in pianists. *Ann. N. Y. Acad. Sci.* 1252, 246–252.

Jeannerod, M. (2003). The mechanism of self-recognition in humans. *Behav. Brain Res.* 142, 1–15.

Kopiez, R., Jabusch, H.-C., Galley, N., Homann, J.-C., Lehmann, A. C., and Altenmuller, E. (2011). No disadvantage for left-handed musicians: the relationship between handedness, perceived constraints and performance-related skills in string players and pianists. *Psychol. Music* 40, 357–384.

Li, J., Yao, L., Hendriks, E., and Wang, J. Z. (2012). Rhythmic brushstrokes distinguish van Gogh from his contemporaries: findings via automated brushstroke extraction. *IEEE Trans. Pattern Anal. Mach. Intell.* 34, 1159–1176.

Loula, F., Prasad, S., Harber, K., and Shiffrar, M. (2005). Recognizing people from their movement. *J. Exp. Psychol. Hum. Percept. Perform.* 31, 210–220.

MacKenzie, C. L., and Van Eerd, D. L. (1990). Rhythmic precision in the performance of piano scales: motor psychophysics and motor programming. *Atten. Perform.* 13, 375–408.

Mosteller, F., and Wallace, D. L. (1964). *Inference and Disputed Authorship: The Federalist*. Reading: Addison-Wesley.

Murgia, M., Hohmann, T., Galmonte, A., Raab, M., and Agostini, T. (2012). Recognising one's own motor actions through sound: the role of temporal factors. *Perception* 41, 976–987.

Prasad, S., and Shiffrar, M. (2009). Viewpoint and the recognition of people from their movements. *J. Exp. Psychol. Hum. Percept. Perform.* 35, 39–49.

Repp, B. H. (1999a). Control of expressive and metronomic timing in pianists. *J. Mot. Behav.* 31, 145–164.

Repp, B. H. (1999b). Effects of auditory feedback deprivation on expressive piano performance. *Music Percept.* 16, 409–438.

Repp, B. H., and Knoblich, G. (2004). Perceiving action identity: how pianists recognize their own performances. *Psychol. Sci.* 15, 604–609.

Rijntjes, M., Dettmers, C., Büchel, C., Kiebel, S., Frackowiak, R. S. J., and Weiller, C. (1999). A blueprint for movement: functional and anatomical representations in the human motor system. *J. Neurosci.* 19, 8043–8048.

Sakai, K., Kitaguchi, K., and Hikosaka, O. (2003). Chunking during human visuomotor sequence learning. *Exp. Brain Res.* 152, 229–242.

Sevdalis, V., and Keller, P. E. (2011). Perceiving performer identity and intended expression intensity in point-light displays of dance. *Psychol. Res.* 75, 423–434.

Stamatatos, E., and Widmer, G. (2005). Automatic identification of music performers with learning ensembles. *Artif. Intell.* 165, 37–56.

van Vugt, F. T., Jabusch, H.-C., and Altenmüller, E. (2012). Fingers phrase music differently: trial-to-trial variability in piano scale playing and auditory perception reveal motor chunking. *Front. Psychol.* 3:495. doi:10.3389/fpsyg.2012.00495

Wagner, C. (1971). "The influence of the tempo of playing on the rhythmic structure studied at pianist's playing scales," in *Medicine and Sport Biomechanics II*, eds J. Vredenbregt and J. Wartenweiler (Basel: Karger), 129–132.

Yamamura, H., Sawahata, Y., Yamamoto, M., and Kamitani, Y. (2009). Neural art appraisal of painter: Dali or Picasso? *Neuroreport* 20, 1630–1633.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 24 January 2013; paper pending published: 12 February 2013; accepted: 04 March 2013; published online: 21 March 2013.

Citation: Van Vugt FT, Jabusch H-C and Altenmüller E (2013) Individuality that is unheard of: systematic temporal deviations in scale playing leave an inaudible pianistic fingerprint. Front. Psychol. ***4****:134. doi: 10.3389/fpsyg.2013.00134*

This article was submitted to Frontiers in Cognitive Science, a specialty of Frontiers in Psychology.

frontiers in
PSYCHOLOGY

ORIGINAL RESEARCH ARTICLE
published: 04 March 2014
doi: 10.3389/fpsyg.2014.00157

Investigating pianists' individuality in the performance of five timbral nuances through patterns of articulation, touch, dynamics, and pedaling

Michel Bernays[1]* and Caroline Traube[2]

[1] IDMIL/SPCL, Schulich School of Music, McGill University, Montreal, QC, Canada
[2] LRGM, OICRM, Faculté de musique, Université de Montréal, Montreal, QC, Canada

Edited by:
Bruno Gingras, University of Vienna, Austria

Reviewed by:
Werner Goebl, University of Music and Performing Arts Vienna, Austria
Steven R. Livingstone, Ryerson University, Canada

***Correspondence:**
Michel Bernays, IDMIL/SPCL, Schulich School of Music, McGill University, 555 Sherbrooke Street West, Montreal, QC H3A 1E3, Canada
e-mail: michel.bernays@mail.mcgill.ca

Timbre is an essential expressive feature in piano performance. Concert pianists use a vast palette of timbral nuances to color their performances at the microstructural level. Although timbre is generally envisioned in the pianistic community as an abstract concept carried through an imaged vocabulary, performers may share some common strategies of timbral expression in piano performance. Yet there may remain further leeway for idiosyncratic processes in the production of piano timbre nuances. In this study, we examined the patterns of timbral expression in performances by four expert pianists. Each pianist performed four short pieces, each with five different timbral intentions (bright, dark, dry, round, and velvety). The performances were recorded with the high-accuracy Bösendorfer CEUS system. Fine-grained performance features of dynamics, touch, articulation and pedaling were extracted. Reduced PCA performance spaces and descriptive performance portraits confirmed that pianists exhibited unique, specific profiles for different timbral intentions, derived from underlying traits of general individuality, while sharing some broad commonalities of dynamics and articulation for each timbral intention. These results confirm that pianists' abstract notions of timbre correspond to reliable patterns of performance technique. Furthermore, these effects suggest that pianists can express individual styles while complying with specific timbral intentions.

Keywords: piano, performance, timbre, individuality, expression, articulation, touch, Bösendorfer CEUS

1. INTRODUCTION

Musical performance holds a crucial role in the art and experience of music. Classical performers in particular can shine their own light upon the composed work and express their creativity. Accordingly, an extensive, empirical body of knowledge has been developed amongst musicians with respect to the art and technique of performance. Notably, it has been so for the piano in the few centuries since the instruments' inception. Guidelines about technique, gesture, touch, and mental approach were provided by teachers and pedagogues in the aim of helping pianists develop their own "sound" and musical expression (Hofmann, 1920; Neuhaus, 1973; Fink, 1992). Individual pianist development is then shaped and oriented by the teacher and the piano school (e.g., Russian, German or French) he/she abides by Lourenço (2010).

A large body of research has been devoted to exploring expressive piano performance, by examining the general performance parameters (in opposition to the musical parameters of pitch, harmony or rhythm Rink et al., 2011) that pianists can use as expressive devices. Expressive control parameters of timing and amplitude have been the most explored by far, for their salience among the performance parameters that pianists can vary and their effect on the perception of emotional expression (Bhatara et al., 2011), and for the relative accessibility of their measurement (e.g., with MIDI digital recording pianos or with acoustical analysis Goebl et al., 2008). They were revealed to follow broad, common expressive strategies that depend on other musical factors. In particular, expressive deviations from the score in timing and dynamics were shown to follow common patterns related to the musical structure (phrasing and local accents) (Repp, 1992; Shaffer, 1995; Parncutt, 2003). Moreover, overlap durations in *legato* articulation were shown to depend on register, tempo, interval size and consonance, and position in an arpeggio (Repp, 1996b), while the melody lead effect (i.e., the melody note in a chord played slightly earlier) was shown as an artifact of the dynamic accentuation of the melodic voice, thus correlated to amplitude (Goebl, 2001). Different computational models of expressive performance were developed (Widmer and Goebl, 2004), based on structurally guided rules of expressive timing and dynamics (MIDI parameters). These rules could be defined in different ways (De Poli, 2004): explicitly (as heuristics), through analysis-by-measurement (Todd, 1992, e.g.,) or analysis-by-synthesis (Friberg et al., 2006); or implicitly, by machine learning, from an input learning set of recorded human performances (Widmer et al., 2009).

However, within the constraints imposed by these structural rules of timing and dynamics, there remains enough space for pianists to bring out their individual expression in performing a piece, as advocated in piano pedagogy and performance. Indeed, idiosyncratic patterns, that would differ between pianists yet remain consistent for each one, were identified below general rules, in the following expressive techniques: in chord asynchronies and melody lead (Palmer, 1996); in temporal deviations around the frame defined by the musical structure (so long as these deviations remain within an "acceptable" range) (Repp, 1990, 1992, 1998); in articulation, for which the general, common trends in overlap durations were obscured by considerable inter-individual variations (Bresin and Battel, 2000); and also in sustain pedal timing (Heinlein, 1929; Repp, 1996c, 1997). On the other hand, in Repp's (1996a) study of 30 performances by 10 graduate students of Schumann's "Träumerei," the expressive dynamics (MIDI velocities) did not appear as a clear bearer of individual differences, yet tended to be more consistent in the repeated performances by each pianist than between performers.

Expressive features of both timing and loudness were used successfully for automatic identification, with machine learning models, of the performer in MIDI (Stamatatos and Widmer, 2005) or audio (Saunders et al., 2008; Wang, 2013) recordings of piano performances. Meanwhile, temporal deviations were identified as individual fingerprints in non-expressive scale playing, yet could not be perceived by human listeners (Van Vugt et al., 2013), indicating that a fine-grained level of individuality in piano performance resides below the level of expressive timing.

More generally, in a review paper, Sloboda (2000) examined the performers' abilities to generate expressively different performances of the same piece of music according to the nature of intended structural and emotional communication, and described how some of these abilities have been shown to have lawful relationships to objective musical and extra-musical parameters. With respect to other keyboard instruments, it was also shown that local tempo variations, onset asynchronies, and especially articulation (overlaps) were highly individual parameters of expressive performance in Baroque organ music (Gingras et al., 2011). How these findings relate to piano performance, however, is a non-trivial issue.

Individuality in expressive piano performance was also examined in light of musical gestures, i.e., timing and dynamic patterns, whose occurrences, distribution and diversity could characterize the individual expressive strategies in 29 case-study performances of Chopin's Mazurka, op.24 no.2 (Rink et al., 2011). Conversely, literal pianists' gestures (body motion, finger movements) in expressive piano performance were shown as highly idiosyncratic, although related to the musical and rhythmic structure of the piece (MacRitchie, 2011; MacRitchie et al., 2013). Likewise, finger kinematics were shown as idiosyncratic enough for performers to be accurately identified, with neural network classifiers, from their finger movements and accelerations during attacks and key presses (Dalla Bella and Palmer, 2011).

However, such pianistic gestures hold other functions (ancillary, figurative) than the actual, effective sound production (Cadoz and Wanderley, 2000). Consequently, their idiosyncratic nature does not necessarily translate into an idiosyncratic expressive sound production—the main concern of this article. Thus, this study only considers the effective gestures applied by pianists to the keyboard and pedals.

Among the expressive musical attributes available to pianists other than sheer timing and loudness, timbre holds a crucial expressive role (Holmes, 2012), which has been widely acknowledged within the pianistic community (Bernays, 2012). Usually envisioned as the inherent characteristic of a sound source or instrument, timbre is also considered by pianists as the subtle quality of sound that they can control through the expressive nuances of their performances. However, it has long been debated whether pianists can actually control timbre as a sound quality of performance. Scientific studies concluded long ago that controlling piano timbre on a single key was limited by the mechanical constraints of the action to sheer keystroke velocity, and thus inseparable from intensity (Hart et al., 1934). The influence of contact noises however (especially finger-key contact) on the timbre of a single piano tone was demonstrated (Goebl and Fujinaga, 2008), suggesting that the type of touch can bear an influence on the timbre of a single tone (Goebl et al., 2005). Yet even so, keyboard control over the timbre of a single tone remains quite limited.

But in a polyphonic, musical context, the expressive performance features of articulation, touch and pedaling can govern subtle tone combinations, in the timing and dynamic balance (*polyphonic touch*) of notes in a chord and in melodic lines (Parncutt and Troup, 2002). Composite timbres thus arise which are, in essence, performer-controlled (Bernays, 2013). Pianists' expressive intentions can thus be conveyed through specific timbral nuances (Sándor, 1995), to the vast palette of which an extensive vocabulary, including numerous adjectival descriptors, has been associated (Bellemare and Traube, 2005; Cheminée, 2006). However, the precise technique and ways of production of piano timbre nuances have generally been subdued to abstraction, mental conception, imitation and aural modeling (Woody, 1999) in piano pedagogy and treatises (Kochevitsky, 1967; Neuhaus, 1973). This abstract approach to teaching the production of piano timbre, in combination with the focus on personal expression, thus suggest that pianists may employ individual, idiosyncratic expressive performance strategies toward producing a specific timbral nuance—whose understanding according to its verbal descriptor may also vary slightly between pianists.

In order to explore pianists' individuality in the production of timbral nuances, piano performance has to be measured and quantified with the high precision required for identifying the subtleties of expressive performance employed in controlling timbre. In the absence or rarity of high-precision measurement tools for piano performance, the intricacies of timbre production have essentially remained out of the reach and/or concern of piano performance studies—with the exception of Ortmann's (1929) investigation, with the help of cumbersome mechanical apparatus, of the relations between piano touch and timbre on a single tone. Going further, with the high-accuracy Bösendorfer CEUS digital piano performance-recording system, this study explores pianists' individuality in the production of piano timbre in a polyphonic, ecologically valid musical context.

Furthermore, the verbalization of piano timbre was studied quantitatively (Bernays and Traube, 2011), according to judgements of semantic similarity between the 14 descriptors of piano timbre most cited by pianists in Bellemare and Traube (2005).

These evaluations were mapped into a semantic space, whose first two, most salient dimensions formed a plan in which descriptors were grouped in five distinct clusters—which was confirmed by hierarchical cluster analysis. In each cluster, the descriptor judged the most familiar was selected. The five most familiar, diverse and representative timbre descriptors thus highlighted—**dry, bright, round, velvety**, and **dark**—appear (in that order) along a circular arc in the semantic plan. These five descriptors defined the timbres for which to seek out individual patterns of production between several pianists.

This method of searching for production patterns in performances driven by verbal descriptors has been employed in studies of emotions in music performance. Verbal descriptors of emotion in music were first categorized by Hevner (1936), in eight groups arranged by similarity in a circular pattern. Studies of emotional expression in music performance have used a subset of emotional descriptors—taken either from Hevner's categories or from the verbal descriptors of basic emotions in a general context (Ekman, 1992)—in order to drive the performers' intentions. Emotional descriptors among "happiness," "anger," "sadness," "tenderness," "fear," "solemnity," as well as emotionally "neutral" (for comparison), were for instance used as emotional instructions to several performers playing various instruments in Gabrielsson and Juslin (1996), Juslin (1997), Juslin and Laukka (2003) (a meta-study), and Quinto et al. (2013). Among other goals, correlations were identified between these emotional expressions and performance parameters of dynamics, tempo, timing, articulation, as well as acoustical features of tone and timbre. In more details, both intensity and tempo were positively correlated with arousal (sadness, tenderness vs. anger, happiness), and high variability in intensity was associated with fear. *Legato* articulation was found to reflect the expression of tenderness or sadness, while a *staccato* articulation expressed happiness. Individual differences between performers in encoding each emotion were also mentioned, yet hardly detailed. In particular, Gabrielsson and Juslin (1996) concluded that the performance rules for communicating emotions depend on the instrument, the musical style, and the performer, as well as on the listener. Meanwhile, Canazza et al. (1997) used sensorial-type adjectives (bright, dark, hard, soft, heavy, light, and normal) as instructions for expressive clarinet performances of the same musical piece, and identified sonological characteristics of these emotions.

Regarding the piano, Madison (2000) explored the expression of happy, sad, angry, and fearful emotions in performances by three pianists, and identified correlations between emotional expression and the degree of variability in patterns of timing, articulation and loudness.

Furthermore, Ben Asher (2013) developed and trained a machine learning algorithm to automatically retrieve in real time the musical expression and emotional intentions in piano performance from the gestures of pianists. The training set used verbal descriptors of basic emotions to drive the performances, and the pianists' high-level gestures were automatically identified from kinaesthetic data.

However, we may argue that the verbal descriptors of basic emotions used in such studies are not comparable, in the context of music performance and pedagogy, with the verbal descriptors of piano timbre nuances we are using in this article. Indeed, the vocabulary that is used by musicians and music teachers to verbally define and communicate the emotional qualities of music in performance is based on more complex or indirect metaphors and analogies, most often attempting to connect emotions and music performance through their shared motor-affective elements and motional aspects, for instance with reference to bodily gesture or vocal intonations (Woody, 2002). Those may be more effective in triggering appropriate actions from the performer, whereas the vocabulary of basic emotions may be comparatively better suited to music perception. On the other hand, the vocabulary describing musical timbre was demonstrated as consensual and meaningful (Faure, 2000). In particular, the vocabulary of verbal description of piano timbre forms a specialized lexicon that holds a distinctive meaning within the context of piano performance (Cheminée, 2006), and was shown as familiar to pianists in a musical context, and largely shared among them (Bellemare and Traube, 2005).

The first research question explored in this study is whether individuality in piano performance manifests itself in the gestures applied by pianists on the keyboard, and if so, which descriptive performance features of dynamics, touch, articulation, and pedaling determined from high-resolution key/pedal position and hammer velocity tracking can reveal individual piano performance patterns. Yet the main research question that we wish to investigate is whether individuality in piano performance, if highlighted by specific performance features, arises in different patterns between performances of different timbral nuances, i.e., whether the characteristics of piano performance that would prove idiosyncratic in comparing different pianists vary depending on the timbral nuance performed.

2. MATERIALS AND METHODS

In order to explore pianists' individuality in the expressive production of piano timbre nuances in a musically relevant framework that could mirror a genuine musical experience, the study was designed with respect to the following steps: selection of the five verbal descriptors **dry, bright, round, velvety** and **dark** as timbral instructions for which to explore performance idiosyncrasies; conception of musical pieces to be expressively performed according to these different timbral nuances; use of non-invasive, high-accuracy piano performance-recording equipment; recording of timbre-colored performances; and extraction therein of meaningful piano performance, articulation, touch and pedaling descriptors.

2.1. MUSICAL PIECES

In order to set a musical context adequate to expressive timbre production in performances, four short solo piano pieces were selected, among 15 specially composed for the study following instructions on the timbral nuances to be expressed (cf. **Figure 1**). Each selected piece could allow for a meaningful, consistent-throughout expression of each of the five timbral nuances, and featured many aspects of piano technique that we wanted to explore. Each just a few bars long (from 4 to 7, with different meters), their duration at score tempo ranged between 12 and 15 s.

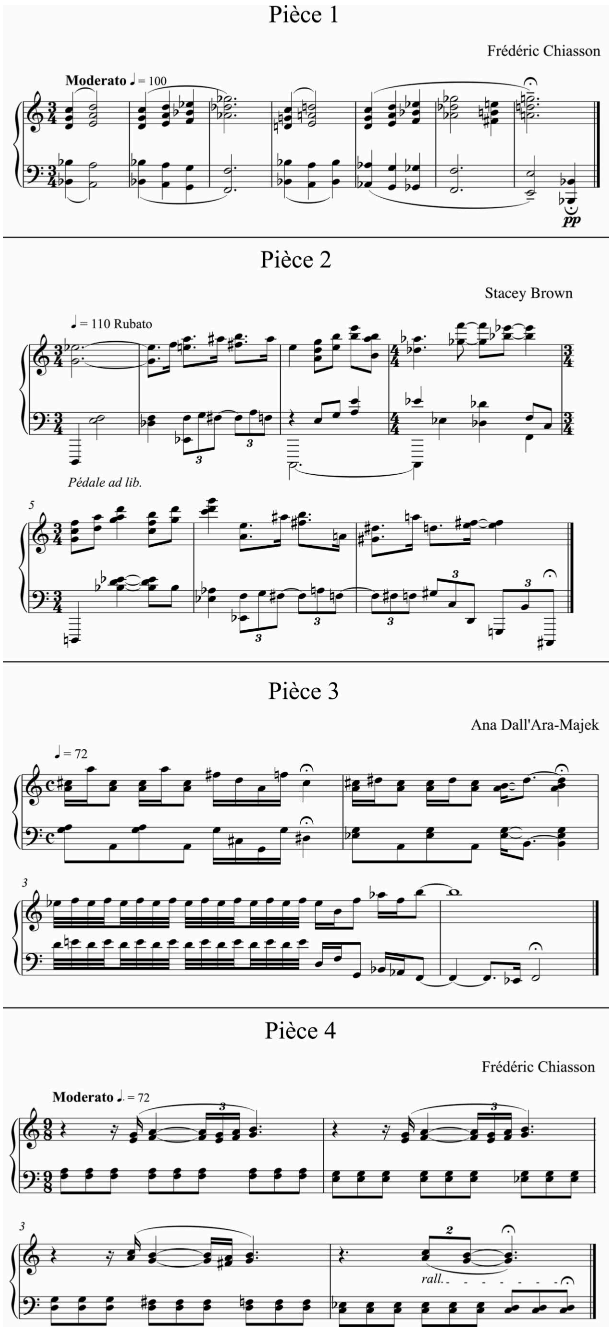


FIGURE 1 | Scores of the four pieces composed and selected for the study.

2.2. EQUIPMENT

To investigate the fine-grained nuances of pianists' performance control and touch that let them express different timbral nuances and could reveal idiosyncratic approaches, highly precise data were required from which to thoroughly assess the intricacies of key strokes. In this aim, we had the opportunity to use the Bösendorfer CEUS piano digital recording and reproducing system. Vastly improving upon the MIDI-based SE reproducing piano, the CEUS system is designed to both record with high accuracy the actions of a pianist on the keyboard, and to reproduce the performance faithfully, with solenoids that activate each key and pedal to mirror the original, recorded performance. In this study, the CEUS system was only used for its recording abilities. Equipped with optical sensors behind the keys, hammers and pedals, microprocessors, electronic boards, and a computer system, it can indeed track key and pedal positions and hammer velocities at high resolution (8-bit) and high sampling rate (500 Hz). The system we used was embedded in the Imperial Bösendorfer Model 290 grand piano installed at BRAMS (International Laboratory for Brain, Music and Sound Research, Montreal, Canada) in a dedicated recording studio.

2.3. PARTICIPANTS

Four pianists participated in the study. They are further referred as pianists A, B, C and D. All four had extensive professional experience and advanced-level piano performance diploma. Pianist A is a 30-year old French female. She studied piano and played professionally in France and Belgium. Pianist B is an 54-year old Italian male. He studied piano in Switzerland and Italy, and played professionally in several countries. Pianist C is a 46 year-old French-Canadian male. He studied piano and played professionally in the Quebec province, Canada. Pianist D is a 22-year old French male. He studied piano and played professionally in France and Canada.

2.4. PROCEDURE

Each participant had received in advance the scores of the pieces and the timbral instructions, and was given time to practice. Rehearsal sessions were allotted on the Bösendorfer piano, to allow for familiarization with the instrument and the room. The timbral instructions were provided with only the five adjectival descriptors. The participants confirmed their familiarity with each descriptor as a piano timbre nuance. They were asked to perform each of the four pieces, with each of the five timbres. Three such runs of 20 performances were conducted successively—twice in an order of pieces and timbres chosen by the participant, and once in randomized forced order—so as to get three performances for each condition (piece × timbre). The participants were allowed to immediately replay a performance if they considered the previous try unsatisfactory. Each of the 60 performances per participant was recorded through the CEUS system. We thus collected 240 CEUS `boe`-format recordings of 4 pianists performing 4 pieces with 5 different timbres, 3 times each.

2.5. PERFORMANCE ANALYSIS AND EXTRACTION OF PERFORMANCE FEATURES

In order to extract meaningful piano performance and touch features from CEUS-acquired data, the PianoTouch `Matlab` toolbox was specifically developed (Bernays and Traube, 2012). From the high-frequency, high-resolution key/pedal positions and hammer velocities, note and chord structures were retrieved, and an exhaustive set of quantified features spanning several broad areas of piano performance and touch were computed:

- Dynamics and attack: maximum hammer velocity (MHV), maximum key depression depth (Amax) and their relations

(ratio of their values, respective timing); attack durations (related to instants of both MHV and Amax), speeds (ratio of MHV or Amax to duration) and percussiveness (pressed vs. struck touch) (Goebl et al., 2005; McPherson and Kim, 2011)
- Articulation: sustain and release durations, synchrony of notes within chords (melody lead; duration, rate and amount of asynchrony at onset and offset), intervals and overlaps between chords (inter-onset and offset-to-onset, overlap durations, number and corresponding amount of depression, with respect to all chords, same-hand chords and other-hand chords)
- Detailed use of the soft and sustain pedals during each chord: duration of use, of full depression and of part-pedaling, timing with regard to chord onsets and offsets, depression (average, maximum, depth at chord onset, offset and MHV).

Each note was thus described by 46 features. Averages and standard deviations of these note features per chord were calculated as chord features. With the addition of 76 chord-specific features, each chord was described by 168 features.

For each performance, averages and standard deviations of chord features were calculated over all the chords in the performance. Only the chord features whose averaging could provide a meaningful description of a performance were conserved. For instance, the absolute instants of chord onsets, while useful in describing a chord, are meaningless in and of itself when averaged over a performance. However, such chord features were indispensable as building blocks for calculating other chord features (e.g., synchrony or attack speed) that can be meaningfully averaged as performance features and compared between different performances. Performance features were thus given by the averages and standard deviations per performance calculated for 100 relevant chord features (out of 168). With the addition of the number of chords and total number of notes per performance, each performance was thus described by 202 performance features.

Moreover, performances features were also calculated over only the chords played with the left hand and over only the chords played with the right hand, and differences in average performance feature values between hands were determined. The 32 pedaling features per chord were not considered in this context. Excluding their averages and standard deviations per performance (64 performance features), there remained 138 performance features to describe each performance with regard to left-hand chords only, to right-hand chords only, and to the differences between hands.

In total, over the four different chord groupings per performance (all chords, left-hand chords only, right-hand chords only, and differences between hands), $202 + 138 \times 3 = 616$ performance features were calculated to characterize each of the 240 recorded performances.

3. RESULTS

3.1. PIANISTS' OVERALL INDIVIDUALITY

First, in order to obtain an overall picture of each pianist's individual and idiosyncratic patterns of articulation, touch, dynamics and pedaling in all the performances whose key and pedal depressions and hammer velocities were recorded, the performance features that could prove characteristic of one pianist's performances in contrast with the others' were sought out over the whole dataset of 240 performances of four different pieces by four pianists highlighting five different timbral nuances.

3.1.1. Statistical method

Statistical analysis of variance was conducted, with regard to the 616 performance features describing each performance, over the 240-performance dataset. Three-Way repeated-measures ANOVAs were performed for each performance feature (as dependent variable), with the performer as random factor, and timbre, piece, and repetition (of the same experimental condition) as fixed-effect factors. Two factors were considered for assessing pianists' overall individuality in performance: the random factor of performer, and the fixed-effect repetitions. Indeed, for a performance feature to reveal pianists' individuality, the effect of the performer has to be significant (i.e., rejecting with at least 95% confidence the null hypothesis of equal variance between pianists), but the feature must also remain consistent between repetitions (i.e., no significant differences between repetitions at the 5% level).

For these two factors, the assumptions required by the ANOVA were tested. For the between-subject factor, assumptions of normal distributions (Kolgomorov–Smirnov test) and homoscedasticity (Levene's test) were tested. In the cases where the ANOVA was significant at the 5% level in rejecting the null hypothesis of equal variance between performers but the assumptions were not met, the non-parametric, one-way Kruskal–Wallis rank analysis of variance was also run, to control for possible type I errors (i.e., to confirm or invalidate significance, depending of the significance at the 5% level of the Kruskal–Wallis test). Effect sizes and statistical power were also calculated. For the repetition, within-subject factor, the Huynh–Feldt correction was applied to the degrees of freedom in order to compensate for possible violations of sphericity (as assessed with Mauchly's test).

With this method, 159 performance features were revealed as both consistent (i.e., not significantly different at the 5% level) between repetitions, and different between the four pianists (i.e., significant at the 5% level in rejecting the null hypothesis of equal variance between them).

3.1.2. Reduced performance space

Principal Component Analysis was applied over the dataset of 240 performances to those 159 selected performance features. Indeed, those are the performance features that were shown to highlight consistent, significant difference between the four performers (within the limits of statistical accuracy, as will be discussed in Section 4); the excluded performance features would only impede the visualization of differences between performers. The values of the selected performance features were normalized (Z-scores per feature over the 240 performances), in order to assess their relative values, so as to enable comparisons between features on an equivalent scale, and so that each pianist's individuality according to each feature could be expressed as a relative deviation from the overall average (Wöllner, 2013). The PCA procedure aimed at defining reduced performance spaces whose dimensions would correspond to the first few principal components and that would illustrate the aspects and features of performances most relevant in highlighting the performers' individuality. The number of principal components required to explain a sufficient part of the variance in the input dataset determined the number of dimensions of each reduced performance space, and the meaning of each dimension in terms of the performance features it represents most saliently was explored in its loadings (i.e., the weights associated to each feature in the linear combination which forms the principal component).

The first three principal components of the PCA applied to the 159 selected performance features over the 240-performance dataset account for 61.4% (31.97, 15.56, and 13.87% respectively) of the input variance. Varimax factor rotation (Kaiser, 1958) was then applied to these first three PCA loadings, in order to optimize the weightings of performance features between the loadings. The average positions (over the three repetitions) of same-condition (pianist, piece and timbre) performances, according to their coordinates in these three rotated dimensions, are presented in **Figure 2**.

According to the varimax-rotated loadings, the first dimension corresponds to performance features of attack and dynamics: hammer velocity and its variations between chords, attack speed, duration, and percussiveness, and key depression depth. For the second dimension, the loadings are principally attributed to performance features of chord, note, sustain, and overlap durations, as well as right-hand note offset timing in chords, and inter-onset intervals. Finally, the third dimension essentially accounts for performance features of articulation: number of overlapping chords, *legato* vs. *staccato* articulation, interval between chords, and melody lead.

In the performance space thus defined, some idiosyncratic, consistent performance tendencies of the four pianists are revealed. Pianist A shows the lowest dynamics and longest/slowest attacks, and the most *legato* articulation. On the other hand, Pianist C uses the highest dynamics and fastest/shortest attacks, while employing rather short notes and detached articulation. As for Pianist B, he tends to employ short notes with the most *staccato* articulation. Finally, Pianist D tends to let keys depressed the longest, with a *legato* articulation.

3.1.3. *Descriptive performance portrait*

Now, in order to obtain a more precise account of the four performers' idiosyncrasies as reflected by specific performance features, the most salient, relevant, meaningful and

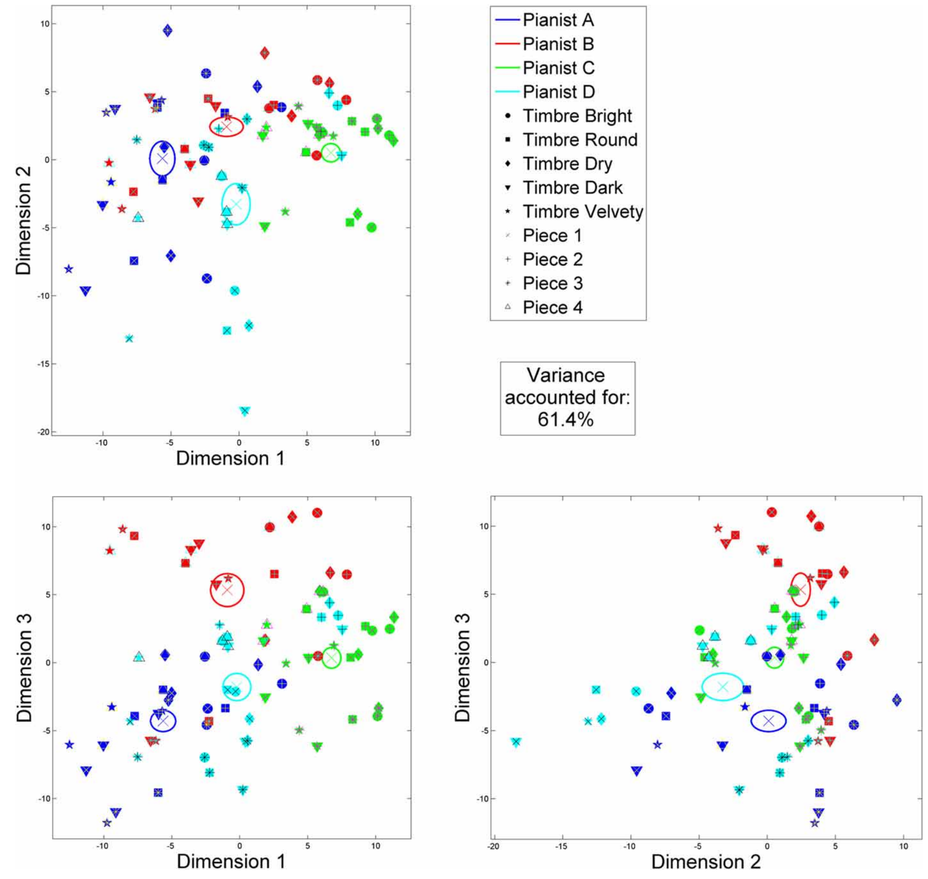


FIGURE 2 | Reduced 3-dimension performance space by PCA and varimax factor rotation applied to 159 significant performance features over the 240-performance dataset: planar projections. For a clearer representation, only the averages of the three repetitions of same-condition performances are plotted. Averages per pianist are indicated by colored crosses, and ±1 SE with ellipses.

non-redundant features among those significant between performers (and consistent between performers' repetitions) were sought out. For this aim, the significant performance features were first divided into the four broad, technically independent categories of: (1) dynamics and attack, (2) articulation, (3) soft pedal, and (4) sustain pedal. Correlations were calculated between all the significant performance features pertaining to the same category, and hierarchical clustering was used to regroup the most correlated (thus redundant) features into a same cluster. For each cluster identified in each category, the most significant, meaningful and interpretable performance features could then be selected, while minimizing the loss of information from the other discarded features. This process allowed for selecting a minimal number of performance features to highlight the differences between the four pianists and draw a unique portrait of each pianist's performing individuality.

Following this method, 16 performance features were selected among the 159 previously identified as significantly differing between performers and consistent between repetitions. With these 16 selected performance features, a minimal, unique and meaningful description of each pianist's individuality in the 240 recorded performances is obtained. The overall performance portraits of each pianist according to these 16 performance features are presented in **Figure 3**.

The 16 selected performance features are described below, and the corresponding statistical scores (ANOVA F-ratio, p-value and effect size) are provided.

- Hammer velocity [$F(3, 19.66) = 13.141$, $p < 10^{-4}$, $\eta^2 = 0.354$]: maximum hammer velocity for each note, as directly measured by the piano sensors; as a direct correlate to intensity, it makes for a descriptor of dynamic level. Values are indicated in 8-bit steps (from 0 to 250).
- Difference in hammer velocity between hands [$F(3, 15.79) = 3.611$, $p = 0.037$, $\eta^2 = 0.234$]: compares hammer velocity between notes played with the right hand and notes played with the left hand, and can thus underline a dynamic emphasis with one hand.
- Variations in hammer velocity [$F(3, 17.01) = 16.603$, $p < 10^{-4}$, $\eta^2 = 0.570$]: describes the range of hammer velocities reached in each performance; values are indicated as ratios of deviation from the average hammer velocity.

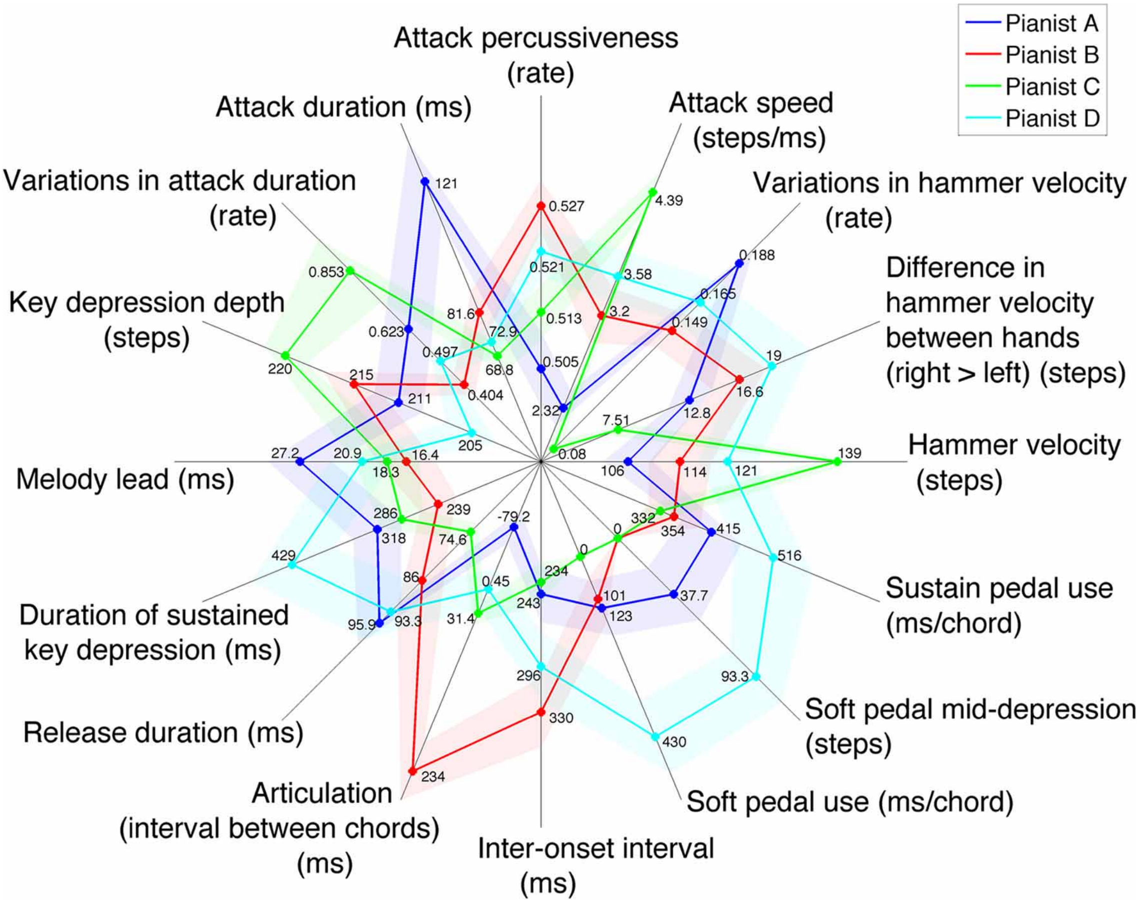


FIGURE 3 | Kiviat chart of the 16 performance features giving a minimal and unique description of four pianists' individual performance patterns. Z-scores per pianist are plotted for each feature with colored dots, with the corresponding unnormalied values indicated alongside. The four colored, dot-linking closed lines portray each pianist's performing style. Shades around each closed line show the ±1.96 SE. intervals (95% confidence interval).

- Attack speed [$F(3, 19.48) = 15.413$, $p < 10^{-4}$, $\eta^2 = 0.420$]: mean attack speed (in 8-bit steps per ms) from the beginning of key depression to the maximum depression reached in the note; although this feature is highly correlated with hammer velocity, some differences can occur, as hammer velocity is defined by the instant key speed at hammer launch instead of the mean attack speed over the keystroke.
- Attack percussiveness [$F(3, 2.38) = 15.496$, $p = 0.043$, $\eta^2 = 0.226$]: an evaluation of keystroke acceleration during attack, this feature indicates the convexity of the keystroke curve: the higher the early key acceleration, the more concave the keystroke curve, and the more percussive the attack—as it corresponds to a key struck rather than pressed (Goebl et al., 2005); a value of 0.5 would indicate a linear attack (on average), while values over 0.5 suggest a concave, more percussive touch.
- Attack duration [$F(3, 14.61) = 6.836, p = 0.004, \eta^2 = 0.333$]: time interval (in ms) from the start of keystroke to its maximum depression; although highly negatively correlated to attack speed (the faster the attack, the shorter its duration), it also depends on nuances of articulation and touch at note onsets, including attack percussiveness and key depression depth.
- Variations in attack duration [$F(3, 11.79) = 4.405, p = 0.027$, $\eta^2 = 0.217$]: variations of attack durations between chords and in time; values are indicated as ratios of deviation from the average attack duration.
- Key depression depth [$F(3, 17.69) = 7.945$, $p = 0.001$, $\eta^2 = 0.320$]: indicates how deep (close to the keybed) each key gets depressed for each note; values are given in 8-bit steps.
- Melody lead [$F(3, 6.38) = 32.003$, $p = 3.1 \cdot 10^{-4}$, $\eta^2 = 0.095$]: time interval (in ms) describing the advance of the first note of a chord on the others; as despite its description in terms of timing, melody lead is essentially a velocity artifact of the dynamic accentuation of the melody note (Goebl, 2001), it is thus a feature of polyphonic, dynamically differentiated touch. Despite a small effect size, the significance of this feature is supported a high statistical power ($\pi = 0.969$), and it was found to be independent from other features of attack, touch and articulation according to hierarchical clustering analysis.
- Duration of sustained key depression [$F(3, 11.99) = 4.744$, $p = 0.021$, $\eta^2 = 0.201$]: time (in ms) for which a key is held depressed, after attack and before release; although tempo can bear an effect on this feature, it is also a descriptor of articulation strategies.
- Release duration [$F(3, 18.41) = 3.545$, $p = 0.035$, $\eta^2 = 0.148$]: time (in ms) taken for releasing the key (from the start of the key moving up to its reaching rest position); this feature essentially accounts for articulation: a note released slowly (thus slowed by the finger) may probably overlap with the next.
- Articulation (interval between same-hand chords) [$F(3, 8.84) = 10.929$, $p = 0.002$, $\eta^2 = 0.565$]: time interval (in ms) from the end of a chord or single note to the start of the next one played with the same hand; negative values indicate *legato*, positive values *staccato*.
- Inter-onset interval [$F(3, 13.28) = 6.151$, $p = 0.008$, $\eta^2 = 0.168$]: time (in ms) elapsed between the start of two consecutive chords; this feature essentially serves as a descriptor of average tempo (Dixon, 2001).
- Soft pedal use [$F(3, 14.94) = 6.356$, $p = 0.005$, $\eta^2 = 0.316$]: average time (in ms) during which the soft pedal is depressed while a chord is being played.
- Soft pedal mid-depression [$F(3, 6.79) = 16.249$, $p = 0.002$, $\eta^2 = 0.380$]: indicates the amount (in 8-bit steps) of part-pedaling during chords; the higher the value, the more the soft pedal was kept only partially depressed.
- Sustain pedal use [$F(3, 16.59) = 3.297$, $p = 0.046$, $\eta^2 = 0.133$]: average time (in ms) during which the sustain pedal is depressed while a chord is being played.

The first eight performance features presented describe dynamics and attack. They reveal that Pianist A used the lowest dynamics (hammer velocity), attack speed and percussiveness, while featuring the longest attacks, and rather shallow key depressions. She also showed the largest variations in intensity between chords. On the other hand, Pianist C applied the highest intensity, fastest attacks and deepest key depressions. His attacks were the longest, but less saliently than his attack speeds could have led to believe. This may be explained by the rather average percussiveness of his attacks: with a pressed touch, key depression starts with zero velocity, which at equal hammer velocity and attack speed may inflate attack duration compared with a struck touch. Pianist C also presented the highest variations in attack duration, yet was the most constant in hammer velocity, which indicates he may have varied his touch percussiveness while always reaching high intensities. Pianist B's attacks are short, deep and the most percussive, as well as the most consistent in duration, yet they are not very fast, and produce low to average intensity (with average consistency). As for Pianist D, he played with average-to-high intensity, attack speed and percussiveness, with equally average-to-high variations in intensity between chords. His attacks were consistently short. Yet his key depressions were the shallowest. Furthermore, all four pianists applied higher intensities with the right hand, but to varying degrees (the least by Pianist C, the most by Pianist D). As a side note, these eight features of attacks and dynamics were significant (in the sense previously indicated) over the chords played by each hand separately, except for attack percussiveness and the difference in hammer velocity between hands (the latter ineligible). Of the six features, all but hammer velocity were more salient in left-hand chords than in right-hand chords, meaning that the differences in attack between the four pianists were more manifest in their left-hand playing than in their right-hand playing.

The following four performance features represented describe articulation, polyphonic touch, and tempo. The melody lead effect was longer on average for Pianist A than for the three others. Pianist A shows the most *legato* articulation and the longest key releases. On the other hand, the duration of her sustaining key depressions is about average, and inter-onset intervals are short, i.e., a fast tempo (actually close to score indications). Likewise, Pianist C played at a fast average tempo (as indicated on the scores) given his short inter-onset intervals. Yet his articulation is *non-legato*, with the shortest key releases and average durations of key sustains. Pianist B played extremely *staccato* compared with the others. His key releases remained of average duration, but his key sustains were the shortest, which left enough space for large intervals between chords despite the slowest tempo he favored. His preference for *staccato* articulation matches well with his short and percussive, yet not very fast, attacks. For Pianist D, key sustains and releases were very long, which may be mostly due to his long inter-onset intervals (slow tempo), whereas his articulation remains essentially *non-legato*.

Finally, the last three performance features describe pedaling. Pianist D used the soft pedal extensively (including part-pedaling), while Pianist C never used the soft pedal. Pianists A and B used the soft pedal sparingly, with some part-pedaling from A but none from B. The sustain pedal was also used extensively by Pianist D, as well as by Pianist A. Pianists B and C used the sustain pedal the least, yet still quite a bit.

3.1.4. Timbre-pianist interaction

In complement to this characterization of the four pianists' general individuality over all performances (regardless of the timbral nuance expressed), we also aimed at determining the idiosyncrasies in pianists' performances that could arise specifically in the expression of each of the five timbral nuances considered. Separate statistical analyses (ANOVA) of performance features were performed to this aim, over each set of performances highlighting the same timbral nuance. However, these separate analyses are only valid for the performance features protected by a significant effect of the timbre-pianist interaction in the general ANOVA.

The effect of the pianist-timbre interaction, with regard to the 616 performance features, was thus examined in the general, three-Way repeated-measures ANOVA (with performer as random factor) over all performances. Violations of sphericity were corrected with the Huynh–Feldt epsilon applied to the degrees of freedom. The effect of the pianist-timbre interaction was significant (at the 5% level for corrected *p*-values) for 149 performance features.

Among these 149 performance features, 86 also showed a significant effect of the pianist. Yet the features of melody lead and variations in attack duration, included in the overall performance portrait as two of the most relevant descriptors of pianists' general individuality, were not found significant in the pianist-timbre interaction. For the variations in attack duration, the non-significant effect of pianist-timbre interaction is supported by non-negligible statistical power ($\pi = 0.456$), which may allow to infer that the variations in attack duration characterize general idiosyncratic traits that are not influenced by the timbral nuance performed. However, given the low statistical power of the non-significant pianist-timbre interaction for melody lead ($\pi = 0.188$), the same inference should not be made.

As for the features significant for the interaction but not for the effect of pianist, they will be discussed in relation with the timbral nuance(s) for which they bear a particular idiosyncratic effect.

3.2. PIANISTS' INDIVIDUALITY IN PERFORMING EACH OF FIVE TIMBRAL NUANCES

The performance strategies and performance features highlighting pianists' individuality in the production of five different timbral nuances were sought out, among the 149 performance features that showed a significant effect of pianist-timbre interaction in the general ANOVA over all performances.

Statistical analyses of variance were conducted separately over each set of 48 performances highlighting one of the five timbral nuances. Five Two-Way repeated measures ANOVAs were performed, with each of the 149 performance features as dependent variable, with the performer as random factor, and piece and repetition (of the same performer-×-piece condition) as fixed-effect factors.

Like in the overall case, for each ANOVA, the between-subject effect of performer was considered. The corresponding assumptions of normal distributions (Kolgomorov–Smirnov test) and homoscedasticity (Levene's test) were tested, and a non-parametric, one-way Kruskal–Wallis rank analysis of variance was performed to confirm or reject the significance of an ANOVA whose assumptions were not met. Effect sizes and statistical power were also calculated. Moreover, the within-subject effect of repetition was taken into account, with the degrees of freedom of the ANOVA corrected for violation of sphericity with the Huynh–Feldt epsilon. A performance feature was then considered as revealing pianists' individuality when the effect of performer was significant at the 5% level (in rejecting the null hypothesis of equal variance between pianists) while the effect of repetition was not (i.e., no significant differences between repetitions).

Consequently, for the sets of performances highlighting each of the bright, dark, dry, round, and velvety timbral nuances, significant differences between pianists and consistence between repetitions were identified in 86, 83, 69, 72, and 97 performance features (respectively).

3.2.1. Reduced performance spaces

For each set of 48 performances highlighting one of the five timbral nuances, Principal Component Analysis was applied to the corresponding performance features selected, whose values were normalized (Z-score per feature over 48 performances).

In each of the five cases, the first two principal components were sufficient to explain a large part of the input variance:

- Bright timbre: 60.6% (41.7 and 18.9% respectively)
- Dark timbre: 60.6% (34.4 and 26.2% respectively)
- Dry timbre: 64.1% (45.1 and 19% respectively)
- Round timbre: 55.5% (39.5 and 16% respectively)
- Velvety timbre: 57.5% (39.9 and 17.6% respectively)

For each of the five timbral nuances, Varimax factor rotation of the first two PCA loadings was attempted, yet did not noticeably improve the clarity of the reduced performance spaces, nor provided a more optimal/interpretable distribution of loadings between features along each principal component, and was thus discarded.

For each of the five timbral nuances, the position of each same-timbre performance according to its coordinates in the two dimensions formed by the corresponding first two principal components is presented in a separate plot in **Figure 4**.

For all five timbres, the loadings of the first dimensions primarily account for overall dynamics (hammer velocity) and

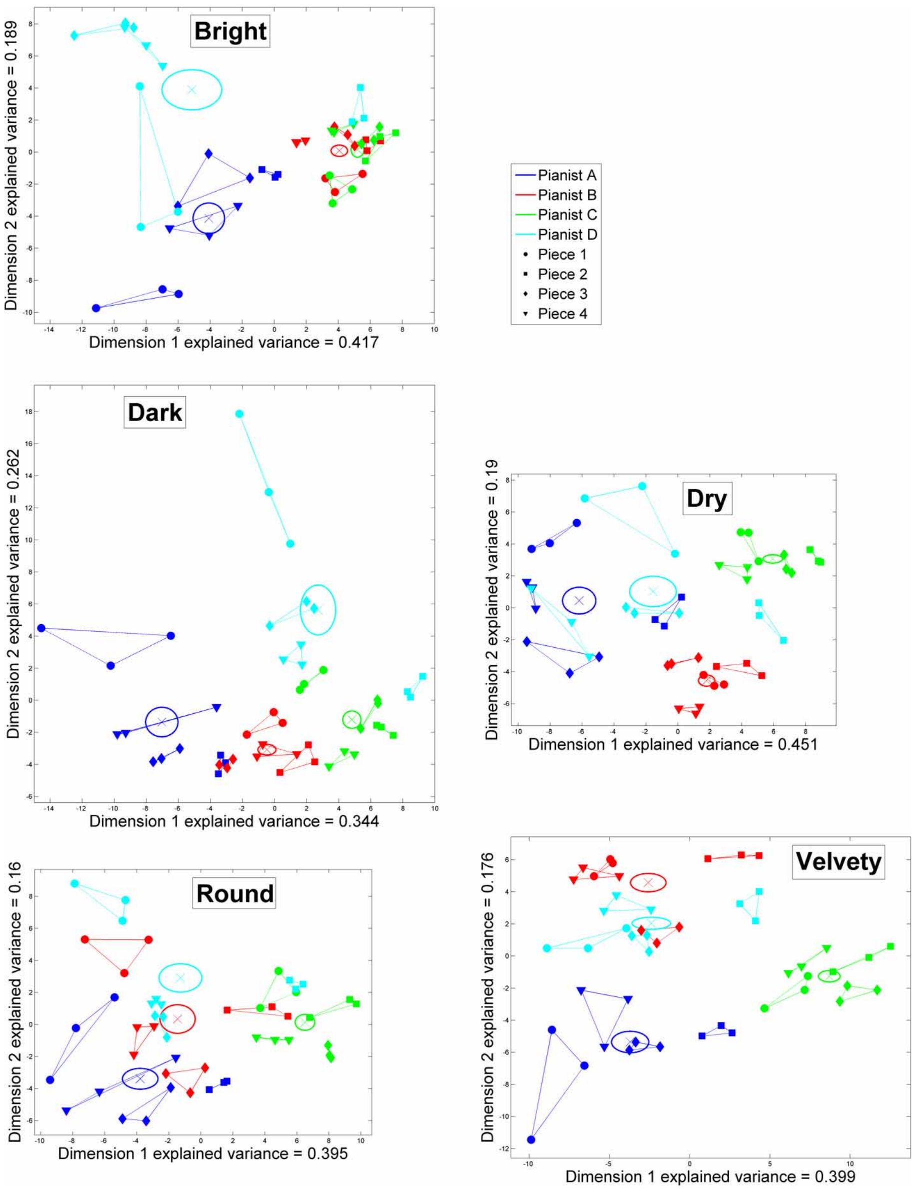


FIGURE 4 | Principal Component Analysis of the performance features highlighting pianists' individuality over each set of 48 performances corresponding to a different timbral nuance (bright, dark, dry, round, or velvety): two-dimensional reduced performance spaces. In each of the five subplots, the colored lines link the same-piece repeated performances by each pianist. Averages per pianist are indicated by colored crosses, and ±1 SE with ellipses.

attack speeds. For all timbres but dark, these features are also predominant with regard to the left hand, whereas for the dark timbre these features are more weighted for the right hand than for the left. Attack durations are also highly weighted, primarily with regard to the left hand (especially for the dark timbre). Variations in hammer velocity also largely contribute to the first dimension, overall and with regard to the left hand (but not the right), and especially for the bright timbre. Other features of touch and articulation contribute, to a lesser degree, to the first dimensions: key depression depth (bright, dry, round), attack percussiveness (dark, velvety), right-hand overlaps (bright); key release durations (round, velvety).

The descriptions of second dimensions according to loadings feature, for all timbres but bright, common trends of articulation, tempo, and pedaling, although in different combinations and weighting order. From highest to lowest weighted features, for dark: note and key sustain durations, sustain pedal use, soft pedal use, and articulation. For dry: sustain pedal use and articulation. For round: inter-onset intervals, sustain pedal use, note and key sustain durations, articulation. For velvety: articulation and soft pedal use. On the other hand, the second dimension for the bright timbre essentially corresponds to right-hand attack speeds and durations, then right-hand hammer velocities, and much less so to articulation.

It must also be mentioned that, in the reduced space of bright-timbre performances, soft pedal use is also accounted for in both dimensions (especially the second). However, only Pianist D used the soft pedal in bright-timbre performances, in the seven performances isolated in the upper left of the performance space. For the other 41 bright-timbre performances, soft pedal features bear no influence.

On average, Pianists B and C tend to be the most consistent for each timbre, both in same-piece repetitions and between pieces (with the exception of Pianist B's round-timbre performances). Difference between pianists are more or less salient depending on timbres and pianists. Pianists A and C are clearly differentiated for all timbres, as their performances occupy different regions of the performance spaces. Pianists A and D differ the most for timbres dark and velvety. Pianist B's dry-timbre performances are clearly singled out, yet for other timbres his performances overlap with different pianists. In particular, Pianists B and C's bright-timbre performances are all but indistinguishable in the corresponding performance, which means that they adopted equivalent strategies as regards the performance features relevant to the individuality-highlighting performance space of the bright timbre.

Moreover, some idiosyncratic tendencies revealed in the overall reduced performance space also appear in the timbre-wise reduced performance spaces, yet may be nuanced depending on the timbral nuance. Indeed, although Pianist A generally shows low dynamics, and slow and long attacks (along the first dimensions), this fact is more salient for dark-timbre performances (and dry to a lesser degree), while intensity, attack speeds and durations are more similar between Pianists A, B and D for round and velvety timbres. Pianists A and D are also quite similar in intensity, attack speeds and durations for a bright timbre, yet only along the first dimension. Indeed, the second dimension accounts for right-hand dynamics/attacks (and with the previously stated effect of soft pedal use by Pianist D in the seven upper-left performances notwithstanding) and highlights a difference between Pianist A and Pianist D's performances of piece no.2. On the other hand, Pianist C's high dynamics, and fast and short attacks, are most salient in velvety-timbre performances, and equivalent to Pianist B's for a bright timbre. Meanwhile, as the performance features associated with the second dimensions differ more largely between timbres, only timbre-wise tendencies can be brought up. For a bright timbre, differences between pianists can be explained by Pianist D's use of the soft pedal on one hand, and by Pianist A's lower right-hand dynamics/attacks on the other hand. For a dark timbre, Pianist D must have used longer notes and more sustain pedal, especially for piece no.1. For a dry timbre, Pianist B is probably set apart by his *staccato* articulation. For a round timbre, a conjecture is more complex to establish. Pianists A and D may be distinguished by tempo (fast vs. slow), while the same difference in tempo between Pianists B and C may be compensated by their different articulations. Finally, for a velvety timbre, articulation may suffice to explain the separation between Pianists A and B, while heavy soft pedal use may largely contribute to the higher-end coordinates of Pianist D's performances along dimension 2.

3.2.2. Descriptive performance portraits

Following the same procedure as in the general case, the performance features most salient, relevant, meaningful, and non-redundant for highlighting pianists' individuality were sought out, for each of the bright, dark, dry, round, and velvety timbral nuances, among the 86, 83, 69, 72, and 97 (resp.) performance features selected for their eliciting significant differences between pianists and consistence between repetitions.

For each timbral nuance, the corresponding selected performance features were divided into the four functionally independent categories of dynamics/attack, articulation, soft pedal and sustain pedal. Correlations were calculated between all performance features within each category, and the most correlated/redundant features were grouped by hierarchical clustering. The most significant, meaningful and interpretable performance feature in each cluster was then selected.

For each of the bright, dark, dry, round, and velvety timbres, the 9, 9, 9, 12, and 13 (resp.) most relevant performance features for describing pianists' individuality were thus selected. The descriptive performance portraits of pianists' individuality for each timbral nuance are presented in **Figure 5**.

Most of the performance features used in these performance portraits of pianists' individuality in the production of each timbral nuance were already featured in the general performance portrait of pianists' individuality, and were accordingly described in Section 3.1.3. However, the following performance features were not previously introduced. In the performance portrait of individuality for a dry timbre:

- Variations in key depression depth: ratio of deviation between chords from the performance average of key depression depth.
- Sustain pedal depression: average depth (in 8-bit steps) of pedal depression per chord.

In the performance portrait of individuality for a round timbre:

- Difference in attack speed between hands: comparison of mean attack speed between notes played with the right hand and notes played with the left hand; values in steps/ms can indicate either faster (> 0) or slower (< 0) attacks with the right hand than the left.

In the performance portrait of individuality for a velvety timbre:

- Soft pedal depression: average depth (in 8-bit steps) of pedal depression per chord.

For each of the five performance portraits of the different timbral nuances, the statistical scores (ANOVA F-ratio, p-value and effect size) corresponding to each of its descriptive feature are provided below.

In the descriptive performance portrait of individuality for a bright timbre (9 features):

- Hammer velocity: $F(3, 10.16) = 16.882$, $p = 2.8 \cdot 10^{-4}$, $\eta^2 = 0.475$

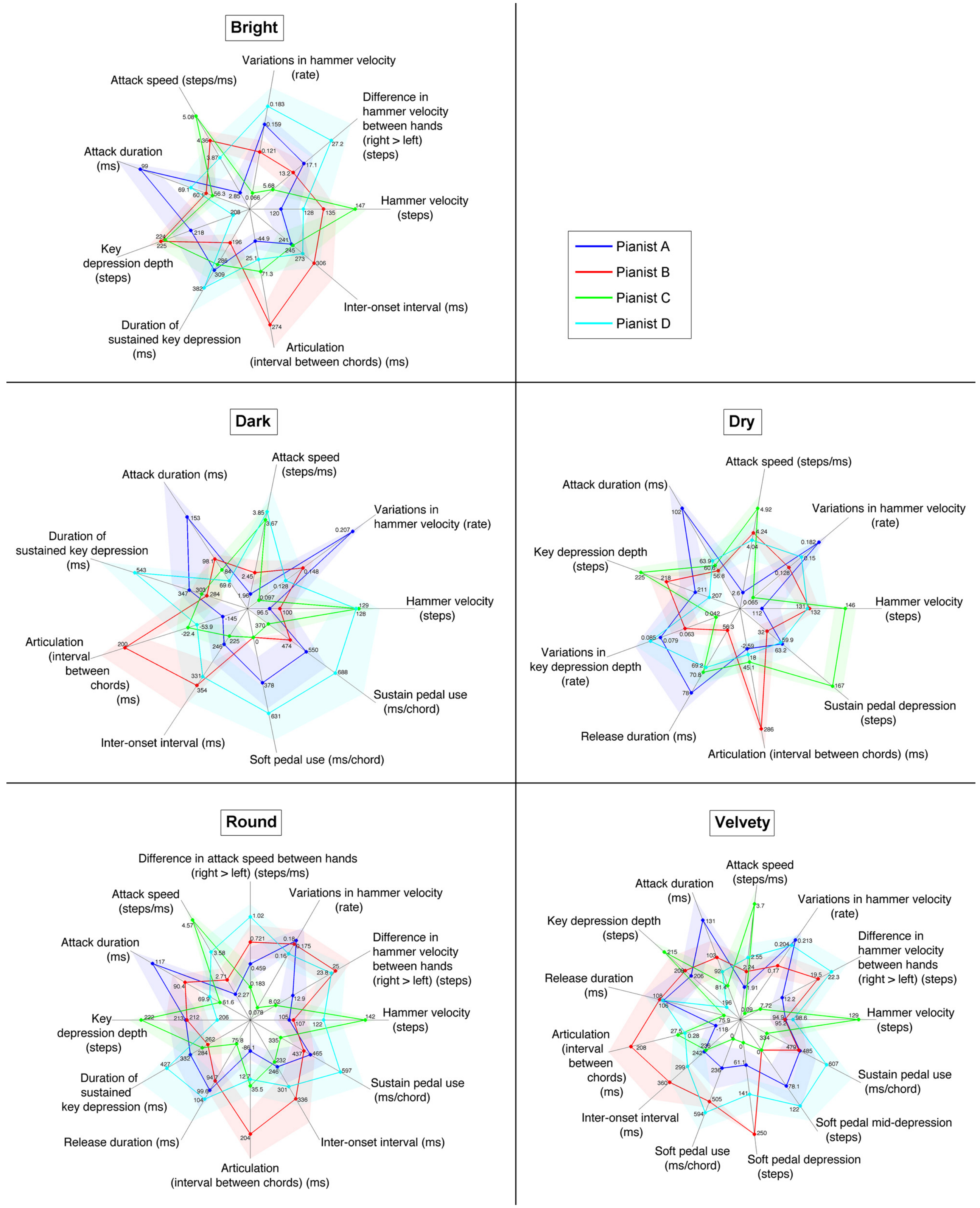


FIGURE 5 | Kiviats charts of the performance features giving a minimal and unique description of four pianists' individual performance patterns in the production of each of five timbral nuances (bright, dark, dry, round, and velvety). Z-scores per pianist are plotted for each feature with colored dots, with the corresponding unnormalized values indicated alongside. The four colored, dot-linking closed lines portray each pianist's performing style. Shades around each closed line show the ±1.96 SE. intervals (95% confidence interval).

- Difference in hammer velocity between hands: $F(3, 8.43) = 5.962$, $p = 0.018$, $\eta^2 = 0.484$
- Variations in hammer velocity: $F(3, 10.84) = 16.060$, $p = 2.6 \cdot 10^{-4}$, $\eta^2 = 0.703$
- Attack speed: $F(3, 10.85) = 24.475$, $p < 10^{-4}$, $\eta^2 = 0.660$
- Attack duration: $F(3, 10.66) = 8.415$, $p = 0.004$, $\eta^2 = 0.581$
- Key depression depth: $F(3, 8.94) = 11.366$, $p = 0.002$, $\eta^2 = 0.597$
- Duration of sustained key depression: $F(3, 9.97) = 5.504$, $p = 0.017$, $\eta^2 = 0.221$
- Articulation (interval between same-hand chords): $F(3, 7.31) = 12.243$, $p = 0.003$, $\eta^2 = 0.638$
- Inter-onset interval: $F(3, 8.03) = 5.463$, $p = 0.024$, $\eta^2 = 0.089$

In the descriptive performance portrait of individuality for a dark timbre (9 features):

- Hammer velocity: $F(3, 11.74) = 22.588$, $p < 10^{-4}$, $\eta^2 = 0.638$
- Variations in hammer velocity: $F(3, 9.55) = 22.551$, $p = p = 1.2 \cdot 10^{-4}$, $\eta^2 = 0.736$
- Attack speed: $F(3, 6.39) = 18.964$, $p = 0.001$, $\eta^2 = 0.600$
- Attack duration: $F(3, 7.02) = 5.454$, $p = 0.030$, $\eta^2 = 0.375$
- Duration of sustained key depression: $F(3, 9.33) = 10.297$, $p = 0.003$, $\eta^2 = 0.343$
- Articulation (interval between same-hand chords): $F(3, 9.02) = 8.324$, $p = 0.006$, $\eta^2 = 0.541$
- Inter-onset interval: $F(3, 12.21) = 7.169$, $p = 0.005$, $\eta^2 = 0.280$
- Soft pedal use: $F(3, 8.54) = 5.876$, $p = 0.018$, $\eta^2 = 0.516$
- Sustain pedal use: $F(3, 9.06) = 7.443$, $p = 0.008$, $\eta^2 = 0.375$

In the descriptive performance portrait of individuality for a dry timbre (9 features):

- Hammer velocity: $F(3, 8.85) = 27.179$, $p < 10^{-4}$, $\eta^2 = 0.558$
- Variations in hammer velocity: $F(3, 3.57) = 28.319$, $p = 0.006$, $\eta^2 = 0.634$
- Attack speed: $F(3, 9.40) = 24.816$, $p < 10^{-4}$, $\eta^2 = 0.619$
- Attack duration: $F(3, 10.43) = 7.148$, $p = 0.007$, $\eta^2 = 0.535$
- Key depression depth: $F(3, 12.08) = 13.489$, $p = 3.7 \cdot 10^{-4}$, $\eta^2 = 0.504$
- Variations in key depression depth: $F(3, 6.26) = 9.163$, $p = 0.011$, $\eta^2 = 0.463$
- Release duration: $F(3, 10.14) = 7.132$, $p = 0.007$, $\eta^2 = 0.433$
- Articulation (interval between same-hand chords): $F(3, 7.62) = 11.342$, $p = 0.003$, $\eta^2 = 0.685$
- Sustain pedal depression: $F(3, 7.18) = 14.834$, $p = 0.002$, $\eta^2 = 0.676$

In the descriptive performance portrait of individuality for a round timbre (12 features):

- Hammer velocity: $F(3, 9.02) = 35.453$, $p < 10^{-4}$, $\eta^2 = 0.589$
- Difference in hammer velocity between hands: $F(3, 8.12) = 9.802$, $p = 0.004$, $\eta^2 = 0.548$
- Variations in hammer velocity: $F(3, 5.52) = 18.583$, $p = 0.003$, $\eta^2 = 0.624$
- Difference in attack speed between hands: $F(3, 8.64) = 4.131$, $p = 0.044$, $\eta^2 = 0.431$
- Attack speed: $F(3, 8.92) = 53.632$, $p < 10^{-4}$, $\eta^2 = 0.686$
- Attack duration: $F(3, 9.47) = 12.217$, $p = 0.001$, $\eta^2 = 0.575$
- Key depression depth: $F(3, 8.98) = 12.875$, $p = 0.001$, $\eta^2 = 0.470$
- Duration of sustained key depression: $F(3, 9.22) = 7.350$, $p = 0.008$, $\eta^2 = 0.211$
- Release duration: $F(3, 10.60) = 6.136$, $p = 0.011$, $\eta^2 = 0.376$
- Articulation (interval between same-hand chords): $F(3, 9.38) = 9.385$, $p = 0.004$, $\eta^2 = 0.569$
- Inter-onset interval: $F(3, 10.91) = 7.121$, $p = 0.006$, $\eta^2 = 0.188$
- Sustain pedal use: $F(3, 10.40) = 8.627$, $p = 0.004$, $\eta^2 = 0.387$

Finally, in the descriptive performance portrait of individuality for a velvety timbre (13 features):

- Hammer velocity: $F(3, 5.17) = 92.653$, $p < 10^{-4}$, $\eta^2 = 0.606$
- Difference in hammer velocity between hands: $F(3, 9.34) = 11.909$, $p = 0.002$, $\eta^2 = 0.650$
- Variations in hammer velocity: $F(3, 6.64) = 35.600$, $p = 1.8 \cdot 10^{-4}$, $\eta^2 = 0.789$
- Attack speed: $F(3, 4.13) = 69.362$, $p = 5.5 \cdot 10^{-4}$, $\eta^2 = 0.635$
- Attack duration: $F(3, 8.76) = 10.662$, $p = 0.003$, $\eta^2 = 0.436$
- Key depression depth: $F(3, 8.11) = 25.073$, $p = 1.9 \cdot 10^{-4}$, $\eta^2 = 0.575$
- Release duration: $F(3, 5.87) = 14.320$, $p = 0.004$, $\eta^2 = 0.489$
- Articulation (interval between same-hand chords): $F(3, 7.41) = 12.912$, $p = 0.003$, $\eta^2 = 0.624$
- Inter-onset interval: $F(3, 8.84) = 9.085$, $p = 0.005$, $\eta^2 = 0.245$
- Soft pedal use: $F(3, 8.93) = 11.448$, $p = 0.002$, $\eta^2 = 0.613$
- Soft pedal depression: $F(3, 10.98) = 20.321$, $p < 10^{-4}$, $\eta^2 = 0.759$
- Soft pedal mid-depression: $F(3, 10.00) = 6.684$, $p = 0.009$, $\eta^2 = 0.542$
- Sustain pedal use: $F(3, 7.53) = 8.579$, $p = 0.008$, $\eta^2 = 0.365$

The complete list of performance features selected as significant, consistent, meaningful and non-redundant in highlighting pianists' individuality in the production of at least one timbral nuance and/or overall is presented in **Table 1**. For each performance feature and each timbral nuance, the table indicates whether the feature was selected in the descriptive portrait, or significant but redundant with others, or not protected by a significant effect of timbre-pianist interaction in the general ANOVA, or whether non-significance was conclusive (with regard to statistical power).

Some of the pianists show consistent patterns across timbres along some of the performance features, which mostly reflect the general descriptive portrait, and also correspond to what could be deduced from the reduced performance spaces. Pianist A always presents the lowest hammer velocities, and the longest and slowest attacks. On the other hand, Pianist C always produced the highest and most regular hammer velocities, as well as fast and short attacks. In the four timbral nuances (all but dark) for which

Table 1 | Performance features most characteristic of pianists' individuality, in performing five timbral nuances and overall.

Timbre: **Performance features**	**Bright**	**Dark**	**Dry**	**Round**	**Velvety**	**All**
ATTACK AND DYNAMICS						
Hammer velocity	O	O	O	O	O	O
Difference in hammer velocity between hands	O	×	–	O	O	O
Variations in hammer velocity	O	O	O	O	O	O
Attack speed	O	O	O	O	O	O
Difference in attack speed between hands	S	–	×	O	×	×
Attack percussiveness	×	S	–	×	S	O
Attack duration	O	O	O	O	O	O
Variations in attack duration	.	.	.	.	.	O
Key depression depth	O	–	O	O	O	O
Variations in key depression depth	S	–	O	×	×	S
ARTICULATION						
Articulation (intervals between chords)	O	O	O	O	O	O
Duration of sustained key depression	O	O	×	O	×	O
Release duration	×	×	O	O	O	O
Inter-onset interval	O	O	×	O	O	O
Melody lead	.	.	.	.	.	O
PEDALS						
Soft pedal use	S	O	S	X	O	O
Soft pedal depression	S	×	×	X	O	–
Soft pedal mid-depression	S	S	×	S	O	O
Sustain pedal use	×	O	X	O	O	O
Sustain pedal depression	×	–	O	–	–	–

The following symbols were used for: O, feature included in the descriptive portrait of the timbral nuance; S, feature significant in highlighting pianists' individuality, but redundant with others; "." (dots), feature not protected by the pianist-timbre interaction in the general ANOVA; the other features were not significant with regard to individuality, "–" (dashes), feature non-significant with regard to individuality, with low statistical power ($\pi < 0.2$), thus inconclusive; x; id., with average statistical power ($0.2 < \pi < 0.8$); X: id., with high statistical power ($\pi > 0.8$), thus conclusive.

key depression depth was selected as a performance significant for pianists' individuality, Pianist C always applied very deep key depressions (close to the keybed), while on the other hand Pianist D always employed the shallowest key depressions. On average, for each of the four timbral nuances, maximum key depressions per note were approximately 10% deeper for Pianist C than for Pianist D. As for articulation, Pianist A always played with the most *legato*, and Pianist B with the most *staccato*. Key depression sustains were also consistently the longest for Pianist D, and the shortest for Pianist B, although this performance feature was only significant in highlighting individuality for three of the five timbral nuances (bright, dark, and round). Finally, inter-onset intervals could significantly portray the pianists' individuality in performing four timbral nuances (all but dry), and were larger for Pianist B (and Pianist D to a lesser degree) than for Pianists A and C, indicating the same differences in average tempo as previously described in the general case over all performances.

Yet otherwise, different performance patterns between pianists arose along performance features in the production of different timbral nuances. Dynamic balance between hands (in hammer velocity) was only significant in the production of bright, round and velvety timbres (although its non-significance for a dry timbre was inconclusive due to low statistical power). Pianist D always largely emphasized the right hand, while Pianist C always used the least right-hand emphasis. Pianist B drastically changed his right-hand dynamic emphasis between timbres, from average (among the four pianists) for bright and velvety timbres, to the most of all for round. Balance between hands in attack speed was also selected as a relevant feature for describing individuality in the production of a round timbre. In contrast with the corresponding dynamic balance between hands, the discrepancy in attack speed between hands (toward the right) for a round timbre is larger (and largest) for Pianist D than Pianist B, although the latter shows a more pronounced dynamic emphasis on the right hand. Dynamic variations within a performance also largely changed between timbres for Pianists A, B, and D. Pianist A's dark performances always featured high dynamic variations, and by far the most of all pianists. On the other hand, Pianist D used even more dynamic variations than her for the production of a bright timbre. Pianist D's dynamic variations are also high for dry, round, and velvety timbres, yet below average for dark. Meanwhile, Pianist B's dynamic variations range from average among pianists for bright, dark and dry timbres, to very high for round performances. Patterns of hammer velocities, attack speeds and durations also change between timbres for Pianists B and D. For a bright timbre, Pianist D's attacks are about average in intensity, speed and duration, while Pianist B's attacks are faster, especially shorter, and brought higher hammer velocities, both than average and than Pianist D's. For the dry timbre, Pianists B and D's attacks are mostly equivalent in intensity, speed and duration, yet although their hammer velocities and attack speeds are average and well below Pianist C's, their attacks are among the shortest (with Pianist C's). For Pianist B, these shortened attacks (despite average speeds) can be explained by his preferred *staccato* articulation, while for Pianist D the same attack characteristics may stem from his shallow key depressions. On the other hand, for round and velvety timbres, Pianist B's attacks are slower and longer than average, while Pianist D's attacks are faster and shorter than Pianists B's and than average (especially for round). These patterns result for Pianist D in hammer velocities average for round, yet quite lower for velvety, while for both timbres Pianist B's hammer velocities are the lowest along with Pianist A's. Finally, for producing a dark timbre, Pianist B remains around low intensities and slow/long attacks, whereas Pianist D employed the fastest and shortest attacks, and the (nearly) highest hammer velocities of all four pianists.

As for key depression depths, besides the consistently deep vs. shallow key depressions for Pianists C and D (resp.), both applied slightly shallower key depressions (while still differing from each other by about 10%.) in performing a velvety timbre than for the other three nuances (bright, dry and round) for which key depression depth was a relevant feature of individuality (key depression depth was inconclusively non-significant in highlighting pianists individuality in dark-timbre performances). While Pianist A remained fairly constant, and average among pianists, in average key depression depth per timbral nuance, Pianist B varied the most in key depression depth between timbral nuances, both in the absolute and with regard to the other pianists, ranging from the deepest key depressions (in performing a bright timbre) to average ones (in performing a round timbre). Moreover, variations in key depression depth (within a performance) were only relevant feature for individuality in the case of a dry timbre. The pianists with the shallower average key depressions (A and D) also show the largest variations (and vice versa), which might be due (at least in part), to a ceiling effect (key depressions cannot get deeper than the keybed; the deeper the key depressions on average, the less room there remains for variations).

Articulation patterns also reflect a different picture of pianists' individuality depending on the timbral nuance expressed.

Key depression sustains were indeed much shorter for Pianist B (than the three others) in the case of a bright timbre, whereas their durations were much closer between Pianists B and C (and Pianist A to a lesser degree) for timbres dark and round.

Key release duration, a significant descriptor of individuality in the production of the three timbres dry, round and velvety, highlighted very different patterns between pianists depending on the timbral nuance. Although Pianist A always featured long key releases, they were only the longest (or more appropriately, the least short) with a dry timbre. Pianist B's key releases were the shortest for a dry timbre, yet the longest for velvety. Pianist D's key releases were also very long (relative to the other pianists) for velvety and especially round timbres, yet only average (and much shorter in the absolute) for dry. On the other hand, Pianist C's key releases, the shortest for round and velvety timbres, were only average among pianists for the dry timbre, although they remained of fairly constant duration in the absolute between timbres (i.e., contrary to the others pianists, Pianist C did not use or show different key release durations in his performing the three different timbral nuances dry, round and velvety).

Articulation, as described by the timing intervals between same-hand chords, presents mostly consistent patterns of individuality between four of five timbral nuances, yet for a dry timbre Pianist A's articulation is not significantly more *legato* than Pianists C and (especially) D, as all three feature an articulation best described as *non-legato* in performing a dry timbre.

Furthermore, the difference in inter-onset intervals (i.e., average tempo) between Pianists B/D and A/C was much more salient for the dark timbre than for round, velvety and (especially) bright.

Lastly, the differences in pedaling strategies between pianists largely vary depending on the timbral nuance performed. In performing a velvety timbre, the soft pedal was used by Pianists A (sparingly), D (massively), and B (constantly, in all velvety performances). Meanwhile, only Pianists A and D used the soft pedal in performing a dark timbre (both to the same extent as for velvety). For the other three timbral nuances, the soft pedal (although it is not presented in the descriptive portraits) was only used in some of Pianist D's performances, and never by the three other pianists.

Finally, the duration of sustain pedal use with regard to chords was significant for timbres dark, round, and velvety, and although idiosyncratic patterns were largely consistent between timbres, with the most use for Pianist D, the least for Pianist C, and average relative use for Pianist A, Pianist B tended to use more sustain pedal (especially with regard to Pianist A) from timbres dark to velvety. Although sustain pedal use was conclusively non-significant in highlighting pianists' individuality for a dry timbre, the amount of sustain pedal depression was significant, and especially highlighted an higher amount of depression for Pianist C.

4. DISCUSSION

In summary, individual strategies between four pianists were successfully revealed, within 240 performances of four pieces with five different timbral nuances, by the fine-grained performance control features extracted from the high-accuracy key, pedal and hammer-tracking data gathered by the CEUS system. The four pianists were thus shown to elicit idiosyncratic patterns of dynamics, attack speed, attack touch (percussiveness, duration, depth, melody lead), articulation, key sustains and releases, average tempo, and pedaling.

Among these performance parameters, some would have been accessible with more rudimentary MIDI (or equivalent) data acquisition. Yet further subtleties of touch, articulation, and pedaling could only be revealed through continuous, high-accuracy key/pedal position tracking data. Although it cannot be determined from this study whether such performance subtleties bear a direct influence on sound production at the piano, they remain inherent to the process of piano playing, and (at least) indirectly involved (through mechanical, physiological, or kinaesthetic functions) in the actual sound production. These subtle control features may thus be considered as valid and valuable descriptors of individuality in piano performance.

In a broad characterization, the individual playing styles of the four pianists show the following salient traits:

- Pianist A's performances had characteristically the lowest dynamics (and high dynamic variability), the longest attacks, the least percussive touch, the longest melody leads (voice accentuation), and the most *legato* articulation—perhaps typical of a French playing style.
- Pianist B favored a very *staccato* articulation, short key sustains despite his playing at the slowest average tempo, and the most percussive touch. This playing style may be related to his upbringing in an Italian piano school that promotes detached playing.
- Pianist C's playing was essentially characterized by the highest (and most steady) intensity, the fastest attacks (yet not as percussive and barely the shortest), and the deepest key depressions. His articulation remained essentially *non-legato*, but with short key releases. He also made a personal choice in never using the soft pedal.
- Lastly, Pianist D's playing was mostly idiosyncratic in his heavy use of both pedals. His playing was also marked by a heavy dynamic emphasis of the right hand, and by generally shallow key depressions. His articulation remained *non-legato*, despite the longest key depression sustains.

Furthermore, with regard to the main hypothesis explored in this study, it was found that pianists' individuality expressed itself differently depending of the timbral nuance performed—within the general frame of their overall performance individuality. In other words, in addition to the general differences between the four pianists and the differences between each timbral nuance (common to all four pianists), the pianists also used some different performance strategies in order to highlight each timbral nuance.

Indeed, amid significant individual differences overall, dynamics were also changed differently by the four pianists in performing different timbral nuances (most saliently for dark-timbre performances, as only Pianist D did not lower his dynamics). Likewise, dynamic variations and balance between hands, as well as attack speed and depth, were altered differently by each pianist between timbral nuances. Articulation was also changed differently by each pianist between timbral nuances, especially in dry-timbre performances (in comparison with the four other timbres), where the tendency toward more *staccato* playing was followed to quite different degrees by each pianist. Finally, the use of the soft pedal was also specific to certain combinations of pianist and timbre.

On the other hand, some performance patterns that are essentially common to all four pianists in the production of different timbral nuances bear some similarity to those highlighted in the expression of different emotions. Given the resembling patterns of intensity and articulation between the strategies of timbre production and those of emotional expression, we may infer that the descriptors of piano timbre may possess some degree of correspondence with verbal descriptors of basic emotions. Velvety and dark timbres may thus be related to sad or tender emotions (low intensity, *legato* articulation), while a dry timbre may reflect happiness (high intensity, *staccato* articulation). This assumption may be supported by comments from two of the four participant pianists, who mentioned that, although undeniably valid and relevant as timbre descriptors, some of the five terms used (especially dry and dark) can actually double as descriptors of musical character—thus closer to reflecting an actual emotional imprint. However, this suggested correspondence between piano performances guided by descriptors of timbre and of emotion only relies on limited (MIDI-accessible) performance parameters. These are far from encompassing the more elaborated descriptions of the performances of different timbral nuances, which may pose the problem of a selection bias in picking the parameters to compare, and leaves several different ways in which the production of timbral nuances may stray from their emotional counterparts.

In conclusion, a novel approach was employed, in which a particular attention was given to respecting a valid musical context, with cohesive, original musical pieces composed for the study, and with timbral expression guided by verbal descriptors shown to the lexicon of piano timbre nuances of common, consensual and meaningful use among pianists. The first hypothesis that pianists' individuality could be revealed in subtle performance control features was confirmed. Moreover, the patterns of individual strategies and differences between the four pianists were found to differ, within the general frame of each pianist's individuality, between the production of different timbral nuances, thus confirming the second hypothesis.

4.1. STATISTICAL VALIDITY OF THE FINDINGS

Some choices were made in the statistical analysis process. First, the significance levels of the ANOVAs were not adjusted for multiple comparisons, despite the 616 dependent variables tested overall and the 149 dependent variables tested per timbral nuance. Indeed, the corrections for multiple comparisons such as Bonferroni-Dunn or Holm-Bonferroni tend to be too conservative, especially when the assumption of independence between dependent variables is not met (as is the case in this study). Consequently, the risk of type II errors was reduced by not using a correction for multiple comparisons. However, we acknowledge the increased possibility of type I errors. Yet the analytic procedures used subsequently were designed so as to significantly reduce (if not eliminate) the risk of misinterpreting the results due to type I errors. Indeed, the contribution of falsely significant features to Principal Component Analysis can be considered as noise (i.e., more or less uniformly distributed between performers), and do not affect the characteristics of pianists' individuality in the corresponding performance spaces—only their legibility. Moreover, the descriptive performance portraits were obtained after careful selection of only one feature within each cluster of highly-correlated features. As the effect size, significance level, and corresponding statistical power in highlighting pianists' individuality were all considered for selecting only the best such feature per cluster, it is improbable that falsely significant features may appear in the descriptive portraits. Moreover, for some clusters where no variable could meet minimum thresholds for both effect size and statistical power ($\eta^2 < 0.2$, $\pi < 0.2$), no feature was selected.

Furthermore, we opted to perform separate ANOVAs for each timbral nuance, instead of relying on pairwise *post-hoc* tests of the pianist-timbre interaction in the general ANOVAs, as the latter can be too conservative, and especially less informative, as they do not distinguish the two-dimensional structure of the interaction effect, i.e, treat all pairs of interaction cases equally (disregarding whether they correspond to the same pianist or timbre). The solution we privileged thus allowed for a more exhaustive exploration and account of the pianist-timbre interaction (i.e., the expression of pianists' individuality in the production of different timbral nuances.

4.2. PERSPECTIVES

The results presented in this article can be set in perspective with the common performance control strategies adopted by the four pianists in order to produce and express the five different timbral nuances, which were explored in Bernays and Traube (2013).

On the other hand, it cannot be determined from this study whether the observed differences in individual performance strategies between the five timbral nuances stem from a different understanding among the four pianists of the timbre descriptors used as performance instructions, or characterize different ways of reaching a common timbral idea. However, the perception and identification of timbre in the audio recordings of the performances analyzed in this article has been investigated (Bernays, 2013). Preliminary results suggest that the different timbral nuances expressed in the performances can be reliably identified (significantly above chance level) by other pianists, which would indicate that, for each of the five timbral nuances, the differences in performance strategy between the four performers may have yielded, despite possible audible differences in sound production, the same perceptual effect as regard timbre identification.

A parallel may be drawn with vocal expression, where nonverbal affective cues remain consistently identifiable between different speakers (even in single vowels), despite large acoustic inter-individual differences, especially in voice quality (i.e., timbre) (Juslin and Scherer, 2005). Likewise, despite inter-individual differences, the timbral nuances expressed at the piano may remain identifiable, and may be categorized according to corresponding adjectival descriptors.

Furthermore, the four musical pieces used in this study were expressly chosen for their different musical characteristics and the different playing styles they would require. Consequently, the performances of all four pieces were analyzed conjointly, with the statistical the effect of the musical piece performed separated from the effects of performer and timbre, in the aim of highlighting individuality in piano performance over an extended, representative set of musical characteristics. It may be worthwhile then to further disentangle the individual playing styles of pianists from the influence of pieces and interpretive goals, as was accomplished by Gingras et al. (2013) in the context of harpsichord performance, by likewise using the analysis methods of complete linear mixed models and correlation-based similarity profiles. The relations between pianists' individuality (as revealed by characteristic performance features) and the musical structure of each piece performed may also be investigated with the software tools included in the PianoTouch toolbox (Bernays and Traube, 2012), in order to assess pianists' individuality as a function of the different musical contexts featured in the pieces.

Finally, the effect of the performance features characteristic of pianists' individuality upon the actual sound production may be explored with acoustical analyses of the audio recordings of the performances.

In the end, this research may help determine the expressive boundaries of individual piano performance within which the same timbral nuance can be produced and perceived.

5. FUNDING

FRQSC (*Fonds de Recherche du Québec – Société et Culture*); CIRMMT (Centre for Interdisciplinary Research in Music Media and Technology); OICRM (*Observatoire Interdisciplinaire de Création et de Recherche en Musique*).

ACKNOWLEDGMENTS

We wish to thank the four pianists who took part in this study, and the composers of the pieces, as well as Bernard Bouchard for his invaluable technical and pianistic help.

REFERENCES

Bellemare, M., and Traube, C. (2005). "Verbal description of piano timbre: exploring performer-dependent dimensions," in *Digital Proceedings of the 2nd Conference on Interdisciplinary Musicology (CIM05)*, (Montreal, QC: Observatoire interdisciplinaire de création et de recherche en musique (OICRM)).

Ben-Asher, M. (2013). *Towards an Emotionally Intelligent Piano: Real-Time Emotion Detection and Performer Feedback via Kinesthetic Sensing in Piano Performance,* Master's thesis, University of Miami, FL.

Bernays, M. (2012). Expression et production du timbre au piano selon les traités: conception du timbre instrumental exprimée par les pianistes et professeurs dans les ouvrages à vocation pédagogique. *Recherche en éducation Musicale* 29, 7–27. Available online at: http://www.mus.ulaval.ca/reem/REEM_29.pdf#page=17

Bernays, M. (2013). *The Expression and Production of Piano Timbre: Gestural Control and Technique, Perception and Verbalisation in the Context of Piano Performance and Practice,* Doctoral dissertation, Université de Montréal, Canada.

Bernays, M., and Traube, C. (2011). "Verbal expression of piano timbre: multidimensional semantic space of adjectival descriptors," in *Proceedings of the International Symposium on Performance Science (ISPS2011)*, eds A. Williamon, D. Edwards and L. Bartel (Utrecht, Netherlands: European Association of Conservatoires (AEC)), 299–304.

Bernays, M., and Traube, C. (2012). "Piano touch analysis: A MATLAB toolbox for extracting performance descriptors from high-resolution keyboard and pedalling data," in *Proceedings of Journées d'Informatique Musicale (JIM2012), Gestes, Virtuosité et Nouveaux Medias*, eds T. Dutoit, T. Todoroff and N. d'Alessandro (Mons, Belgium: UMONS/numediart), 55–64.

Bernays, M., and Traube, C. (2013). "Expressive production of piano timbre: touch and playing techniques for timbre control in piano performance," in *Proceedings of the 10th Sound and Music Computing Conference (SMC2013)*, ed R. Bresin (Stockholm, Sweden: KTH Royal Institute of Technology), 341–346.

Bhatara, A., Tirovolas, A. K., Duan, L. M., Levy, B., and Levitin, D. J. (2011). Perception of emotional expression in musical performance. *J. Exp. Psychol. Hum. Percept. Perform.* 37, 921–934. doi: 10.1037/a0021922

Bresin, R., and Battel, G. U. (2000). Articulation strategies in expressive piano performance. *J. New Music Res.* 29, 211–224. doi: 10.1076/jnmr.29.3.211.3092

Cadoz, C., and Wanderley, M. M. (2000). "Gesture–music," in *Trends in Gestural Control of Music*, eds M. M. Wanderley and M. Battier (Paris: IRCAM), 71–94.

Canazza, S., De Poli, G., Rinaldin, S., and Vidolin, A. (1997). Sonological analysis of clarinet expressivity. *Music Gestalt Comput. Stud. Cogn. Syst. Musicol.* 1317, 431–440. doi: 10.1007/BFb0034131

Cheminée, P. (2006). "Vous avez dit 'clair'?" Le lexique des pianistes, entre sens commun et terminologie. *Cahiers du LCPE: Dénomination, désignation et catégories* 7, 39–54. Available online at: http://www.lam.jussieu.fr/Publications/CahiersLCPE/cahier7.pdf#page=39

Dalla Bella, S., and Palmer, C. (2011). Rate effects on timing, key velocity, and finger kinematics in piano performance. *PLoS ONE* 6:e20518. doi: 10.1371/journal.pone.0020518

De Poli, G. (2004). Methodologies for expressiveness modelling of and for music performance. *J. New Music Res.* 33, 189–202. doi: 10.1080/0929821042000317796

Dixon, S. (2001). Automatic extraction of tempo and beat from expressive performances. *J. New Music Res.* 30, 39–58. doi: 10.1076/jnmr.30.1.39.7119

Ekman, P. (1992). An argument for basic emotions. *Cogn. Emot.* 6, 169–200. doi: 10.1080/02699939208411068

Faure, A. (2000). *Des Sons Aux Mots, Comment Parle-t-on du Timbre Musical?,* Doctoral dissertation, Ecole des Hautes Etudes en Sciences Sociales, Paris, France.

Fink, S. (1992). *Mastering Piano Technique: A guide for students, teachers, and performers*, Portland, OR: Amadeus Press.

Friberg, A., Bresin, R., and Sundberg, J. (2006) Overview of the KTH rule system for musical performance. *Adv. Cogn. Psychol.* 2, 145–161. doi: 10.2478/v10053-008-0052-x

Gabrielsson, A., and Juslin, P. N. (1996) Emotional expression in music performance: Between the performer's intention and the listener's experience, *Psychol. Music* 24, 68–91. doi: 10.1177/0305735696241007

Gingras B., Asselin P.-Y., and McAdams, S. (2013). Individuality in harpsichord performance: disentangling performer- and piece-specific influences on interpretive choices, *Front. Psychol.* 4:895. doi: 10.3389/fpsyg.2013.00895

Gingras, B., Lagrandeur-Ponce, T., Giordano, B. L., and McAdams, S. (2011). Perceiving musical individuality: performer identification is dependent on performer expertise and expressiveness, but not on listener expertise. *Perception* 40, 1206–1220. doi: 10.1068/p6891

Goebl, W. (2001). Melody lead in piano performance: expressive device or artifact?, *J. Acoust. Soc. Am.* 110, 563–572. doi: 10.1121/1.1376133

Goebl, W., Bresin, R., and Galembo, A. (2005). Touch and temporal behavior of grand piano actions. *J. Acoust. Soc. Am.* 118, 1154–1165. doi: 10.1121/1.1944648

Goebl, W., Dixon, S., DePoli, G., Friberg, A., Bresin, R., and Widmer, G. (2008). "'Sense' in expressive music performance: Data acquisition, computational studies, and models," in *Sound to Sense - Sense to Sound: A State of the Art in Sound and Music Computing*, eds P. Polotti and D. Rocchesso (Berlin: Logos), 195–242.

Goebl, W., and Fujinaga, I. (2008) "Do key-bottom sounds distinguish piano tones?," in *Proceedings of the 10th International Conference on Music Perception and Cognition (ICMPC 10)*, (Sapporo: Hokkaido University), 292.

Hart, H. C., Fuller, M. W., and Lusby, W. S. (1934). A precision study of piano touch and tone. *J. Acoust. Soc. Am.* VI, 80–94. doi: 10.1121/1.1915706

Heinlein, C. P. (1929). A discussion of the nature of pianoforte damper-pedalling together with an experimental study of some individual differences in pedal performance. *J. Gen. Psychol.* 2, 489–508. doi: 10.1080/00221309.1929.9918087

Hevner, K. (1936). Experimental studies of the elements of expression in music. *Am. J. Psychol.* 48, 246–268. doi: 10.2307/1415746

Hofmann, J. (1920). *Piano Playing with Piano Questions answered*. Philadelphia, PA: Theodore Presser Co.

Holmes, P. A. (2012). An exploration of musical communication through expressive use of timbre: The performer's perspective. *Psychol. Music* 40, 301–323. doi: 10.1177/0305735610388898

Juslin, P. N. (1997). Emotional communication in music performance: a functionalist perspective and some data. *Music Percept.* 14, 383–418. doi: 10.2307/40285731

Juslin, P. N., and Laukka, P. (2003). Communication of emotions in vocal expression and music performance: different channels, same code?, *Psychol. Bull.* 129, 770–814. doi: 10.1037/0033-2909.129.5.770

Juslin, P. N., and Scherer, K. R. (2005). "Vocal expression of affect," in *The New Handbook of Methods in Nonverbal Behavior Research*, eds J. A. Harrigan, R. Rosenthal, and K. R. Scherer (New York, NY: Oxford University Press), 65–135.

Kaiser, H. F. (1958). The varimax criterion for analytic rotation in factor analysis. *Psychometrika* 23, 187–200. doi: 10.1007/BF02289233

Kochevitsky, G. (1967). *The art of piano playing: a scientific approach*. (Secaucus, NJ: Sunny-Brichard).

Lourenço, S. (2010). European Piano Schools: Russian, German and French classical piano interpretation and technique. *J. Sci. Technol. Arts* 2, 6–14. doi: 10.7559/citarj.v2i1.7

MacRitchie, J. (2011). *Elucidating Musical Structure through Empirical Measurement of Performance Parameters*, Doctoral dissertation, University of Glasgow, UK.

MacRitchie, J., Buck, B., and Bailey, N. J. (2013). Inferring musical structure through bodily gestures. *Musicae Scientiae* 17, 86–108. doi: 10.1177/1029864912467632

Madison, G. (2000). Properties of expressive variability patterns in music performances. *J. New Music Res.* 29, 335–356. doi: 10.1080/09298210008565466

McPherson, A., and Kim, Y. (2011). "Multidimensional gesture sensing at the piano keyboard," in *CHI '11: Proceedings of the 2011 annual conference on Human factors in computing systems (Vancouver, BC)*, New York, NY: ACM, 2789–2798. doi: 10.1145/1978942.1979355

Neuhaus, H. (1973). *The Art of Piano Playing*, London, UK: Barrie and Jenkins (Translated from Russian by K. A. Leibovitch).

Ortmann, O. R. (1929). *The Physiological Mechanics of Piano Technique: An Experimental Study of the Nature of Muscular Action as Used in Piano Playing and of the Effects Thereof Upon the Piano Key and the Piano Tone*, New York, NY: E.P. Dutton.

Palmer, C. (1996). On the assignment of structure in music performance. *Music Percept.* 14, 23–56. doi: 10.2307/40285708

Parncutt, R. (2003). "Accents and expression in piano performance," in *Perspektiven und Methoden einer Systemischen Musikwissenschaft*, eds K. W. Niemöller and B. Gätjen (Frankfurt: Peter Lang), 163–185.

Parncutt, R., and Troup, M. (2002) "Piano," in *The Science and Psychology of Music Performance: Creating Strategies for Teaching and Learning*, eds R. Parncutt, and G. McPherson (New York, NY: Oxford University Press), 285–302. doi: 10.1093/acprof:oso/9780195138108.001.0001

Quinto, L., Thompson, W. F., and Taylor, A. (2013). The contributions of compositional structure and performance expression to the communication of emotion in music. *Psychol. Music*. doi: 10.1177/0305735613482023. [Epub ahead of print].

Repp, B. H. (1990). Patterns of expressive timing in performances of a Beethoven minuet by nineteen famous pianists. *J. Acoust. Soc. Am.* 88, 622–641.

Repp, B. H. (1992) Diversity and commonality in music performance: an analysis of timing microstructure in Schumanns "Träumerei." *J. Acoust. Soc. Am.* 92, 2546–2568. doi: 10.1121/1.399766

Repp, B. H. (1996a) The dynamics of expressive piano performance: Schumann's "Träumerei" revisited. *J. Acoust. Soc. Am.* 100, 641–650. doi: 10.1121/1.415889

Repp, B. H. (1996b). Patterns of note onset asynchronies in expressive piano performance. *J. Acoust. Soc. Am.* 100, 3917–3932. doi: 10.1121/1.417245

Repp, B. H. (1996c). Pedal timing and tempo in expressive piano performance: a preliminary investigation. *Psychol. Music* 24, 191–221. doi: 10.1177/0305735696242011

Repp, B. H. (1997). Expressive timing in a Debussy Prelude: a comparison of student and expert pianists. *Musicae Scientiae* 1, 257–268.

Repp, B. H. (1998). A microcosm of musical expression. I. Quantitative analysis of pianists' timing in the initial measures of Chopin's Etude in E major. *J. Acoust. Soc. Am.* 104, 1085–1100. doi: 10.1121/1.423325

Rink, J., Spiro, N., and Gold, N. (2011). "Motive, gesture, and the analysis of performance," in *New Perspectives on Music and Gesture*, eds A. Gritten and E. King (Aldershot: Ashgate Publishing), 267–292.

Saunders, C., Hardoon, D. R., Shawe-Taylor, J., and Widmer, G. (2008). Using string kernels to identify famous performers from their playing style. *Intel. Data Anal.* 12, 425–440. Available online at: http://iospress.metapress.com/content/p76205178g37346x/

Shaffer, L. H. (1995). Musical performance as interpretation. *Psychol. Music* 23, 17–38. doi: 10.1177/0305735695231002

Sloboda, J. A. (2000). Individual differences in music performance. *Trends Cogn. Sci.* 4, 397–403. doi: 10.1016/S1364-6613(00)01531-X

Stamatatos, E., and Widmer, G. (2005). Automatic identification of music performers with learning ensembles. *Artif. Intel.* 165, 37–56. doi: 10.1016/j.artint.2005.01.007

Sándor, G. (1995). *On Piano Playing*, New York, NY: Schirmer Books.

Todd, N. P. M. (1992). The dynamics of dynamics: a model of musical expression. *J. Acous. Soc. Am.* 91, 3540–3550. doi: 10.1121/1.402843

Van Vugt, F. T., Jabusch, H. C., and Altenmller, E. (2013). Individuality that is unheard of: systematic temporal deviations in scale playing leave an inaudible pianistic fingerprint. *Front. Psychol.* 4:134. doi: 10.3389/fpsyg.2013.00134

Wang, C. I. (2013). *Quantifying pianist style – An investigation of performer space and expressive gestures from audio recordings*. Doctoral dissertation, New York University.

Widmer, G., Flossmann, S., and Grachten, M. (2009). YQX plays Chopin, *AI Magazine* 30, 35–48.

Widmer, G., and Goebl, W. (2004). Computational models of expressive music performance: the state of the art. *J. New Music Res.* 33, 203–216. doi: 10.1080/0929821042000317804

Woody, R. H. (1999). The relationship between explicit planning and expressive performance of dynamic variations in an aural modeling task. *J. Res. Music Edu.* 4, 331–342. doi: 10.2307/3345488

Woody, R. H. (2002). Emotion, imagery and metaphor in the acquisition of musical performance skill. *Music Edu. Res.* 4, 213–224. doi: 10.1080/1461380022000011920

Wöllner, C. (2013). How to quantify individuality in music performance? studying artistic expression with averaging procedures. *Front. Psychol.* 4:361. doi: 10.3389/fpsyg.2013.00361

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 31 August 2013; accepted: 08 February 2014; published online: 04 March 2014.

Citation: Bernays M and Traube C (2014) Investigating pianists' individuality in the performance of five timbral nuances through patterns of articulation, touch, dynamics, and pedaling. Front. Psychol. ***5****:157. doi: 10.3389/fpsyg.2014.00157*

This article was submitted to Cognitive Science, a section of the journal Frontiers in Psychology.

frontiers in
PSYCHOLOGY

ORIGINAL RESEARCH ARTICLE
published: 22 November 2013
doi: 10.3389/fpsyg.2013.00853

Getting into the musical zone: trait emotional intelligence and amount of practice predict flow in pianists

Manuela M. Marin[1]* and Joydeep Bhattacharya[2]

[1] Department of Basic Psychological Research and Research Methods, University of Vienna, Vienna, Austria
[2] Department of Psychology, Goldsmiths, University of London, London, UK

Edited by:
Narayanan Srinivasan, University of Allahabad, India

Reviewed by:
Massimiliano Palmiero, University of L'Aquila, Italy
Barbara Colombo, Catholic University of the Sacred Heart, Italy
Saskia Jaarsveld, University of Kaiserslautern, Germany

***Correspondence:**
Manuela M. Marin, Department of Basic Psychological Research and Research Methods, University of Vienna, Liebiggasse 5, Vienna, A-1010, Austria
e-mail: manuela.marin@univie.ac.at

Being "in flow" or "in the zone" is defined as an extremely focused state of consciousness which occurs during intense engagement in an activity. In general, flow has been linked to peak performances (high achievement) and feelings of intense pleasure and happiness. However, empirical research on flow in music performance is scarce, although it may offer novel insights into the question of why musicians engage in musical activities for extensive periods of time. Here, we focused on individual differences in a group of 76 piano performance students and assessed their flow experience in piano performance as well as their trait emotional intelligence. Multiple regression analysis revealed that flow was predicted by the amount of daily practice and trait emotional intelligence. Other background variables (gender, age, duration of piano training and age of first piano training) were not predictive. To predict high achievement in piano performance (i.e., winning a prize in a piano competition), a seven-predictor logistic regression model was fitted to the data, and we found that the odds of winning a prize in a piano competition were predicted by the amount of daily practice and the age at which piano training began. Interestingly, a positive relationship between flow and high achievement was not supported. Further, we explored the role of musical emotions and musical styles in the induction of flow by a self-developed questionnaire. Results suggest that besides individual differences among pianists, specific structural and compositional features of musical pieces and related emotional expressions may facilitate flow experiences. Altogether, these findings highlight the role of emotion in the experience of flow during music performance and call for further experiments addressing emotion in relation to the performer and the music alike.

Keywords: optimal experience, altered states of consciousness, music performance, autotelic personality, emotion

INTRODUCTION

Professional musicians often spend many months on practicing a musical piece, aiming at mastering its technical and interpretative challenges in order to prepare for a perfect performance in front of an audience. One possible explanation for performers' motivation to take on such intense musical practice on a daily basis for many years is the experience of flow (Csikszentmihalyi, 2002; Lamont, 2009; Custodero, 2012). Flow, or optimal experience, can be broadly defined as a psychological state involving the positive experience of being fully engaged in the successful pursuit of an activity (Csikszentmihalyi, 1990), and due to its intrinsically rewarding nature, flow seems to motivate humans to keep returning to the flow-inducing action and meeting greater challenges. Csikszentmihalyi (1990) developed a nine-dimensional flow construct. Based on these dimensions, flow is characterized by

> challenge-skill balance (feeling competent enough to meet the high demands of the situation), action-awareness merging (doing things spontaneously and automatically without having to think), clear goals (having a strong sense of what one wants to do), unambiguous feedback (knowing how well one is doing during the performance itself), concentration on the task at hand (being completely focused on the task at hand), sense of control (having a feeling of total control over what one is doing), loss of self-consciousness (not worrying what others think of oneself), transformation of time (having the sense that time passes in a way that is different from normal), and autotelic experience (feeling the experience to be extremely rewarding). (Martin and Jackson, 2008, p. 146)

The construct of flow is conceptually similar to the construct of peak experience and peak performance, psychological states characterized by intense positive feelings and personal fulfillment (Maslow, 1968). These two constructs share many qualities with each other (Privette, 1983; Privette and Bundrick, 1991), and in fact, flow has been shown to be related to peak performances and high achievements across disciplines ranging from sports (e.g., Jackson, 1999), music performance (O'Neill, 1999; Sawyer, 2006; but also see Wrigley and Emmerson, 2013), to compositional creativity and meaningfulness (MacDonald et al., 2006; Baker and MacDonald, 2013). Here, we investigate individual differences among pianists with regard to flow experiences and their relationship to trait emotional intelligence and peak performance, respectively.

Most activities can be flow-inducing, irrespective of whether they are work or leisure-based (Csikzentmihalyi and

Csikzentmihalyi, 1988). While there are considerable variations in the flow-inducing tasks and contextual settings, the flow experience itself is surprisingly similar across a range of demographic variables like culture, ethnicity, socioeconomic background and age (Csikzentmihalyi and Csikzentmihalyi, 1988; Clarke and Haworth, 1994; Moneta, 2004; Asakawa, 2010). However, large individual differences do exist in the characteristics of flow experience like its frequency and strength. Csikszentmihalyi (1990) has already proposed that certain personality traits, such as curiosity, persistence and low self-centeredness, may be characteristics of people who can easily achieve flow states. These personality traits may constitute what is known as an "autotelic personality" (Nakamura and Csikszentmihalyi, 2009). Autotelic persons look for more challenges (Logan, 1988), are less anxious, more motivated, and show higher 'playfulness' (Tan and Chou, 2010). However, we must point out here that not much is known about what constitutes an autotelic personality (Busch et al., 2013), and in fact, its existence is yet to be supported by substantial empirical evidence.

The limited number of studies on individual differences and dispositional flow vary in terms of types of individual differences and activities under investigation. For instance, the balance between skills and demands required by an activity only induced flow in those individuals that were characterized by an internal locus of control (Keller and Blomann, 2008; Mosing et al., 2012b). Locus of control is a personality construct (Rotter, 1966) that refers to people's beliefs regarding the action-outcome relationship. An internal locus of control is characterized by the belief that outcomes depend on controllable factors, such as attitude, preparation and effort, whereas an external locus of control is reflected in the belief that an outcome depends on the environment, luck and knowing the right people (Rotter, 1966; Levenson, 1981; Lefcourt, 1991). Since associations with happiness (e.g., Larson, 1989; DeNeve and Cooper, 1998) as well as with positive mental health (Naditch et al., 1975; Presson and Benassi, 1996) have been reported in individuals with a high internal locus of control, the finding by Keller and Blomann (2008) suggests that the underlying mechanism explaining this relationship may be rooted in varying degrees of sensitivity to the level of control during the pursuit of an activity. Individuals high in internal locus of control may enjoy the activity more when facing challenges and thus enter into flow states more easily.

Need for achievement was also identified as a personal characteristic that fosters flow experience through challenge-skill balance (Eisenberger et al., 2005). More generally, positive relationships between flow and mental toughness in sports (Crust and Swann, 2013), with personality traits reflecting a high need to learn and low need for activity in videogaming (Seger and Potts, 2012), as well as with self-control (Kuhnle et al., 2012), novelty seeking and persistence (Teng, 2011) were reported, respectively. A negative relationship between flow proneness and neuroticism was found with regard to activities in everyday life (Ullen et al., 2012). Furthermore, flow proneness and intelligence were not associated in a study involving adult twin pairs (Ullen et al., 2012).

Given the widely-studied relationship between flow, motivation, high achievement and individual differences in sports (for a review see Swann et al., 2012) and work-related activities (e.g., Csikszentmihalyi and LeFevre, 1989; Eisenberger et al., 2005; Nakamura and Csikszentmihalyi, 2009), it is surprising to note that the investigation of flow in music performance and composition has received comparatively little attention since Csikszentmihalyi's introduction of the flow concept (1975). One of the first studies on flow with regard to music was conducted by O'Neill (1999), who investigated the development of performance skills in adolescent musicians and their relation to flow by using the Experience Sampling Method. She found a positive relationship between high achievement in music performance and the number of experienced flow states. Custodero (2005) provided evidence for the existence of flow-like states in infants and children by investigating different musical learning environments. Moreover, MacDonald et al. (2006) revealed a positive relationship between creativity, flow and the quality of group compositions in university students. In a similar vein, the degree of flow experiences has been found to be positively correlated with the meaningfulness of songs created during therapeutic songwriting (Baker and MacDonald, 2013). In a longitudinal study, Fullagar et al. (2013) showed that high degrees of flow were accompanied with low experiences of performance anxiety in music performance students. The emergence of flow has also been examined in the context of choir singing and conducting (Custodero, 2002; Bloom and Skutnick-Henley, 2005; Freer, 2009) as well as in ensemble playing (Kraus, 2003; Sawyer, 2006).

More recently, researchers began to explore the psychophysiological underpinnings of flow states in pianists, revealing that flow is associated with decreased heart period, blood pressure and heart rate variability as well as with increased activity of the zygomaticus major muscle and respiratory depth (De Manzano et al., 2010). Thomson and Jaque (2011) investigated psychophysiological responses during flow states in performing musicians and found decreased cardiac autonomic balance and regulatory capacity.

Flow can also be experienced during music listening. For instance, Lamont (2009) discussed flow in the context of students' self-reported peak experiences in music listening and performance, exploring the ways in which music can lead to happiness. Similarly, Diaz (2011) was interested in investigating flow and its relation to mindfulness in the context of music listening and found that there may be a phenomenological difference between flow and aesthetic response. Flow was also associated with music listening in sports and exercise in elite-athletes (Laukka and Quick, 2011). Other recent studies addressed the issue of how to assess flow experiences in musicians (Martin and Jackson, 2008; Sinnamon et al., 2012; Wrigley and Emmerson, 2013), suggesting that flow scales developed for fields other than music, such as the flow scales by Jackson and Marsh (1996) and by Jackson and Eklund (2002), are reliable tools applicable to the domain of music. Taken together, research on musical flow offers valuable insights into questions relevant for psychologists, teachers, therapists and performers/composers alike.

Research on musicians' personalities has largely focused on differences between musicians of various instrument groups and musical styles (e.g., Buttsworth and Smith, 1995; Kemp, 1996; Cribb and Gregory, 1999; Langendörfer, 2008; Hernandez et al., 2009; Vuust et al., 2010), the relationship between personality and performance anxiety (e.g., Cooper and Wills, 1989; Marchant-Haycox and Wilson, 1992; Kenny et al., 2004), as well as the relationship between personality and creativity (e.g., Gibson et al., 2009; Charyton and Snelbecker, 2010). To our knowledge, studies focusing primarily on general aspects of personality (i.e., not music-specific traits such as performance anxiety) and their relation to flow experiences in musicians have not been conducted so far. Specifically, we explored whether there is something inherent in the personality of pianists (see e.g., Chmurzynska, 2012) that could help understand why some pianists reach flow states more often and easily compared to others.

Wrigley and Emmerson (2013) found in their study that flow experiences may depend on the family of specific instrument(s). Their results indicated that piano players reported lower levels of flow on average compared to brass and string players. Due to these differences of flow experiences observed in different instrument groups, we consider it as appropriate to focus our research on pianists. Furthermore, the piano is a common instrument in Western populations and also widely studied within the field of music performance research.

Flow is usually related to peak performances and subsequent happiness, and is as such a highly emotional experience. However, in music research, only a few studies have addressed flow theory in relation to emotion, which is rather surprising because emotions play a crucial role in musical communication (Juslin and Sloboda, 2010), possibly constituting an essential difference between the domains of sports and music (Sinnamon et al., 2012). For example, Bakker (2005) found support for the cross-over of flow between music teachers and their students by building on emotional contagion theory (Hatfield et al., 1994). The degree of flow experienced in teachers was correlated with the teachers' intrinsic work motivation, which was positively associated with the degree of flow experienced among their students.

Fritz and Avsec (2007) reported that positive emotional aspects of subjective well-being and dispositional aspects of flow were positively correlated in music students, whereas flow and satisfaction with life were less strongly correlated. The authors thus concluded that flow is more related to affective than cognitive aspects of subjective well-being. However, a great deal remains unknown about flow and its relation to emotion in music performance research. It can be surmised that the experience of flow during musical activities may depend on at least two "affective factors": first, on the musical piece and its emotional characteristics, and second, on the performer's personality and his or her emotional intelligence. Both of these aspects were investigated in the current study by using self-report measures.

Emotional intelligence can be generally defined as "[…] the ability to process emotion-laden information competently and to use it to guide cognitive activities like problem solving and to focus energy on required behaviors" (Salovey et al., 2009). However, it is now standard in the field to differentiate between two constructs, namely trait emotional intelligence and ability emotional intelligence (for a review see Petrides, 2011). Trait emotional intelligence is measured via self-report and conceptualized as a personality trait, whereas ability emotional intelligence refers to emotion-related cognitive abilities and is measured via maximum-performance tests. Since the present study focuses on flow theory and emotion, trait emotional intelligence was chosen as a possible facet of a pianist's personality that may be associated with flow. This approach was also motivated by the fact that a positive relationship between trait emotional intelligence and length of musical training was reported earlier for a group of music students (Petrides et al., 2006). The primary goal of our study was to investigate whether trait emotional intelligence could predict dispositional flow in pianists. We hypothesized that a higher degree of trait emotional intelligence would be associated with a higher degree of dispositional flow.

Since a relationship between flow and superior performances and achievement was previously found by others (O'Neill, 1999; MacDonald et al., 2006; Baker and MacDonald, 2013), another goal of our study concerned the modeling of high achievement in piano performance as measured by having won prizes at piano competitions. It seemed plausible to assume that the experience of flow in piano performance would predict success in piano competitions. Finally, a set of self-developed questions explored whether there are specific characteristics of a musical piece that may induce flow states more easily than others. In particular, exploratory questions referred to emotions expressed and induced by the music, the musical style since emotional communication is associated with certain musical styles more than with others (Kallinen, 2005; Robinson, 2005), and also to the compositional style. If emotional intelligence and dispositional flow were related in the performer, these additional questions would help better understand the underlying mechanisms between flow and emotion in music performance and initiate future experiments.

METHODS AND MATERIALS

PARTICIPANTS

Participants were 76 piano performance students (including 45 females) who, at the time of this study, were pursuing a professional career as a musician. Seventy-three students were enrolled in a classical performance degree and three in a Jazz performance degree at English-speaking institutions of higher education (university or music conservatory). Fifty-six students were undergraduates. The participants had the following nationalities: UK ($n = 29$), US ($n = 24$), Australia ($n = 17$) and Canada ($n = 6$). The mean age was 21.7 years ($SD = 3.7$). Our participants started their piano training on average at 6.8 years ($SD = 2.8$) of age, played the piano as a first instrument for 14.0 years ($SD = 5.0$), and practiced on average 3.3 h a day ($SD = 2.1$) at the time of the study. Forty-five participants had previously won at least one prize in a piano competition. Thirty-seven participants indicated that they preferred to play the piano alone rather than together with others. Participants also estimated to improvise on average 1.8 h per week ($SD = 3.2$) on the piano. Twenty-one students reported regularly playing other instruments besides the piano. Our participants were thus considered to have a high degree of musical training and involvement with music at the time of the study.

MATERIALS

The questionnaire comprised two standardized tests, one on dispositional flow and one on trait emotional intelligence, as well as two self-developed questionnaires, one on flow and musical characteristics referring to emotion and musical style and one on the socio-demographic and musical backgrounds (musical training, musical preference, amount of practice etc.). The order of the administration of these separate questionnaires remained the same across all participants and was as follows: socio-demographic and musical background, flow scale, self-developed questions on flow and musical characteristics, and trait emotional intelligence scale.

The Dispositional Flow Scale-2 (DFS-2) (Jackson and Eklund, 2004) comprises 36 items referring to the nine-dimensional nature of flow and has been reliably applied (Cronbach alpha = 0.92) to assess flow in music performance (Sinnamon et al., 2012). Answers are collected on 5-point scales (1 = never to 5 = always) and require specific instructions depending on the activity under investigation. They were as follows: "Please answer the following questions in relation to your experience of practicing/playing a piano solo piece that you know by heart and which could be performed in a concert next week." These concrete instructions were chosen to make participants think of a common and realistic situation of their lives as musicians and to enhance the comparability across their responses. Moreover, earlier research suggested that flow occurs more often at the last stages of practicing a new musical piece (Kraus, 2003). Furthermore, these instructions were considered as appropriate since we also aimed at investigating flow and peak performance.

The short form of the Trait Emotional Intelligence Questionnaire (TEIQue-SF) (Petrides and Furnham, 2006) measures global trait emotional intelligence by collecting responses to 30 items on 7-point scales (1 = completely disagree to 7 = completely agree).

The self-developed questionnaire on flow in the context of musical emotions and musical style involved questions about (i) flow and piano performance, (ii) flow and musical emotions, and (iii) flow and musical styles and composers. Depending on the type of question, answers involved yes/no responses, numeric or verbal responses, or responses on rating scales (either ranging from "yes, agree," "yes, somehow agree," "no, somehow disagree," "no, don't agree" and "don't agree," or from "always," "frequently," "sometimes," "rarely," "never" to "don't know"). The questions were developed by the first author, a musicologist, but also discussed with two professional pianists in order to ensure that all questions were meaningful to musicians and comprehensive. Participants were allowed to skip questions if they preferred not to respond to some of these questions, which was rarely the case.

Specifically, six questions assessed the number of flow states during piano performance and music listening (e.g., Would you agree that flow states in piano performance can only be reached when the piece is nearly ready for a public performance?), the relationship between flow and motivation (Would you agree that the experience of flow keeps you motivated to practice the piano and to become better?), flow and life-satisfaction (Do you experience a high degree of life satisfaction after the experience of flow in piano performance?) as well as the occurrence of flow by defining flow according to Csikszentmihalyi's concept (1990) prior to these questions: "Flow refers to an altered state of consciousness where one becomes so deeply immersed in a task that all else seems to disappear. This state is characterized by total concentration on the task at hand, clear goals, and unselfconscious action. Self-reports of flow include a transformation of our perception of time and self-awareness as well as a sense of fulfillment and feelings of intense happiness after a flow performance, referring to the intrinsically rewarding experience that flow brings to the individual."

Several questions addressed the possible relationship between musical emotions and flow. Two questions probed the relationship between happiness and flow (Do you experience intense happiness and enjoyment WHILE being in a flow state in piano performance? Do you experience intense happiness and enjoyment shortly AFTER the experience of flow in piano performance?), two the general role of musical emotions in flow induction (Musical pieces are expressive of different types of emotions. From your own experience, do you feel that flow states are more easily induced by certain types of emotions expressed by a piece? Musical pieces can induce different types of emotions in you. From your own experience, do you feel that flow states are more easily induced when you feel certain types of emotions while playing a piece?), and two further questions referred to changing emotions in a musical piece (Would you agree that flow states appear less frequently when the emotional content of a piece is varying a lot over the course of a piece?) and general liking for certain emotions and their effect on flow states (Do you feel that flow states are more easily reached when the piece induces emotions in you that you particularly like in general (can be either positive or negative emotions?).

Two questions asked for specific ratings of flow in the context of musical emotions following Russell's circumplex model of affect (1980). The instructions ensured that participants understood the difference between felt and perceived emotions, a crucial distinction with regard to the study of musical emotions (Gabrielsson, 2002). The questions regarding flow in the context of Russell's circumplex model of affect were posed as follows: 1) "Emotions can be described by arousal (calm vs. activated) and pleasantness (pleasant vs. unpleasant). From your experience, please rate how often one of the following emotional states EXPRESSED by a piece has led to flow in piano performance. Note: You did not necessarily feel these emotions yourself while playing a piece," followed by the specific emotions "low-arousing pleasant" "high-arousing pleasant," "low-arousing unpleasant" and "high-arousing unpleasant" and the respective rating scale. The second question referred to felt emotions: "From your experience, please rate how often one of the following emotional states INDUCED by a piece has led to flow in piano performance. Note: These following emotions were not necessarily expressed by a piece, but they describe your feelings while playing a piece," again followed by the specific emotions "low-arousing pleasant," "high-arousing pleasant," "low-arousing unpleasant" and "high-arousing unpleasant" and the respective rating scales.

The last five questions probed whether there was an association between musical styles, musical emotions and flow during piano performance. For example, participants were asked whether they had experienced flow states more often with certain musical styles than with others (Do you feel that you experience flow states more often when playing certain musical styles?), and further, to indicate the musical style that has most frequently induced flow states. Questions regarding the familiarity with and preference of musical styles complemented this section. Finally, participants were asked whether they could name any composers whose pieces had reliably induced flow during piano practice over a long period of time.

PROCEDURE

Music departments and piano professors in the UK, United States, Canada and Australia were contacted via email by the first author and invited to distribute the link to the online questionnaire among their students. This method of recruitment was chosen in order to ensure that participants were in fact piano performance students coming from higher institutions. Answering all the questionnaires took around 34 min on average, and participants could choose to participate in a prize draw. The data was collected between February and November 2010. This study was approved by the local ethics committee of the Department of Psychology at Goldsmiths, University of London, and followed the guidelines of the Declaration of Helsinki.

STATISTICAL ANALYSIS

Statistical analyses were conducted in IBM SPSS Statistics version 19 (SPSS Inc., Chicago, IL, USA) and in Matlab R2010b (The MathWorks, Inc., Natick, Massachusetts, USA). In order to control for type 1 error, we report adjusted p-values calculated for the non-parametric correlation analysis following the sequential Bonferroni-Holm procedure (Holm, 1979). This procedure is a sequentially rejective version of the simple Bonferroni correction for multiple comparisons, which strongly controls the family-wise error rate at level alpha. Howell (2002) recommends the Bonferroni-Holm procedure for multiple testing of several correlations from the same matrix. For regression analyses, it was ensured that all assumptions (no multicollinearity between the predictors, independence, homoscedasticity and normality of the errors) were met. Mediation regression analyses were computed using the SPSS macro "PROCESS" (Hayes, 2012). All statistical tests were two-tailed at an alpha level of 0.05 if not otherwise indicated.

RESULTS

RELATIONSHIP BETWEEN TRAIT EMOTIONAL INTELLIGENCE AND FLOW EXPERIENCE

Three univariate outliers with 2 *SD* above or below the mean score were removed in the averaged DFS-2 scores and in the averaged TEIQue-SF scores, respectively. Reliability analyses (Cronbach, 1951) were conducted for each standardized scale after the removal of the outliers. For the DFS-2, Cronbach's alpha coefficient was calculated on all 36 items and yielded a value of $\alpha = 0.89$ ($N = 73$). Individual analyses for each of the nine flow subscales revealed similarly high values between $\alpha = 0.71$ for the subscale of action-awareness merging and $\alpha = 0.90$ for the subscale of clear goals. Note that these values exceed Nunnally's (1978) criterion of 0.70 for acceptable reliability. Internal consistency was also assessed for the TEIQue-SF and considered as sufficiently high with a value of $\alpha = 0.83$ ($N = 73$). Basic descriptives of the flow and trait emotional intelligence scales are displayed in **Table 1**.

A comparison with the mean DFS-2 scores for each subscale as reported in Sinnamon et al. (2012) shows that, similar to their results reported for an sample of elite music performance students ($N = 80$), the mean rating for the subscale of Loss of Self-consciousness was the lowest among the nine subscales. In fact, the current result of 2.78 is similar to the reported mean value of 2.64 by Sinnamon et al. (2012), suggesting that Loss of Self-consciousness is a dimension of flow that may not be so relevant for flow in music performance. In the Sinnamon et al. (2012) study, the ranking of the nine subscales for the elite sample (studying music performance on a full-time basis) showed that Clear Goals (4.28), Autotelic Experience (4.19), Clear Feedback (3.96) and Challenge-skill Balance (3.92) were the four dimensions with the highest mean ratings. In our sample of piano performance students, Clear Goals (3.74), Autotelic Experience (3.66), Challenge-skill Balance (3.53) and Transformation of Time (3.50) were the dimensions with the highest mean ratings, indicating an overlap of three out of four dimensions between these two studies involving music performance students.

As a next step, the relationships between the individual flow subscales and the global flow score were investigated by correlation analyses. Shapiro-Wilk normality tests indicated significant deviations from normality for six out of the nine subscales after the removal of outliers 2 *SD* above or below the mean. Therefore, non-parametric Spearman-Rho correlations (r_s) were computed on the unaltered original scores (**Table 2**). The findings suggest that all nine subscales were moderately to highly correlated with the average global flow score. The subscale of Autotelic Experience was most highly correlated with global flow [$r_{s(74)} = 0.80$], followed by Sense of Control [$r_{s(74)} = 0.72$], Challenge-skill Balance [$r_{s(74)} = 0.70$] and Total Concentration [$r_{s(74)} = 0.68$]. The subscales of Transformation of Time [$r_{s(74)} = 0.46$], Loss of Self-consciousness [$r_{s(74)} = 0.43$], and Unambiguous Feedback [$r_{s(74)} = 0.47$] showed only moderate correlations with

Table 1 | Descriptive statistics of the DFS-2 scores, its nine subscales, and of TEIQue-SF (*N* = 73).

	M	*SD*	Min	Max	α
Mean flow score	3.37	0.38	2.67	4.25	0.89
Challenge-skill balance	3.53	0.60	2.50	5.00	0.80
Merging of action and awareness	3.21	0.60	2.00	5.00	0.71
Clear goals	3.74	0.79	1.75	5.00	0.90
Unambiguous feedback	3.50	0.65	1.50	5.00	0.84
Total concentration	3.24	0.58	2.00	5.00	0.77
Sense of control	3.14	0.52	2.00	5.00	0.74
Loss of self-consciousness	2.78	0.84	1.00	5.00	0.86
Transformation of time	3.50	0.78	1.75	5.00	0.81
Autotelic experience	3.66	0.80	2.00	5.00	0.87
Mean traitEI score	4.83	0.60	3.57	5.87	0.83

mean (M), standard deviation (SD), minimum (Min), maximum (Max), and Cronbach's alpha (α).

Table 2 | Spearman-Rho correlations between the global mean DFS-2 score and the mean scores of the nine flow subscales (*N* = 76).

Measure	Challenge-skill balance	Merging of action and awareness	Clear goals	Unambiguous feedback	Total concentration	Sense of control	Loss of self-consciousness	Trans-formation of time	Autotelic experience
Merging of action and awareness	0.22								
Clear goals	0.51*	0.12							
Unambiguous feedback	0.35	0.09	0.41*						
Total concentration	0.45*	0.23	0.39*	0.37					
Sense of control	0.50*	0.29	0.45*	0.37	0.56*				
Loss of self-consciousness	0.09	0.20	0.01	0.10	0.30	0.22			
Transformation of time	0.15	0.52*	0.13	−0.15	0.17	0.20	0.01		
Autotelic experience	0.63*	0.43*	0.37	0.26	0.49*	0.56*	0.32	0.24	
Mean flow score	0.70*	0.53*	0.59*	0.47*	0.68*	0.72*	0.43*	0.46*	0.80*

**$p < 0.05$ after Bonferroni-Holm correction; all dfs = 74.*

the average flow score. These results are reflected by the generally low inter-correlations between these subscales with all other subscales, suggesting that not all dimensions contributed equally strongly to the overall flow scores in pianists.

To predict overall flow experience during piano playing, a multiple stepwise linear regression analysis was conducted with the following predictors: *traitEI* (trait emotional intelligence), *practice* (daily amount of piano practice), *training* (overall duration of piano training), *age piano* (age of first piano lesson), *age* and *gender* (males = 1, females = 2). The average DFS-2 score, *flow*, was entered as the dependent variable. Outliers with 2 *SD* above and below the mean were removed from all variables prior to the analysis and cases were deleted list-wise; this resulted in 62 participants for the regression analysis. Correlations between the predictors are shown in **Table 3** and regression coefficients in **Table 4**. The basic descriptive values were as follows ($N = 62$): *traitEI* ($M = 4.81$, $SD = 0.61$), *practice* ($M = 3.05$ h, $SD = 1.49$), *training* ($M = 13.90$ years, $SD = 3.53$), *age piano* ($M = 6.35$ years, $SD = 2.00$), *age* ($M = 21.00$ years, $SD = 2.5$), *gender* (23 males, 39 females), *flow* ($M = 3.34$, $SD = 0.38$).

After two steps, the model was found to be successful in predicting flow experiences, $F_{(2,\ 59)} = 12.47, p < 0.001$. Two predictors, namely daily amount of practice and trait emotional intelligence, explained 27.0% of the overall variability of the flow scores (adjusted $R^2 = 0.27$). The sizes and significances of β-values indicated that daily amount of practice was the most important predictor, but that trait emotional intelligence contributed significantly to an improvement of the model in a second step, $\beta = 0.29$, $t_{(55)} = 2.37$, $p = 0.021$. In other words, the results were in line with our hypothesis that trait emotional intelligence and flow experience are positively correlated. The positive linear association between amount of practice, trait emotional intelligence and flow is depicted in **Figure 1**.

Next we explored the underlying relationships between trait emotional intelligence, the amount of daily practice and flow. This analysis was motivated by previous research showing that musical training is correlated with emotional intelligence

Table 3 | Pearson product-moment correlations between the average flow score and six predictors (*N* = 62).

Measure	Gender	Age	Practice	Training	Age piano	TraitEI
Age	−0.18					
Practice	−0.12	0.25				
Training	−0.03	0.49	0.12			
Age piano	−0.07	−0.12	−0.07	−0.70		
TraitEI	0.03	0.09	0.44	0.05	0.02	
Flow	−0.10	0.08	0.48	0.11	−0.05	0.45

Practice, daily amount of piano practice; training, overall duration of piano training; age piano, age of first piano lesson; traitEI, mean TEIQue-SF score; flow, mean DFS-2 score.

Table 4 | Summary of stepwise regression analysis for six variables predicting flow in piano performance students (*N* = 62).

Variable	*B*	*SE B*	β
STEP 1			
Constant	2.97	0.10	
Practice	0.12	0.03	0.48***
Adjusted R^2	0.22		
F	17.92***		
STEP 2			
Constant	2.2	0.34	
Practice	0.09	0.03	0.35**
TraitEI	0.18	0.08	0.29*
Adjusted R^2	0.27		
F	12.47***		
ΔR^2	0.07		

**$p < 0.05$, **$p < 0.01$, ***$p < 0.001$; B, non-standardized regression coefficient; β, standardized regression coefficient, SE, standard error; practice, daily amount of piano practice in hours; traitEI, mean TEIQue-SF score.*

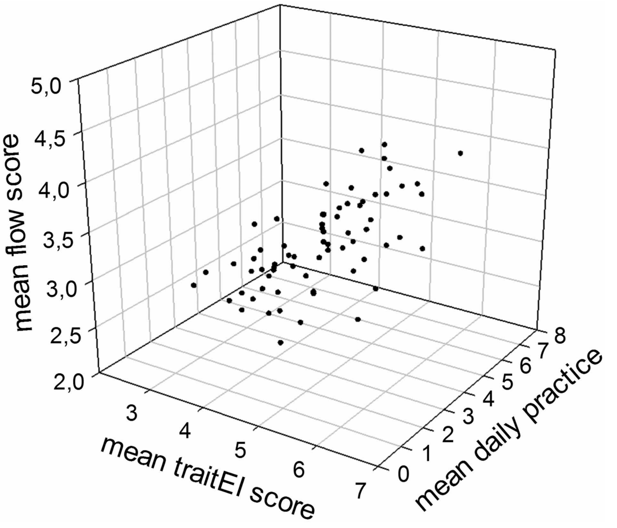


FIGURE 1 | Relationship between mean trait emotional intelligence scores, average amount of daily practice and mean dispositional flow scores in piano performance students ($N = 62$).

Table 5 | Indirect effect of daily amount of practice on flow experience through trait emotional intelligence ($N = 68$).

	Model predicting traitEI (ME)	
	Coeff	***SE***
Constant	4.28	0.15
Practice (X)	0.17***	0.05
Summary of model predicting ME	$R^2 = 0.17$***	
	Model predicting flow experience (Y)	
Constant	2.09***	0.32
Practice (X)	0.08*	0.03
TraitEI (ME)	0.21**	0.07
Summary of model predicting Y	$R^2 = 0.29$***	
	Indirect effect	**CI 95%**
	0.04	0.01 0.07

X, independent variable, ME, mediator variable, Y, dependent variable, Coeff, coefficient, CI, confidence interval. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.

(Petrides et al., 2006) and the recognition of emotional prosody in speech (Lima and Castro, 2011), with even some causal evidence for an effect of musical training (Thompson et al., 2004), and that musicians respond differently to musical emotions compared to non-musicians (Dellacherie et al., 2011; Marin et al., 2012). We decided to test a mediation model based on a bootstrapping approach (Hayes, 2012) with amount of daily practice as the independent variable (X), flow experience as the dependent variable (Y) and emotional intelligence as a mediator (ME). The analysis was conducted by the SPSS macro PROCESS (Hayes, 2012). First, a mediator model was computed, a dependent variable model was computed in a second step, and finally a confidence interval for the indirect effect was computed applying a bias-corrected resampling bootstrap technique with 5000 resamples. All relationships between the variables were modeled as linear and visually inspected prior to the analysis.

Table 5 summarizes the results of the mediation regression analysis. The model predicting trait emotional intelligence was significant, $F_{(1,\ 66)} = 13.70, p < 0.001$, indicating that daily practice and trait emotional intelligence were positively correlated with each other. The model predicting flow experience was also significant, $F_{(2,\ 65)} = 13.43, p < 0.001$, explaining around 30% of the variance in flow experience. We observed a significant direct effect of daily practice on flow experience, indicating how much a unit change in practice affects flow experience independent of its effect on trait emotional intelligence. Furthermore, there was a positive correlation between emotional intelligence and flow experience, after controlling for daily practice. Last, we found a significant positive indirect effect, implying that an increase of daily practice led to an increase in flow experience through the effect of daily practice on emotional intelligence. The mediation effect size (R^2) was small (0.12); therefore, these results should be interpreted with caution. For example, exchanging the independent variable with the mediator and vice versa did not change the overall results of the model. An alternative path model assuming that an increase in flow experiences increases levels of emotional intelligence through the effect of flow on practice was also tested and revealed a similar pattern of results. Since the intercorrelations between the three variables were similar in direction and strength, it was difficult to assess which model may be the correct one. At present, concrete theories on the relationship between flow, emotion and practice are lacking and thus cannot guide mediation analysis. Therefore, the current results should be regarded as exploratory.

RELATIONSHIP BETWEEN FLOW AND HIGH ACHIEVEMENT

Another testable hypothesis of interest concerned the relationship between flow and high achievement in piano performance. Therefore, to predict the likelihood that a piano student has won a prize in a piano competition (as a measure of high achievement), a binary logistic regression model (stepwise forward using the likelihood ratio statistic) was fitted to the data with seven predictors: *traitEI*, *practice*, *training*, *gender*, *age*, *age piano* and *flow*. Non-prize winners ($N = 24$) were coded as 1 and prize winners ($N = 38$) as 2. All assumptions for this statistical analysis were also verified. A test of the final model after two steps vs. a model with intercept only was statistically significant, $\chi^2(2) = 17.09$, $p < 0.001$. The model was able to correctly classify 73.7% of those who won a prize and 54.2% of those who did not, for an overall success rate of 66.1%. **Table 6** shows the logistic regression coefficient, Wald test and odds ratio for each of the two significant predictors of the model, namely *practice* and *age piano*. The odds ratio of *practice* indicates that for each one hour increase in piano practice per day, there is a doubling of the odds that the piano performance student would win a prize, when other variables are controlled. In other words, the odds increase around 111% for a change of 1 h of practice. This interpretation of the result is reliable because the respective 95% confidence interval was >1 (lower boundary = 1.28, upper boundary = 3.47). The second

Table 6 | Regression coefficients and overall model evaluation for a logistic regression analysis using seven predictors to model high achievement in piano performance ($N = 62$).

			Wald's			e^{β}
Predictor	***B***	***SE B***	**X^2**	***df***	***p***	**(odds ratio)**
STEP 1						
Constant	−1.63	0.72	5.12	1	0.024*	NA
Practice	0.74	0.25	8.72	1	0.003**	2.09
STEP 2						
Constant	0.67	1.24	0.29	1	0.589	NA
Practice	0.75	0.25	8.63	1	0.003**	2.11
Age piano	−0.37	0.17	4.54	1	0.033*	0.69
Test			**X^2**	***df***	***p***	
STEP 1						
Overall model evaluation Score test			11.93	1	0.001**	
Goodness-of-fit test Hosmer and Lemeshow			6.86	6	0.334	
STEP 2						
Overall model evaluation Score test			17.09	2	<0.001***	
Goodness-of-fit test Hosmer and Lemeshow			11.82	8	0.160	

*NA = non-applicable; Step 1: Cox and Snell $R^2 = 0.18$, Nagelkerke $R^2 = 0.24$; Step 2: Cox and Snell $R^2 = 0.24$, Nagelkerke $R^2 = 0.33$; *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.*

significant predictor in the model was negative and referred to the age at which piano performance students began their piano training. The odds ratio was 0.69 and the respective 95% confidence interval was <1 (lower boundary = 0.50, upper boundary = 0.97), meaning that the odds of winning a prize in a competition were 0.69 lower for those who started their piano training one year later, or that there is a 31% decrease in the odds for winning a prize for each one-year increase in the age at which piano training began. For interpretational purposes one can also invert the odds ratio for this negative predictor, which shows that, for each one-year decrease in the age at which piano training began, the odds of winning a prize in a piano competition increase by a multiplicative factor of 1.44 (44%). In summary, although amount of practice and the age at which the lessons began were shown to be significant predictors, our hypothesis that flow experiences predicts high achievement in piano performance was not corroborated by the current data.

FLOW, MUSICAL EMOTIONS AND MUSICAL STYLES

The analysis of the self-developed questionnaire on flow and emotion in music performance showed that 69 out of 76 pianists had experienced flow as defined by the nine-dimensional concept of flow. Sixty-three pianists estimated to experience on average 5.00 flow states ($SD = 6.37$) per month during piano playing and 7.56 flow states per month ($SD = 8.97$) during music listening. A Wilcoxon signed-rank test ($N = 56$) showed that the number of estimated flow states during music listening was significantly higher than the one during piano performance, $T = 262$, $p = 0.004$, $r = -0.39$. A Spearman-Rho correlation further indicated that there was a significant positive correlation between the number of flow states experienced during piano performance and music listening in a typical month, $r_{s(55)} = 0.50$, $p < 0.001$. The majority of participants also reported that flow experiences kept them motivated to practice the piano and to become better. The frequency of answers was as follows: 65.2% "yes, agree," 20.3% "yes, somehow agree," 5.8% "no, somehow disagree," 1.5% "no, don't agree" and 7.2% of the pianists did not know. Moreover, the majority of pianists ($N = 68$) somehow or completely agreed that flow states in piano performance can only be reached when the piece is nearly ready for a public performance: 19.1% "yes, agree," 48.5% "yes, somehow agree," 14.7% "no, somehow disagree," 11.8% "no, don't agree" and 5.9% of the pianists did not know. Last, we were also interested in whether flow and life satisfaction were linked in piano performance students and the data revealed the following answers ($N = 66$): 43.9% "always," 30.3% "frequently," 18.2% "sometimes," 1.5% "rarely," 0% "never" and 6.1% "don't know."

Several items of the questionnaire referred to the relationship between flow and emotion in piano performance. Specifically, two questions addressed whether flow experiences are accompanied with intense feelings of happiness, differentiating between happiness *during* and *after* flow states piano performance. For the question referring to happiness during flow states, the pattern of results was as follows: 39.7% "always," 20.6% "frequently," 22.1% "sometimes," 10.3% "rarely", 2.9% "never" and 4.4% "don't know." Happiness after flow states was even more common, as shown by the following replies: 46.4% "always," 30.4% "frequently," 13.0% "sometimes," 2.9% "rarely," and 7.3% "don't know."

Next, a set of items referred to musical emotions, that is, emotions that are expressed or induced by the musical structure, and 62 out of 69 participants who experienced flow during piano performance agreed that flow states are more easily induced by certain types of emotions expressed by a musical piece than by others. In a similar vein, 61 out of 69 participants also responded that flow states depend on the nature of emotions induced by music. **Table 7** summarizes responses to how often emotions varying in arousal and pleasantness, which were either expressed or induced by the music, led to flow in the current sample of pianists. For both expressed and induced musical emotions, low-arousing unpleasant emotions were not so frequently associated with flow states than other emotions, such as high-arousing pleasant and unpleasant emotions.

Two other questions addressed emotion and flow. First, participants ($N = 69$) were asked to indicate whether they would agree that flow states appear less frequently when the emotional content of a piece varies a lot over the course of a piece. Thirty-six point two per cent of the pianists answered "no, somehow disagree," 17.4% with "no, don't agree," 24.6% with "yes, somehow agree," 11.6% with "yes, agree" and 10.1% with "don't know," which suggests the existence of two subgroups in our sample, those who agree (36.2%) and those who do not (53.6%). Second, participants ($N = 69$) indicated whether they felt that flow states are more easily reached when the piece induces emotions that they particularly like in general (can be either positive or negative emotions). Here, the pattern of results was more clear and showed that the majority of pianists agreed (44.9%) or somehow agreed (37.7%), whereas only 13.0% somehow disagreed and 4.35% did not know. Taken together, these exploratory results are in line with the view that flow is a highly emotional experience, and further, suggest that musical emotions may play an important role in the induction of flow in performing artists.

A final set of items probed whether musical emotions and flow experience were associated through the musical style during piano performance. Participants ($N = 67$) reported whether they experienced flow more often when playing certain musical styles. Most participants agreed that the musical style played a role in flow states. The frequency of answers was as follows: 35.8% "yes, agree," 35.8% "yes, somehow agree," 10.5% "no, somehow disagree," 7.5% "no, don't agree" and 10.5% of the pianists did not know. Furthermore, participants ($N = 68$) were asked to select the musical style in which they had experienced flow in piano performance most frequently. Pianists associated most frequently the Romantic style with flow (64.7%), followed by Classical (13.2%), Contemporary (8.8%), Baroque (2.9%), Other (10.3%) and Jazz (0%). In order to see whether this finding conforms to the pianists' preference for and familiarity with these musical styles, two other questions relating to the musical background were analyzed. First, a question referred to pianists' ($N = 76$) most favorite musical style in piano performance and the pattern of results was as follows: Romantic (57.9%), Contemporary (14.5%), Classical (10.5%), Baroque (6.6%), Jazz (5.3%) and Other (5.3%). Second, pianists ($N = 76$) had to indicate which musical style they played most frequently during the last five years: Romantic (42.1%), Classical (31.6%), Contemporary (11.8%), Baroque (4.0%), Jazz (2.6%) and Other (7.9%). In summary, for this specific sample of piano performance students, the Romantic style was the most familiar, preferred and also most flow-inducing. This finding corresponds to pianists' ($N = 68$) large agreement on the question whether flow states are more easily reached when playing pieces that they particularly like: 69.1% "yes, agree," 25.0% "yes, somehow agree," 2.9% "no, somehow disagree," 1.5% "no, don't agree" and 1.5% of the pianists did not know.

Table 7 | Emotions varying in arousal and pleasantness and their frequency of being related to flow states.

Musical emotions	Very often	Often	Sometimes	Never
EXPRESSED EMOTIONS ($N = 66$)				
Low-arousing pleasant	11	15	39	1
High-arousing pleasant	18	25	22	1
Low-arousing unpleasant	8	13	33	12
High-arousing unpleasant	13	19	28	6
INDUCED EMOTIONS ($N = 65$)				
Low-arousing pleasant	9	26	28	2
High-arousing pleasant	16	28	19	2
Low-arousing unpleasant	5	17	36	7
high-arousing unpleasant	13	16	31	5

Next we analyzed the possible link between familiarity, preference and flow induction with regard to musical styles at an individual level (**Table 8**). For thirty-one participants (45.6%) out of 68, the most frequently played musical style was also the most flow-inductive style, regardless of the type of musical style. The results further indicated that a high number of pianists ($n = 25$) selected the Romantic style as the most frequently played and most flow-inductive. However, the data also showed that 14 pianists who frequently played the classical style chose the Romantic style as the most flow-inductive style, suggesting that the Romantic style may be more flow-inductive than other styles. Similarly, we assessed whether there was a relationship between preference for a musical style and frequent flow induction. For forty-two (61.8%) out of 68 participants, the most favorite musical style was also the most flow-inductive style, regardless of the type of musical style. Romantic music was the most preferred musical style and also the most flow-inductive style for

Table 8 | Relationships between the most frequent occurrence of flow and the frequency of playing a musical style and the preference for a musical style, respectively ($N = 68$).

Most flow	Baroque	Classical	Romantic	Contemporary	Jazz	Other
MOST FREQUENTLY PLAYED MUSICAL STYLE						
Baroque	0	1	0	1	0	0
Classical	0	3	3	2	1	0
Romantic	1	14	25	1	0	3
Contemporary	1	1	1	3	0	0
Jazz	0	0	0	0	0	0
Other	0	1	1	1	1	3
MOST FAVORITE MUSICAL STYLE						
Baroque	0	1	0	0	0	1
Classical	1	2	2	2	2	0
Romantic	1	3	34	5	1	0
Contemporary	1	0	2	3	0	0
Jazz	0	0	0	0	0	0
Other	0	0	3	0	1	3

34 piano performance students. Ten participants who had not chosen Romantic music as the most favorite style indicated that Romantic music was frequently flow-inductive.

Finally, we asked pianists ($N = 69$) to name composers whose pieces had reliably induced flow states in the past (different pieces by the same composer over a longer period of time). Pianists could name as many composers as they wished. Fifteen pianists did not name any composer. The responses of the other pianists were counted and those composers that were only named once were added to the category "Other." Note that pianists could name more than one composer and all responses were considered in the analysis. **Figure 2** depicts that Frédéric Chopin (1810–1849) was clearly the most often named composer, mentioned by 25 pianists, followed by Beethoven (13), Debussy (12) and J. S. Bach (8). These findings not only show that pianists were able to relate composers to flow experiences, but also that there was high agreement among pianists that Chopin's music is particularly flow-inducing.

DISCUSSION

Given the apparent paucity of research on flow and individual differences in music performance, the present study sought to investigate flow in relation to trait emotional intelligence in piano performance students. The rationale for this approach lies in the facts that being in a flow state is regarded as a highly emotional and intrinsically rewarding experience (Csikszentmihalyi, 1990), that music is strongly communicative of emotions (Juslin and Sloboda, 2010), and that being able to effectively deal with musical emotions may thus underlie the proneness of achieving a flow state during music performance. Further support for our approach to study flow in relation to emotion was recently provided by a study suggesting that the proneness to experience flow may be associated with personality dimensions that are under dopaminergic control and be reflected in low impulsiveness, stable emotion and positive affect (De Manzano et al., 2013). This is in line with findings by Montag et al. (2011), who observed that during listening to pleasant and unpleasant music individual differences with regard to self-transcendence modulate activity in the ventral striatum, which is part of the reward-circuitry. Last, recent research on the underlying genetic architecture of individual differences with respect to general flow proneness indicated that the same genetic factors may influence flow experienced across domains, whereas specific environmental factors may explain differences in flow proneness in different domains (Mosing et al., 2012a,b). Based on these findings we hypothesized that there is a positive association between trait emotional intelligence and flow experiences among musicians.

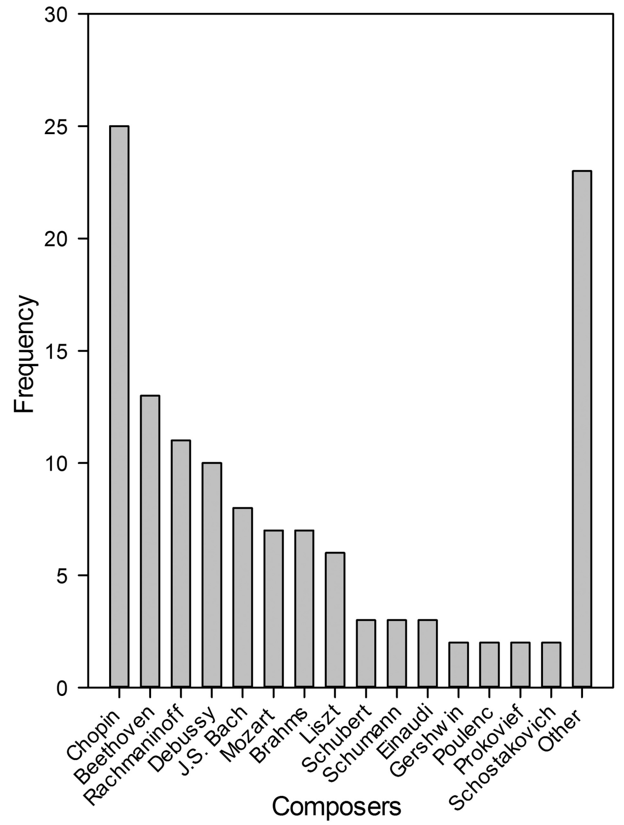


FIGURE 2 | Frequency distribution of composers who repeatedly induced flow states in a sample of piano performance students (*N* = 69). More than one name could be given.

Our correlational study comprised a sample of undergraduate and postgraduate piano performance students ($N = 76$), implying that these students were professionally engaged with piano playing. We did not test for pianists' ability to deal with musical emotions but used a general personality test (Petrides and Furnham, 2006) to predict the disposition to achieve flow states (Jackson and Eklund, 2002) during piano performance. A stepwise regression analysis, including trait emotional intelligence, gender, age, age of first piano training, duration of piano training, and daily amount of piano practice as predictors, revealed that besides the amount of piano practice, trait emotional intelligence was the only other significant predictor in the model. In other words, the higher the trait emotional intelligence of a piano performance student, the more prone is s/he to experience flow. The positive association between the two variables is in line with models of emotional intelligence that claim that the ability to get into a flow state is a sign of high emotional intelligence (Goleman, 1995). It remains to be seen whether this positive relationship between trait emotional intelligence and dispositional flow can also be observed under experimental conditions and in domains outside music. Music performance may be a kind of activity in which emotional communication plays a larger role than in other physical and cognitive activities.

Our regression model demonstrated a positive relationship between trait emotional intelligence, daily amount of practice and flow, and yielded an adjusted R^2 of around 0.27. Thus, it can be argued that the model needs to be extended and improved by including other predictors. Given that gender, age and predictors related to musical training were not predictive of flow in this rather homogeneous sample of piano performance students, it seems pertinent to assume that other personality features that are not covered by our study may also contribute to flow experience. Therefore, future models could include those traits that have been predictive of flow in domains outside music. For instance, locus of control (Keller and Blomann, 2008) as well as self-control (Kuhnle et al., 2012), novelty seeking and persistence (Teng, 2011) have been associated with flow experiences. Moreover, the investigation of mediation effects in a set of personality traits may be a promising avenue for future research on the existence of an autotelic personality among musicians.

The underlying relationship between daily amount of piano practice, trait emotional intelligence and dispositional flow was further examined by fitting a mediator model to the data. Research on the relationship between amount of musical training and emotional responses to musical emotions (Dellacherie et al., 2011; Marin et al., 2012 but see Bigand et al., 2005) and emotional prosody in speech (Thompson et al., 2004; Lima and Castro, 2011; Thompson et al., 2012; but see Trimmer and Cuddy, 2008) is somewhat relevant for the current research and was thus taken as a conceptual starting point for modeling effects of the amount of daily practice on flow through trait emotional intelligence. Our mediator model was significant, but similar results were also obtained for an alternative path model (flow—practice—emotional intelligence). This clearly limits the interpretation of the suggested mediator model and more (experimental) research is needed to elucidate the relationship between practice behavior, trait emotional intelligence and flow during music performance.

A set of self-developed questions corroborated the hypothesis that the ease of experiencing flow is related to the emotions expressed and induced by a musical piece. The majority of participants (around 89%) acknowledged the role of musical emotions in flow induction. The results further suggest that pleasant and unpleasant high-arousing musical emotions are more associated with the experience of flow than unpleasant low-arousing musical emotions, a finding which was valid for both expressed and induced musical emotions. Future experiments may explore the role of specific types of musical emotions and musicians' ability to deal with them with regard to the nine flow dimensions proposed by Csikszentmihalyi (1990). One testable hypothesis that directly follows from the current results is that high-arousing pleasant musical emotions may be more strongly associated with the dimension of autotelic experience because the latter is usually accompanied with enjoyment and happiness, which are both characterized by high arousal. In other words, it is possible that a congruency between musical emotions and emotions inherent to autotelic experience may facilitate the latter state. From our perspective, the current research could also be extended by adding tests on emotional ability involving musical stimuli, which may offer additional insight into unresolved questions regarding the role of emotions in flow induction.

Since it is known that musical styles vary in their degree of emotional expressivity (Kallinen, 2005), we also explored whether the degree of induced flow may depend on the musical style. Our data suggests that pianists largely support the view that flow experiences occur more often with certain musical styles (around 72% "agreed" or "somehow agreed" with this statement). The majority of our participants associated Romantic music, and particularly the music by Frédéric Chopin, with flow experiences. However, although the Romantic era and its music are generally regarded as being strongly expressive of emotions (Robinson, 2005), this musical style was also the most familiar and preferred one among pianists. Further analyses based on individual relationships between these variables revealed that for 45.6% of the participants the most familiar musical style was also the most flow-inductive style, regardless of the type of musical style. Nonetheless, 14 participants who most frequently played classical music associated the Romantic style with flow experience. Accordingly, there is mild evidence that familiarity may not be the sole explanation in the flow-musical style relationship. We further observed a link between the most favorite musical style and flow in 61.8% of the participants, regardless of the specific musical style. Of course, further studies would be necessary to disentangle effects of familiarity and preference on flow from those that are due to the musical structure of the style.

A binary logistic regression model was computed to predict high achievement among piano performance students (i.e., having won prizes at piano competitions). Our hypothesis that enhanced levels of experienced flow may predict high achievement in piano performance could not be supported by the current data. Instead, the logistic regression model indicated that the amount of daily practice and the age at which piano training began were the only significant predictors. This result is essentially in line with research on professional achievement in music performance which regards experience and practice as crucial for superior expert performance (Sloboda et al., 1996; Lehmann and Ericsson, 1997; Gabrielsson, 2003), but which also suggests that superior music performance may be a multifaceted phenomenon that is conceptually complex and difficult to model (Hallam, 1997; Ericsson, 2006). For instance, recent research has shown that visual information largely influences judgments of musical performances in competitions (Tsay, 2013), which may partly explain why flow did not predict high achievement in the current sample of pianists. A related issue concerns the possible link between high achievement and some external locus of control, which may counteract the positive relationship between internal locus of control, flow and high achievement. Our finding that the age of the first piano training was predictive of success in piano performance is in line with results indicating that there may be a sensitive period in early childhood where musical practice in the form of motor training may lead to benefits for performance in adulthood (e.g., Watanabe et al., 2007; Penhune, 2011).

A dissociation between high degrees of flow and high achievement has been previously reported in sports (Jackson, 1999). In a similar vein, Wrigley and Emmerson (2013) investigated flow states during a music performance examination and did not find that students who further progressed in their studies experienced flow more often than those who did not. Privette (1983) discussed the differences and similarities between the constructs of peak experience, peak performance and flow. She suggested that, for example, the notion of playfulness may be essential to flow but not to peak performance and further, that a strong sense of self is common for peak performances but not for flow. Although there is a substantial overlap between these constructs, differences on one dimension of the construct, in combination with effects of social context (e.g., practice vs. performance vs. exam), may explain discrepancies in research results and should thus be considered in future research.

The current study also provided some insights into the question of whether all nine dimensions of the flow concept developed by Csikszentmihalyi (1990) contribute equally well to the global flow score as assessed by the frequently used dispositional and state flow scales developed by Jackson and colleagues (Jackson and Marsh, 1996; Jackson and Eklund, 2002). In general, we found positive correlations between all nine subscales and the global DFS-2 scale, indicating that all different dimensions of dispositional flow play a role in flow experienced during music performance. Sinnamon et al. (2012), also assessing dispositional flow, reported that the DFS-2 subscales of Transformation of Time and Loss of Self-consciousness correlated more weakly with the other flow subscales in a sample of music students. It is interesting to note that whereas the current study asked piano performance students to think of performing a piece that is already well-mastered, Sinnamon et al. (2012) asked their participants to rate the items based on their experience of performing music in general. In both cases, the dimensions of Transformation of Time and Loss of Self-consciousness appeared as being less correlated with other flow dimensions, corroborating previous findings in the domain of sports (e.g., Jackson, 1996; Jackson et al., 2001; Jackson and Eklund, 2002). This finding may also indicate that the specific instructions of current studies on dispositional flow in music performance did not largely affect the inter-relationships between the subscales and the relationship with the global flow score. Moreover, our results suggest that the subscale of Unambiguous Feedback is another dimension of flow that is less correlated with other subscales, which is in line with reports by Sinnamon et al. (2012). Finally, previous research on musical flow involving the Flow State Scale-2 (Jackson and Marsh, 1996) reported that the subscale Transformation of Time was among the weakest predictors of global flow state and that Autotelic Experience, Sense of Control and Challenge-skill Balance were among the strongest (Wrigley and Emmerson, 2013). Our current results are similar to these findings. In summary, these results illustrate that the inter-relationships between the global flow scale and its subscales may be similar when flow is assessed in different scenarios of music performance (dispositional vs. state). However, more empirical evidence is needed to support this claim.

In conclusion, this study highlights the role of emotions in flow experience in two ways. First, individual differences regarding trait emotional intelligence predict dispositional flow, and second, pianists acknowledge the role of musical emotions in the induction of flow. Both findings warrant further experimental investigations for generalizations, by including other instrument groups and artistic activities, such as dancing and painting, in which emotional communication is also vital.

AUTHOR CONTRIBUTIONS

Manuela M. Marin and Joydeep Bhattacharya conceived and designed the research. Manuela M. Marin collected and analyzed the data. Manuela M. Marin and Joydeep Bhattacharya wrote the article.

ACKNOWLEDGMENTS

We thank Bruno Gingras for comments on an earlier draft of the manuscript.

REFERENCES

Asakawa, K. (2010). Flow experience, culture, and well-being: how do autotelic Japanese college students feel, behave, and think in their daily lives? *J. Happiness Stud.* 11, 205–223. doi: 10.1007/s10902-008-9132-3

Baker, F. A., and MacDonald, R. A. (2013). Flow, identity, achievement, satisfaction and ownership during therapeutic songwriting experiences with university students and retirees. *Music Sci.* 17, 197–229. doi: 10.1177/1029864913476287

Bakker, A. B. (2005). Flow among music teachers and their students: the crossover of peak experiences. *J. Vocat. Behav.* 66, 26–44. doi: 10.1016/j.jvb.2003.11.001

Bigand, E., Vielliard, S., Madurell, F., Marozeau, J., and Dacquet, A. (2005). Multidimensional scaling of emotional responses to music: the effect of musical expertise and of the duration of the excerpts. *Cogn. Emot.* 19, 1113–1139. doi: 10.1080/02699930500204250

Bloom, A. J., and Skutnick-Henley, P. (2005). Facilitating flow experiences among musicians. *Am. Music Teacher* 54, 24–28.

Busch, H., Hofer, J., Chasiotis, A., and Campos, D. (2013). The achievement flow motive as an element of the autotelic personality: predicting educational attainment in three cultures. *Eur J. Psychol. Educ.* 28, 239–254. doi: 10.1007/s10212-012-0112-y

Buttsworth, L. M., and Smith, G. A. (1995). Personality of Australian performing musicians by gender and by instrument. *Pers. Indiv. Differ.* 18, 595–603. doi: 10.1016/0191-8869(94)00201-3

Charyton, C., and Snelbecker, G. E. (2010). General, artistic, and scientific creativity attributes of engineering and music students. *Creat. Res. J.* 19, 213–225. doi: 10.1080/10400410701397271

Chmurzynska, M. (2012). "Personality conditions of pianists' achievements," in *Proceedings of the 12th International Conference on Music Perception and Cognition and 8th Triennal Conference of the European Society of the Cognitive Sciences of Music,* eds. E. Cambouropoulos, C. Tsougras, P. Mavromatis, and K. Pastiadis (Thessaloniki: Aristotle University of Thessaloniki), 214–221.

Clarke, S. G., and Haworth, J. T. (1994). Flow experience in the daily lives of sixth-form college students. *Br. J. Psychol.* 85, 511–523. doi: 10.1111/j.2044-8295.1994.tb02538.x

Cooper, C. L., and Wills, G. I. D. (1989). Popular musicians under pressure. *Psychol. Music* 17, 22–36. doi: 10.1177/0305735689171003

Cribb, C., and Gregory, A. H. (1999). Stereotypes and personalities of musicians. *J. Psychol. Inter. Appl.* 133, 104–114. doi: 10.1080/00223989909599725

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika* 16, 297–334. doi: 10.1007/BF02310555

Crust, L., and Swann, C. (2013). The relationship between mental toughness and dispositional flow. *Eur. J. Sport Sci.* 13, 215–220. doi: 10.1080/17461391.2011.635698

Csikszentmihalyi, M. (1975). *Beyond Boredom and Anxiety*. San Francisco, CA: Jossey-Bass.

Csikszentmihalyi, M. (1990). *Flow: The Psychology of Optimal Performance*. New York, NY: Cambridge University Press.

Csikszentmihalyi, M. (2002). *Flow: The Classic Work on How to Achieve Happiness*. London: Rider.

Csikzentmihalyi, M., and Csikzentmihalyi, I. (eds.). (1988). *Optimal Experience: Psychological Studies of Flow in Consciousness*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511621956

Csikszentmihalyi, M., and LeFevre, J. (1989). Optimal experience in work and leisure. *J. Per. Soc. Psychol.* 56, 815–822. doi: 10.1037/0022-3514.56.5.815

Custodero, L. (2002). Seeking challenge, finding skill: flow experience in music education. *Arts Educ. Policy Rev.* 103, 3–9. doi: 10.1080/10632910209600288

Custodero, L. (2005). Observable indicators of flow experience: a developmental perspective of musical engagement in young children from infancy to school age. *Music Edu. Res.* 7, 185–209. doi: 10.1080/14613800500169431

Custodero, L. (2012). "The call to create: flow experience in musical learning and teaching," in *Musical Imaginations*, eds D. Hargreaves, D. Miella, and R. MacDonald (Oxford: Oxford University Press), 369–384.

Dellacherie, D., Roy, M., Hugueville, L., Peretz, I., and Samson, S. (2011). The effect of musical experience on emotional self-reports and psychophysiological responses to dissonance. *Psychophysiology* 48, 337–349. doi: 10.1111/j.1469-8986.2010.01075.x

De Manzano, O., Cervenka, S., Jucaite, A., Hellenas, O., Farde, L., and Ullen, F. (2013). Individual differences in the proneness to have flow experiences are linked to dopamine D2-receptor availability in the dorsal striatum. *Neuroimage* 67, 1–6. doi: 10.1016/j.neuroimage.2012.10.072

De Manzano, O., Theorell, T., Harmat, L., and Ulle, F. (2010). The psychophysiology of flow during piano playing. *Emotion* 10, 301–311. doi: 10.1037/a0018432

DeNeve, K. M., and Cooper, H. (1998). The happy personality: a meta-analysis of 137 personality traits and subjective well-being. *Psychol. Bull.* 124, 197–229. doi: 10.1037/0033-2909.124.2.197

Diaz, F. M. (2011). Mindfulness, attention, and flow during music listening: an empirical investigation. *Psychol. Music* 41, 42–58. doi: 10.1177/0305735611415144

Eisenberger, R., Jones, J. R., Stinglhamber, F., Shanock, L., and Randall, A. T. (2005). Flow experiences at work: for high achievers alone? *J. Organiz. Behav*. 26, 755–775. doi: 10.1002/job.337

Ericsson, K. A. (2006). "The influence of experience and deliberate practice on the development of superior expert performance," in *Cambridge Handbook of Expertise and Expert Performance*, eds K. A. Ericsson, N. Charness, P. Feltovich, and R. R. Hoffman (Cambridge: Cambridge University Press), 685–706. doi: 10.1017/CBO9780511816796

Freer, P. (2009). Boys' descriptions of their experiences in choral music. *Res. Stud. Music Educ*. 31, 142–160. doi: 10.1177/1321103X09344382

Fritz, B. S., and Avsec, A. (2007). The experience of flow and subjective well-being of music students. *Psiholoska Obzorja/Horizons Psychol.* 16, 5–17.

Fullagar, C. J., Knight, P. A., and Sovern, H. S. (2013). Challenge/skill balance, flow, and performance anxiety. *Appl. Psychol. Int. Rev*. 62, 236–259. doi: 10.1111/j.1464-0597.2012.00494.x

Gabrielsson, A. (2002). Emotion perceived and emotion felt: same or different? *Music Sci*. Special issue 2011–2002, 123–147.

Gabrielsson, A. (2003). Music performance research that the millennium. *Psychol. Music* 31, 221–272. doi: 10.1177/03057356030313002

Gibson, C., Folley, B. S., and Park, S. (2009). Enhanced divergent thinking and creativity in musicians: a behavioural and near-infrared spectroscopy study. *Brain Cogn.* 69, 162–169. doi: 10.1016/j.bandc.2008.07.009

Goleman, D. (1995). *Emotional Intelligence*. New York, NY: Bantham Book.

Hallam, S. (1997). "What do we know about practising? Towards a model synthesising the research literature," in *Does Practice Make Perfect? Current Theory and Research on Instrumental Music Performance*, eds Jorgensen, H. and A. C. Lehmann (Oslo: Norwegian State Academy of Music), 179–231.

Hatfield, E., Cacioppo, J., and Rapson, R. L. (1994). *Emotional Contagion*. New York, NY: Cambridge University Press.

Hayes, A. F. (2012). *PROCESS: A Versatile Computational Tool for Observed Variable Mediation, Moderation, and Conditional Process Modeling [White paper]*. Available online at: http://www.afhayes.com/

Hernandez, D., Russo, S. A., and Schneider, B. (2009). The psychological profile of a rock band: using intellectual and personality measures with musicians. *Med. Probl. Perform. Ar*. 24, 71.

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scand. J. Stat*. 6, 65–70.

Howell, D. C. (2002). *Statistical Methods for Psychology*. Pacific Grove, CA: Wadsworth.

Jackson, S. A. (1996). Toward a conceptual understanding of the flow experience in elite athletes. *Res. Q. Exercise Sport* 67, 76–90. doi: 10.1080/02701367.1996.10607928

Jackson, S. A. (1999). "Joy, fun, and flow state in sport," in *Emotions in Sports,* ed Y. L. Hannin (Illinois, IL: Human Kinetics), 135–156.

Jackson, S. A., and Eklund, R. C. (2002). Assessing flow in physical activity: the flow state scale-2 (FSS-2) and dispositional flow scale-2 (DFS-2). *J. Sport Exerc. Psychol*. 24, 133–150.

Jackson, S. A., and Eklund, R. C. (2004). *The Flow Scales Manual*. Morgantown, WV: Fitness Information Technology.

Jackson, S. A., and Marsh, H. W. (1996). Development and validation of a scale to measure optimal experience: the flow state scale. *J. Sport Exerc. Psychol.* 18, 17–35.

Jackson, S. A., Thomas, P. R., Marsh, H. W., and Smethurst, C. J. (2001). Relationships between flow, self-concept, psychological skills, and performance. *J. Appl. Sport Psychol.* 13, 154–178. doi: 10.1080/104132001753149865

Juslin, P. N., and Sloboda, J. A. (2010). *Handbook of Music and Emotion*. Oxford: Oxford University Press.

Kallinen, K. (2005). Emotional ratings of music excerpts in the western art music repertoire and their self-organization in the Kohonen neural network. *Psychol. Music* 33, 373–393. doi: 10.1177/0305735605056147

Keller, J., and Blomann, F. (2008). Locus of control and the flow experience: an experimental analysis. *Eur. J. Pers.* 22, 589–607. doi: 10.1002/per.692

Kemp, A. E. (1996). *The Musical Temperament: Psychology and Personality of Musicians.* New York, NY: Oxford University Press. doi: 10.1093/acprof:oso/9780198523628.001.0001

Kraus, B. N. (2003). Musicians in flow. Optimal experience in the wind ensemble rehearsal. Dissertation abstract international section A. *Hum. Soc. Sci.* 64, 839.

Kuhnle, C., Hofer, M., and Kilian, B. (2012). Self-control as predictor of school grades, life balance, and flow in adolescents. *Br. J. Educ. Psychol*. 82, 533–548. doi: 10.1111/j.2044-8279.2011.02042.x

Lamont, A. (2009). "Strong experiences of music in university students," in *Proceedings of the 7th Triennial Conference of the European Society of the Cognitive Sciences of Music*, eds J. Louhivuori, T. Eerola, S. Saarikallio, T. Himberg and P.-S. Eerola (Jyväskylä: University of Jyväskylä), 250–259.

Langendörfer, F. (2008). Personality differences among orchestra instrumental groups: Just a stereotype? *Pers. Individ. Differ*. 44, 610–620. doi: 10.1016/j.paid.2007.09.027

Larson, R. (1989). Is feeling "in control" related to happiness in daily life? *Psychol. Rep.* 64, 775–784. doi: 10.2466/pr0.1989.64.3.775

Laukka, P., and Quick, L. (2011). Emotional and motivational uses of music in sports and exercise: a questionnaire study among athletes. *Psychol. Music* 41, 198–215. doi: 10.1177/0305735611422507

Lefcourt, H. M. (1991). "Locus of control," in *Measures of Personality and Social Psychological Attitudes*, eds J. P. Robinson, P. R. Shaver, and L. S. Wrightman (San Diego, CA: Academic Press), 413–499.

Lehmann, A. C., and Ericsson, K. A. (1997). Research on expert performance and deliberate practice: Implications for the education of amateur musicians and music students. *Psychomusicology* 16, 40–58. doi: 10.1037/h0094068

Kenny, D. T., Davis, P., and Oates, J. (2004). Music performance anxiety and occupational stress among opera chorus artists and their relationship with state and trait anxiety and perfectionism. *J. Anx. Disord*. 18, 757–777. doi: 10.1016/j.janxdis.2003.09.004

Levenson, H. (1981). "Differentiating among internality, powerful others, and chance," in *Research With the Locus of Control Construct*, ed H. Lefcourt (New York, NY: Academic Press), 15–63.

Lima, C. F., and Castro, S. L. (2011). Speaking to the trained ear: musical expertise enhances the recognition of emotions in speech prosody. *Emotion* 11, 1021–1031. doi: 10.1037/a0024521

Logan, R. (1988). "Flow and solitary ordeals," in *Optimal Experience: Psychological Studies of Flow in Consciousness*, eds M. Csikszentmihalyi and I. Csikszentmihalyi (Cambridge: Cambridge University Press), 172–180. doi: 10.1017/CBO9780511621956.010

MacDonald, R., Byrne, C., and Carlton, L. (2006). Creativity and flow in musical composition: an empirical investigation. *Psychol. Music* 34, 292–306. doi: 10.1177/0305735606064838

Marchant-Haycox, S. E., and Wilson, G. D. (1992). Personality and stress in performing artists. *Pers. Individ. Differ*. 13, 1061–1068. doi: 10.1016/0191-8869(92)90021-G

Marin, M. M., Gingras, B., and Bhattacharya, J. (2012). Crossmodal transfer of arousal, but not pleasantness, from the musical to the visual domain. *Emotion* 12, 618–631. doi: 10.1037/a0025020

Martin, A. J., and Jackson, S. A. (2008). Brief approaches to assessing task absorption and enhanced subjective experience: examining 'short' and 'core' flow in diverse performance domains. *Motiv. Emot.* 32, 141–157. doi: 10.1007/s11031-008-9094-0

Maslow, A. (1968). *Toward a Psychology Being*. Oxford: D. Van Nostrand Reinhold Company.

Moneta, G. B. (2004). The flow experience across cultures. *J. Happiness Stud.* 5, 115–121. doi: 10.1023/B:JOHS.0000035913.65762.b5

Montag, C., Reuter, M., and Axmacher, N. (2011). How one's favorite song activates the reward circuitry of the brain. Personality matters! *Behav. Brain Res*. 225, 511–514. doi: 10.1016/j.bbr.2011.08.012

Mosing, M. A., Magnusson, P. K. E., Pedersen, N. L., Nakamura, J., Madison, G., and Ullen, F. (2012a). Heritability of proneness for psychological flow experiences. *Pers. Individ. Differ*. 53, 699–704. doi: 10.1016/j.paid.2012.05.035

Mosing, M. A., Pedersen, N. L., Cesarini, D., Johannesson, M., Magnusson, P. K. E., Nakamura, J., et al. (2012b). Genetic and environmental influences on the relationship between flow proneness, locus of control and behavioural inhibition. *PLoS ONE* 7:e47958. doi: 10.1371/journal.pone.0047958

Naditch, M. P., Gargan, M., and Michael, L. (1975). Denial, anxiety, locus of control, and the discrepancy between aspirations and achievements as components of depression. *J. Abnorm. Psychol.* 84, 1–9. doi: 10.1037/h0076254

Nakamura, J., and Csikszentmihalyi, M. (2009). "Flow theory and research," in *Handbook of Positive Psychology*, eds S. J. Lopez and C. R. Snyder (Oxford: Oxford University Press), 195–206.

Nunnally, J. C. (1978). *Psychometric Theory, 2nd Edn*. New York, NY: McGraw-Hill.

O'Neill, S. (1999). Flow theory and the development of musical performance skills. *Bull. Coun. Res. Music. Educ.* 141, 129–134.

Penhune, V. B. (2011). Sensitive periods in human development: evidence from musical training. *Cortex* 47, 1126–1137. doi: 10.1016/j.cortex.2011.05.010

Petrides, K. V. (2011). "Ability and trait emotional intelligence," in *The Blackwell-Wiley Handbook of Individual Differences,* eds T. Chamorro-Premuzic, A. Furnham, and S. von Stumm, S. (New York, NY: Wiley), 656–678.

Petrides, K. V., and Furnham, A. (2006). The role of trait emotional intelligence in a gender-specific model of organizational variables. *J. Appl. Soc. Psychol.* 36, 552–569. doi: 10.1111/j.0021-9029.2006.00019.x

Petrides, K. V., Niven, L., and Furnham, A. (2006). The trait emotional intelligence of ballet dancers and musicians. *Psicothema* 18(Suppl.), 101–107.

Presson, P. K., and Benassi, V. A. (1996). Locus of control orientation and depressive symptomatology: a meta-analysis. *J. Soc. Behav. Pers.* 11, 201–212.

Privette, G. (1983). Peak experience, peak performance and flow: a comparative analysis of positive human experiences. *J. Pers. Soc. Psychol.* 45, 1361–1368. doi: 10.1037/0022-3514.45.6.1361

Privette, G., and Bundrick, C. M. (1991). Peak experience, peak performance, and flow. *J. Soc. Behav. Pers.* 6, 169–188. doi: 10.1037/0022-3514.45.6.1361

Robinson, J. (2005). *Emotion and its Role in Literature, Music, and Art*. Oxford: Oxford University Press.

Rotter, J. B. (1966). Generalized expectancies for internal versus external control of reinforcement. *Psychol. Monogr-Gen. A.* 80, 1–28. doi: 10.1037/h0092976

Russell, J. A. (1980). A circumplex model of affect. *J. Pers. Soc. Psychol.* 39, 1161–1178. doi: 10.1037/h0077714

Salovey, P., Mayer, J. D., Caruso, D., and Yoo, S. H. (2009). "The positive psychology of emotional intelligence," in *The Oxford Handbook of Positive Psychology*, eds S. J. Lopez and C. R. Snyder (New York: Oxford University Press), 237–248.

Sawyer, R. K. (2006). Group creativity: musical performance and collaboration. *Psychol. Music* 34, 148–165. doi: 10.1177/0305735606061850

Seger, J., and Potts, R. (2012). Personality correlates of psychological flow states in videogame play. *Curr. Psychol.* 31, 103–121. doi: 10.1007/s12144-012-9134-5

Sinnamon, S., Moran, A., and O'Connell, M. (2012). Flow among musicians: measuring peak experiences of student performers. *J. Res. Music Educ.* 60, 6–25. doi: 10.1177/0022429411434931

Sloboda, J. A., Davidson, J. W., Howe, M. J. A., and Moore, D. M. (1996). The role of practice in the development of expert musical performance. *Br. J. Psycho.*87, 287–309. doi: 10.1111/j.2044-8295.1996.tb02591.x

Swann, C. Keegan, R. J. Piggott, D., and Crust, L. (2012). A systematic review of the experience, occurrence, and controllability of flow states in elite sport. *Psych. Sport Exerc.* 13, 807–819. doi: 10.1016/j.psychsport.2012.05.006

Tan, F. B., and Chou, J. P. C. (2010). Dimensions of autotelic personality and their effects on perceived playfulness in the context of mobile information and entertainment services. *Australas. J. Info. Syst.* 17, 5–22.

Teng, C. (2011). Who are likely to experience flow? Impact of temperament and character on flow. *Pers. Individ. Differ*. 50, 863–868. doi: 10.1016/j.paid.2011.01.012

Thompson, W. F., Marin, M. M., and Stewart, L. (2012). Reduced sensitivity to emotional prosody in congenital amusia rekindles the musical protolanguage hypothesis. *Proc. Natl. Acad. Sci. U.S.A.* 109, 19027–19032 doi: 10.1073/pnas.1210344109

Thompson, W. F., Schellenberg, E. G., and Husain, G. (2004). Decoding speech prosody: do music lessons help? *Emotion* 4, 46–64. doi: 10.1037/1528-3542.4.1.46

Thomson, P., and Jaque, V. (2011). "Psychophysiological study: ambulatory measures of the ANS in performing artists," in *Proceedings of International Symposium on Performance Science*, eds A. Williamon, D. Edwards, and L. Bartel (Utrecht: European Association of Conservatoires (AEC)), 149–154.

Trimmer, C. G., and Cuddy, L. L. (2008). Emotional intelligence, not music training, predicts recognition of emotional speech prosody. *Emotion* 8, 838–849. doi: 10.1037/a0014080

Tsay, C.-J. (2013). Sight over sound in the judgment of music performance. *Proc. Natl. Acad. Sci. U.S.A.* 110, 14580–14585. doi: 10.1073/pnas.1221454110

Ullen, F., de Manzano, Ö., Almeida, R., Magnusson, P. K. E., Pedersen, N. L., Nakamura, J., et al. (2012). Proneness for psychological flow in everyday life: associations with personality and intelligence. *Pers. Individ. Differ*. 52, 167–172. doi: 10.1016/j.paid.2011.10.003

Vuust, P., Gebauer, L., Hansen, N. C., Jorgensen, S. R., Moller, A., and Linnet, J. (2010). Personality influences career choice: sensation seeking in professional musicians. *Music Educ. Res*. 12, 219–230. doi: 10.1080/14613801003746584

Watanabe, D., Savion-Lemieux, T., and Penhune, V. B. (2007). The effect of early musical training on adult motor performance: evidence for a sensitive period in motor learning. *Exp. Brain Res.* 176, 332–340. doi: 10.1007/s00221-006-0619-z

Wrigley, W. J., and Emmerson, S. B. (2013). The experience of the flow state in live music performance. *Psychol. Music* 41, 97–118. doi: 10.1177/0305735611418552

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 08 September 2013; accepted: 28 October 2013; published online: 22 November 2013.

Citation: Marin MM and Bhattacharya J (2013) Getting into the musical zone: trait emotional intelligence and amount of practice predict flow in pianists. Front. Psychol. ***4****:853. doi: 10.3389/fpsyg.2013.00853*

This article was submitted to Cognitive Science, a section of the journal Frontiers in Psychology.

ORIGINAL RESEARCH ARTICLE
published: 28 November 2013
doi: 10.3389/fpsyg.2013.00895

Individuality in harpsichord performance: disentangling performer- and piece-specific influences on interpretive choices

Bruno Gingras*[1]*, *Pierre-Yves Asselin*[2] and *Stephen McAdams*[2]

[1] *Department of Cognitive Biology, University of Vienna, Vienna, Austria*
[2] *Schulich School of Music, McGill University, Montreal, QC, Canada*

Edited by:
Lauren Stewart, Goldsmiths University of London, UK

Reviewed by:
Floris T. van Vugt, University of Music, Drama and Media Hannover, Germany
Morwaread M. Farbood, New York University, USA

****Correspondence:***
Bruno Gingras, Department of Cognitive Biology, University of Vienna, Althanstrasse 14, A-1090 Vienna, Austria
e-mail: brunogingras@gmail.com

Although a growing body of research has examined issues related to individuality in music performance, few studies have attempted to quantify markers of individuality that transcend pieces and musical styles. This study aims to identify such meta-markers by discriminating between influences linked to specific pieces or interpretive goals and performer-specific playing styles, using two complementary statistical approaches: linear mixed models (LMMs) to estimate fixed (piece and interpretation) and random (performer) effects, and similarity analyses to compare expressive profiles on a note-by-note basis across pieces and expressive parameters. Twelve professional harpsichordists recorded three pieces representative of the Baroque harpsichord repertoire, including three interpretations of one of these pieces, each emphasizing a different melodic line, on an instrument equipped with a MIDI console. Four expressive parameters were analyzed: articulation, note onset asynchrony, timing, and velocity. LMMs showed that piece-specific influences were much larger for articulation than for other parameters, for which performer-specific effects were predominant, and that piece-specific influences were generally larger than effects associated with interpretive goals. Some performers consistently deviated from the mean values for articulation and velocity across pieces and interpretations, suggesting that global measures of expressivity may in some cases constitute valid markers of artistic individuality. Similarity analyses detected significant associations among the magnitudes of the correlations between the expressive profiles of different performers. These associations were found both when comparing across parameters and within the same piece or interpretation, or on the same parameter and across pieces or interpretations. These findings suggest the existence of expressive meta-strategies that can manifest themselves across pieces, interpretive goals, or expressive devices.

Keywords: music performance, individuality, expressive strategies, harpsichord, interpretation, concordance

INTRODUCTION

Over the last few decades, a growing body of research has examined issues related to individuality in musical performance (e.g., Repp, 1992; see Sloboda, 2000 for a review). Computational methods have led to the development of higher-level descriptors to capture and identify recurrent expressive gestures associated with a given performer (Widmer and Goebl, 2004; Saunders et al., 2008). However, few studies have attempted to quantify markers of individuality that transcend specific pieces and musical styles. Indeed, it seems likely that, among the factors which influence a performer's interpretive choices, some derive from performer-specific tendencies, including kinematic and neuromuscular "fingerprints" (Dalla Bella and Palmer, 2011; Van Vugt et al., 2013), whereas others stem from stylistic considerations related to the piece (or genre) being performed. In order to identify which performance characteristics are reliable markers of a performer's artistic individuality across genres and styles, it is necessary, as a first step, to disentangle these two contributions. Nevertheless, it has proven difficult, for several reasons, to untangle these factors. One obvious issue is that pieces vary in length, texture, and meter. Another issue is that these markers of artistic individuality may plausibly encompass several expressive parameters, such as articulation, velocity, or timing, instead of being restricted to a single expressive device. To identify such expressive "meta-strategies," it is necessary to adopt a statistical approach suitable for analyzing parameters that are measured in different units. Thus, there is a need for a robust methodological approach that allows us to obtain valid statistical inferences even when comparing individual performance profiles across pieces and expressive parameters.

Stamatatos and Widmer (2005) showed, by developing a machine-learning approach based on a set of classifiers that could reliably differentiate among 22 pianists playing two pieces composed by Chopin, that performer-specific characteristics that are not tied to a particular piece could be identified from a symbolic representation (MIDI data) of the expressive parameters associated with each note. More recently, similar methods were successfully applied to the recognition of performers in commercial jazz recordings (Ramirez et al., 2010) and violin recordings (Ramirez et al., 2011) on the basis of the audio signal. In contrast to these studies, which focused mostly on the development of efficient algorithms for the automatic recognition of performers, the present article aims to expand this field of research in a different direction, by developing reliable and statistically rigorous methods for discriminating between piece-specific and performer-specific stylistic influences and for detecting commonalities in expressive patterns across pieces and interpretations.

Although a substantial body of empirical research has focused on piano performance (see Gabrielsson, 2003 for a review), there is a dearth of quantitative studies on expressive strategies in harpsichord performance. However, the study of harpsichord performance is particularly relevant in that it affords an opportunity to compare and extend the findings from piano performance research to other keyboard instruments that may favor different expressive strategies, as well as to musical genres that have been comparatively neglected in performance research. Here, we analyzed a set of recordings of three pieces played by twelve professional harpsichordists on an authentic Italian-style harpsichord equipped with a MIDI console which allowed the precise measurement of performance parameters. The three pieces selected for this study were representative of the Baroque harpsichord repertoire and covered a broad stylistic range: the third variation from the *Partita No. 12 sopra l'aria di Ruggiero* by Girolamo Frescobaldi (1583–1643), the *Prélude non mesuré No. 7*, an unmeasured prelude by Louis Couperin (1626–1661), and *Les Bergeries*, a rondo by François Couperin (1668–1733). The variation from the *Partita No. 12* (hereafter *Partita*) exemplifies the polyphonic, contrapuntal writing of the early Baroque period. The *Prélude non mesuré* (hereafter *Prélude*) belongs to a semi-improvised French harpsichord genre in which the notated score specifies the ordering and pitch height of the notes, but does not indicate measures, nor individual note durations in most cases (including the *Prélude*), thus giving performers more freedom to form their own interpretation and making this a particularly appropriate genre for research on individuality in performance. Finally, the *Bergeries* is typical of the early eighteenth century French harpsichord school, with François Couperin being probably one of its greatest exponents.

Besides examining recordings of three different pieces, we also compared different interpretations of the same piece by the same set of performers. Indeed, performers were invited to record three different interpretations of the *Partita*, each emphasizing a different melodic line (corresponding respectively to the soprano, alto, and tenor parts). This afforded us an opportunity to evaluate the impact of following an explicit interpretive strategy on the expression of individuality in addition to investigating piece-related effects. Four expressive parameters were analyzed for all performances: articulation (corresponding to the amount of overlap between successive notes, from *staccato* to *legato*), note onset asynchrony (defined as the difference in onset time between events that are notated as synchronous in the score), timing (variations in tempo), and velocity (key press velocity). In line with Stamatatos and Widmer (2005), we extracted these expressive parameters from the MIDI data corresponding to the recordings of the performances. As with organ performance (Gingras, 2008; Gingras et al., 2010), the harpsichord affords no or very little timbre differentiation (excluding registration changes), and dynamic differentiation remains limited (Penttinen, 2006). Thus, most of the expressive features available to harpsichordists, such as articulation, onset asynchrony, and tempo variations, involve the manipulation of timing-related parameters, making the study of expressivity in harpsichord performance ideally suited for the type of MIDI-based quantitative analysis that we propose here.

We used two statistical approaches to investigate expressive individuality in harpsichord performance. The first approach consists in analyzing global piece- or performer-specific trends by examining average expressive tendencies over entire performances, whereas the second approach corresponds to a comparison of expressive profiles at the note-by-note level. Both methods provide complementary information when analyzing expressive patterns in performance (Palmer, 1989; Moelants, 2000). With the first approach, we sought to isolate and quantify the influence of the piece being performed (or the interpretive strategy being followed), as well as the impact of the performer's own stylistic individuality, on the average levels associated with each specific expressive parameter. For instance, this method could be used to determine whether there were significant differences in the mean velocity levels associated with different performers, pieces, or interpretations. One drawback of this approach is that, because it focuses on statistically significant differences observed on mean values representing the average level of an expressive parameter for each performance, it is not suitable for analyzing differences in expressive profiles that are only manifested at the note-by-note level, a problem for which our second approach was better suited. Our aim was twofold with this second approach: first, we sought to determine whether we could detect within-piece concordance among the expressive profiles corresponding to different expressive parameters, when considering performances of the same piece (and similarly when comparing performances following the same interpretive goal in the case of the *Partita*). For instance, we wanted to evaluate whether two performers who display similar articulation profiles when playing the same piece also tend to display similar timing profiles, and whether the reverse is also true for performers who display dissimilar expressive profiles. Second, we examined within-parameter concordance across pieces (or interpretations) when considering profiles associated with a single expressive parameter. For example, we investigated whether two performers who display similar articulation profiles when playing one piece also tend to display similar articulation profiles when playing another piece.

The first approach described here corresponds essentially to an analysis of variance, or more generally to a broad category of statistical methods defined as *general linear models*. Here, because we were interested specifically in isolating the contribution of each individual performer (modeled as a *random effect*) and of each piece or interpretive goal (modeled as a *fixed effect*) to the observed variance for each expressive parameter, we used *linear mixed models* (LMMs) to obtain maximum likelihood (ML) estimates of the "piece" (or "interpretation") and "performer" effects (Laird and Ware, 1982; Laird et al., 1987; Lindstrom and Bates, 1988). LMMs are a particularly appropriate statistical tool to address these issues because they can fit a variety of covariance structures and allow for the specification of both random intercepts (i.e., fitting individual intercepts for each performer, corresponding to the overall mean values across all pieces for a given expressive parameter), and random slope effects (fitting individual effects associated with each piece for each performer) (West et al., 2007). Although random slope effects are often neglected, Schielzeth and Forstmeier (2009) have shown that ignoring random slope effects tends to overestimate fixed effects in mixed-model designs.

The second approach outlined above is akin to a *similarity analysis* on expressive profiles. Here, we used the correlation between pairs of expressive profiles as a similarity metric. As a normalized and dimensionless similarity metric, the correlation coefficient is appropriate for comparing variables with different units or scales, such as different expressive parameters, and is especially useful for comparing profiles or sequences (Hubert, 1979). Thus, correlation coefficients are among the most effective measures for detecting similarity in gene expression profiles (Yona et al., 2006), a research question which has many parallels with the similarity analysis of expressive profiles in music performance. Unlike the parametric Pearson correlation coefficient, non-parametric correlation coefficients such as Spearman's rho and Kendall's tau are not sensitive to outliers and are less affected by the shape of the statistical distribution of the data, making them more widely applicable as similarity indices. Indeed, a recent study identified Spearman's rho and Kendall's tau as being among the most effective measures for identifying gene coexpression networks (Kumari et al., 2012). Furthermore, non-parametric correlations were shown to be more efficient than parametric measures for detecting stylistic similarity between texts (Popescu and Dinu, 2009).

In contrast to Spearman's rho which is mathematically equivalent to Pearson's coefficient computed on ranks, Kendall's tau is a measure of concordance, corresponding to the probability of agreement on the sign of the difference between pairs of values (Newson, 2002). Therefore, Kendall's tau is especially useful if the direction of the change between two points is more important than the ranking of the absolute values of the points comprising a given sequence or profile, and has been shown to perform better than either Pearson's or Spearman's coefficients when correlating psychiatric symptom ratings (Arndt et al., 1999) and when comparing the rate and direction of change in ecological communities (Huhta, 1979). Because we were specifically interested in the degree of concordance between performers' expressive patterns in the present study, we chose to use Kendall's tau correlation coefficient to assess the pairwise similarity between expressive profiles. These pairwise correlations were then used to generate similarity matrices, calculated separately for each expressive parameter and for each piece (and for each interpretation in the case of the *Partita*). Comparisons were first conducted to assess within-piece concordance between similarity matrices computed for all expressive parameters obtained from a single piece. In a second step, similarity matrices computed for all three pieces on the same expressive parameter were compared to assess the degree of within-parameter concordance between expressive profiles associated with different pieces. The same procedure was then repeated to compare different interpretations of the *Partita*. Lastly, to evaluate the impact of the choice of correlation coefficient on our results, we compared the outcomes of similarity analyses employing Spearman's rho vs. Kendall's tau as a similarity metric.

RESULTS

LINEAR MIXED MODEL ANALYSES

Comparisons across pieces

For each of the four expressive parameters (articulation, asynchrony, timing, and velocity), mean values were computed over each performance, separately for each piece (see section Performance Data Analysis in Materials and Methods for computational details). All the analyses of variance presented in this section were conducted on the mean values thus obtained (shown in **Figure 1**). LMMs were built using the step-up approach (Snijders and Bosker, 1999; Raudenbush and Bryk, 2002), beginning with an unconditional means model with only intercepts for fixed and random effects. For the purpose of conducting comparisons across pieces, we retained only the *Partita* recordings emphasizing the highest melodic line (soprano). Repeated-measures LMMs were used because each piece was recorded twice by each performer, with individual performers (12) treated as random effects and pieces (3) treated as a fixed effect. The potential effect of repetition (comparing the first and second recordings of each piece), as well as the interaction between piece and repetition, were also considered as fixed effects. Note that the models for asynchrony did not include the *Prélude* whose score does not include any note onsets notated as synchronous. Furthermore, the effect of piece was not considered in the case of timing given that durations were zero-centered for each piece to allow for meaningful comparisons across pieces (see section Performance Data Analysis in Materials and Methods).

Fixed effects were first added to the models, followed by random effects. Both random intercepts and random slope effects were considered. At each step, the improvement to the fit of the model was assessed by likelihood tests using ML estimation when comparing models that differed only in the specification of the fixed effects, and restricted maximum likelihood (REML) estimation when comparing models that differed only in the specification of the random effects (Morrell, 1998; Verbeke and Molenberghs, 2000). The following paragraphs outline the model building steps. Detailed tests of significance are only provided for the final models (see **Table 1**) since all further analyses were conducted on the final models. However, a summary of the *p*-values obtained during the model-building steps is given below where relevant.

In comparison to the baseline model including only intercepts for fixed and random effects, the addition of a fixed effect of piece significantly improved the fit of the models for articulation and velocity ($p < 0.001$ in both cases), but was only marginally significant in the case of asynchrony ($p = 0.08$). The effect of piece was nevertheless included in all three models to facilitate comparisons

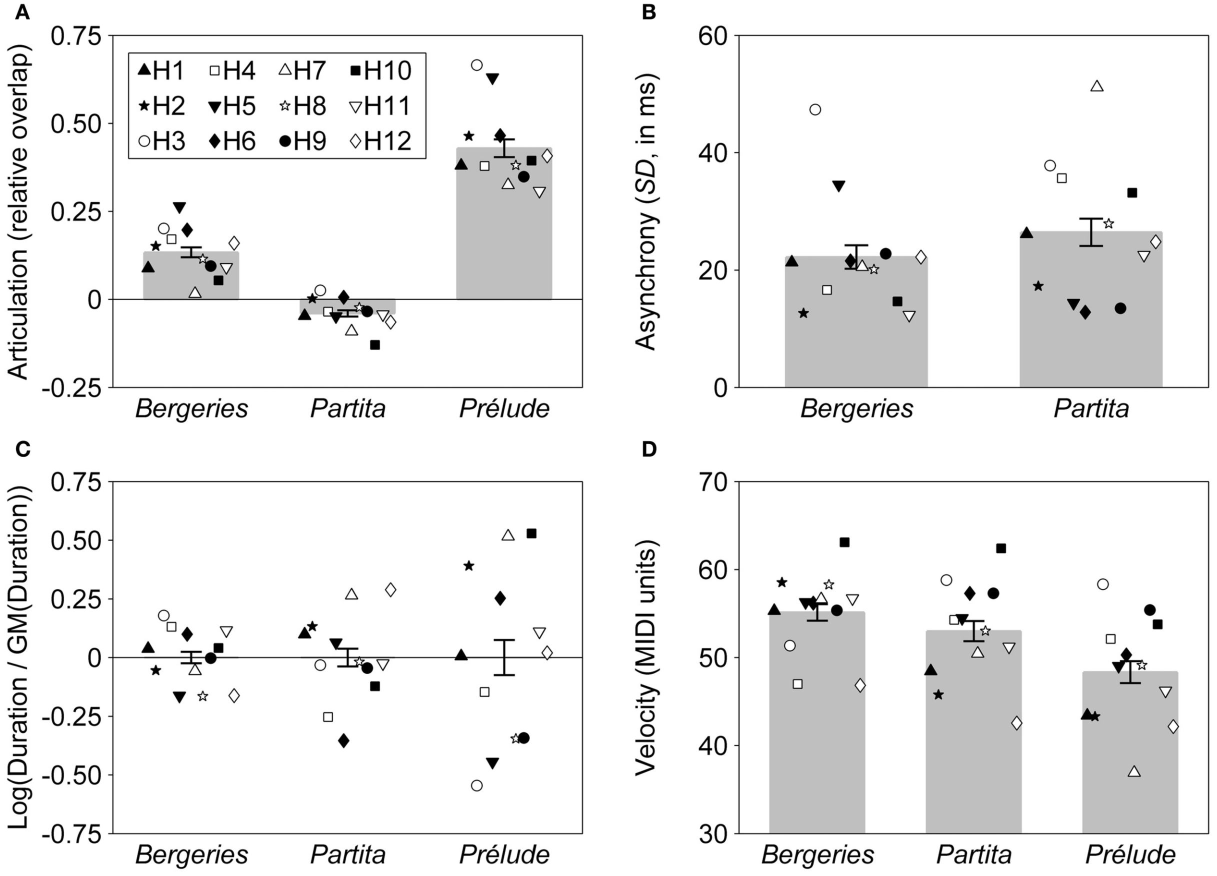


FIGURE 1 | Mean values for each of the four expressive parameters, for all three pieces. Each individual harpsichordist (identified as H1, H2, …, H12) is represented by a unique symbol. Each symbol represents the average of two recordings by the same performer. Error bars represent the standard error of the mean. **(A)** Articulation, measured as relative overlap (negative values correspond to a detached articulation and positive values to a *legato* articulation). **(B)** Asynchrony, measured as the standard deviation of onset times for nominally synchronous notes (in milliseconds). **(C)** Timing, measured as the logarithm of the ratio of the duration of the piece to the geometric mean (GM) of the duration of all performances of the same piece. **(D)** Velocity, measured in MIDI units (16–100).

between models (Cheng et al., 2009). On the other hand, adding the effect of repetition or the interaction between piece and repetition did not improve the fit of the models (all p-values > 0.41 for repetition, and all p-values > 0.27 for the interaction between piece and repetition). Therefore, the models obtained at the end of this step incorporated a fixed effect of piece (except in the case of timing) and a random intercept.

In a second step, random effects were added. In order to ascertain that random effects, corresponding to individual effects associated with each performer, were significant, we first compared the fit of the models obtained at the end of the first step with equivalent models including only fixed effects (no random intercept). Indeed, models including a random intercept fitted the data significantly better than models incorporating only fixed effects (all p-values < 0.05). Subsequently, the inclusion of a random effect of piece was also considered. Adding a random effect of piece improved the fit for all models (all p-values < 0.01), leading to our final models, which included a fixed effect of piece, a random intercept, and a random effect of piece (**Table 1**). Note that in the case of the models for asynchrony and timing, the inclusion of a random piece effect resulted in a non-significant random intercept, suggesting that most of the between-performers variance observed for these two expressive parameters was captured by the random piece effect (we will revisit this point below). Nevertheless, the random intercept was kept in all final models in order to facilitate comparisons between models.

Finally, we sought to directly quantify the variance explained by the fixed (piece) and random (performer) effects in our models. In contrast to traditional general linear models, there is no standard formula for computing the proportion of variance (R^2) explained by the various parameters of a linear mixed model. In this paper, we use a promising approach for estimating R^2 in generalized LMMs (GLMMs, which include LMMs) that was proposed by Nakagawa and Schielzeth (2013). This method can be used to obtain the proportion of variance explained by the fixed effects in a model [defined as "marginal" R^2, or $R^2_{\mathrm{GLMM}(m)}$ in Nakagawa and Schielzeth's notation], and the proportion of variance explained by both fixed and random effects ["conditional"

Table 1 | Linear mixed models comparing across recordings of the three pieces.

Expressive parameter	Fixed effects	Random effects (performer)		R^2_{GLMM}	
	Piece	Intercept (overall mean)	Piece (slope)	Marginal (fixed)	Conditional (fixed and random)
Articulation	$F_{(2,\ 22)} = 223.05$, $\mathbf{p < 0.001}$	$\chi^2(1) = 8.76$, $\mathbf{p = 0.003}$	$\chi^2(1) = 7.75$, $\mathbf{p = 0.005}$	0.836	0.920
Asynchrony†	$F_{(1,\ 11)} = 1.02$, $p = 0.335$	$\chi^2(1) = 0.10$, $p = 0.756$	$\chi^2(1) = 58.31$, $\mathbf{p < 0.001}$	0.038	0.426
Timing*	N/A	$\chi^2(1) = 0.001$, $p = 0.978$	$\chi^2(1) = 95.59$, $\mathbf{p < 0.001}$	N/A	0.197
Velocity	$F_{(2,\ 22)} = 7.85$, $\mathbf{p = 0.003}$	$\chi^2(1) = 4.82$, $\mathbf{p = 0.028}$	$\chi^2(1) = 78.85$, $\mathbf{p < 0.001}$	0.210	0.625

*The significance of the fixed effect of piece was assessed with Type III F-tests conducted on the final models, whereas the significance of the random intercept and slope effects was assessed with likelihood tests using REML estimation. Statistically significant p-values are indicated in bold. For each expressive parameter, the corresponding marginal and conditional R^2_{GLMM} values were computed on a random-intercept model that was equivalent to the final model but with the random slope effect excluded (see Nakagawa and Schielzeth, 2013). †Asynchrony values were not computed for the Prélude, whose score does not include notes that should be played together. *The fixed effect of piece was not considered for timing, given that all values were zero-centered for each piece to allow for meaningful comparisons across pieces.*

R^2, notated as $R^2_{\text{GLMM(c)}}$]. The proportion of variance explained by random effects alone can be estimated by comparing both quantities. Note that Nakagawa and Schielzeth's formula does not account for random slope effects (here, random piece effects). However, R^2 values obtained for random-slope models are usually very similar to those obtained for analogous random-intercept models when the same fixed effects are fitted (Snijders and Bosker, 1999). Therefore, we have followed Nakagawa and Schielzeth's (2013) suggestion of computing R^2_{GLMM} values for random-slope models on analogous random-intercept models. The R^2 values reported in **Table 1** are thus only an approximation of the R^2 values for the final models, which include a random slope effect.

A comparison of the marginal R^2 values obtained for the different expressive parameters shows that the fixed effect of piece was dominant in the case of articulation, explaining more than 80% of the total variance, suggesting that the overall articulation pattern (detached or *legato*) was mostly a function of the specific piece to be performed, with performer-associated effects playing only a minor role (**Figure 1A**). On the other hand, the fixed piece effect had only a moderate influence on velocity (**Figure 1D**) and was negligible in the case of asynchrony (**Figure 1B**). Random effects (individual differences between performers), which are discussed in greater detail below, played a much larger role for these two expressive parameters than for articulation.

Post-hoc tests (pairwise comparisons, all *p*-values Bonferroni-corrected) were conducted for articulation and velocity in order to compare the estimated marginal means for each piece. In the case of articulation, pairwise comparisons showed that the *Prélude* was played significantly more *legato* than both the *Bergeries*, $t_{(1,\ 22)} = 13.15$, $p < 0.001$, and the *Partita*, $t_{(1,\ 22)} = 20.89$, $p < 0.001$. The *Bergeries* was also played significantly more *legato* than the *Partita*, $t_{(1,\ 22)} = 7.73$, $p < 0.001$, giving the following ordering from more detached to more *legato* articulation: *Partita* < *Bergeries* < *Prélude* (**Figure 1A**). Regarding velocity, the *Prélude* was played with significantly less velocity than both the *Partita*, $t_{(1,\ 22)} = 2.66$, $p = 0.043$, and the *Bergeries*, $t_{(1,\ 22)} = 3.87$, $p = 0.003$, with no significant difference between the latter two (**Figure 1D**).

Statistically significant random intercepts correspond to a systematic tendency by some performers to display a given expressive feature to a lesser or greater extent than their colleagues, across all pieces. For the four expressive parameters surveyed here, significant random intercepts were only found for articulation and velocity, corresponding to a systematic tendency by some performers to play more detached (or more *legato*), or with a smaller (or greater) velocity than their colleagues, across all pieces (**Figures 1A,D**). On the contrary, the non-significant random intercepts for the timing and asynchrony models indicate that none of the performers in our sample tended to play significantly slower or faster, or with a lesser or greater degree of asynchrony, than their colleagues when averaging across all pieces (**Figures 1B,C**).

A significant random piece effect indicates that the effect associated with a given piece is not uniform across all performers, or, in other words, that different performers respond differently to a given piece. Significant random piece effects were found for all four expressive parameters, with large effects in the case of asynchrony, timing, and velocity. The weaker random piece effect for articulation is linked to the strong fixed effect observed for this parameter, which suggests that performers tended to respond more uniformly to piece effects in the case of articulation than for other parameters such as asynchrony and velocity, for which the magnitude of the fixed effect was comparatively smaller.

LMMs allow random effects to be predicted for individual performers (Littell et al., 2006, chapter 8). A summary of the significant intercept and piece random effects at the performer

Table 2 | Individual random effects associated with each performer.

Expressive parameter	Intercept (overall mean)	Piece (slope)		
		Bergeries	***Partita***	***Prélude***
Articulation ($df = 36$)	Detached: H7* *Legato*: H3**,H5*	n.s.	n.s.	*Legato*: H3**, H5*
Asynchrony† ($df = 24$)	n.s.	More: H3***, H5**	Less: H5*, H6**, H9* More: H7***	N/A
Timing ($df = 36$)	n.s.	Slower: H3**, H4*, H11* Faster: H5**, H8**, H12**	Slower: H2*, H7***, H12*** Faster: H4***, H6***, H10*	Slower: H2***, H6***, H7***, H10***, H11* Faster: H3***, H4*, H5***, H8***, H9***
Velocity ($df = 36$)	Less: H12* More: H10*	Less: H3*, H4** More: H2*	Less: H2*, H12*	Less: H7** More: H3**

Individual performers are identified by codes H1 to H12. The significance of the random intercept and piece effects predicted for each individual performer was assessed using two-tailed t-tests. The denominator degrees of freedom are indicated for each expressive parameter in the leftmost column. $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$; n.s., no significant effect. †Asynchrony values were not computed for the Prélude, whose score does not include notes that should be played together.

level is provided in **Table 2**, with individual harpsichordists identified by codes H1 to H12. In line with the results reported previously, no significant random intercepts were found for asynchrony and timing, and only two performers showed significant random piece effects for articulation. Significant random piece effects were especially prevalent for timing, notably in the case of the *Prélude*, suggesting a greater degree of individual variability in the choice of tempi (**Figure 1C**). Furthermore, we can also see in **Table 2** that some performers displayed a greater degree of expressive individuality, as indicated by a large number of significant random effects, than others who showed few or no significant effects (see also **Figure 1**). For instance, significant effects were associated with performer H3 for all four expressive parameters, but no effects reached significance for performer H1.

Finally, to control for the fact that all the between-pieces comparisons conducted here employed the interpretation of the *Partita* emphasizing the highest melodic line (soprano), we repeated the LMM analyses described above using the interpretations of the *Partita* emphasizing the alto and tenor parts in turn. We obtained very similar results to those shown in **Table 1**, both for the *F*-tests on the fixed piece effect and for the likelihood tests on the random intercept and piece effects, with identical outcomes for the significance tests and similar *F* ratios and chi-square values in all cases. This result suggests that the choice of the interpretation of the *Partita* for the purpose of conducting comparisons across pieces had very little bearing on the results of the analyses presented here.

Comparisons across interpretations of the Partita

Because performers recorded three different interpretations of the *Partita*, we also analyzed the contribution of the interpretive goal and of performers' individual specificities to the variance observed on the mean values for each of the four expressive parameters across interpretations of the *Partita*. Following the procedure described in the preceding section, repeated-measures LMMs were built using the step-up approach, beginning with an unconditional means model with only intercepts for fixed and random effects, treating individual performers (12) as random effects and interpretations (3) as a fixed effect. Once again, repetition (comparing the first and second recordings of each interpretation), as well as the interaction between interpretation and repetition, were considered as fixed effects. Given that the timing comparisons were conducted across interpretations of the same piece in this case, we used the untransformed durations of the performances here (see Performance Data Analysis in Materials and Methods).

In comparison to the baseline model including only intercepts for fixed and random effects, the addition of a fixed effect of interpretation significantly improved the fit of the model for asynchrony ($p = 0.04$) and marginally for articulation ($p = 0.10$), but not for either timing or velocity (both p-values > 0.16). The effect of interpretation was nevertheless included in all four models to facilitate comparisons between models. In contrast to what was observed when comparing across pieces, adding a fixed effect of repetition significantly improved the fit of the model for asynchrony ($p = 0.01$), but not for the other parameters (all other p-values > 0.19). Again, the effect of repetition was added to all four models. The addition of the interaction between interpretation and repetition did not significantly improve the fit of any model (all p-values > 0.11). Thus, the models obtained at the end of this step incorporated fixed effects of interpretation and repetition as well as a random intercept.

Random effects were then examined. We confirmed that models including a random intercept fitted the data significantly better than models incorporating only fixed effects (all p-values < 0.001). Subsequently, the inclusion of a random effect of interpretation was also considered. Adding a random effect of interpretation improved the fit for articulation and timing (both

p-values < 0.01), but not for asynchrony or velocity (both p-values > 0.12). The random effect of interpretation was included in all four models. Because a fixed effect of repetition was included in the models, we also considered a random effect of repetition, but its addition did not improve the fit of any models (all p-values > 0.12). Hence, our final models included fixed effects of interpretation and repetition, a random intercept, and a random effect of interpretation (**Table 3**).

R^2_{GLMM} values were computed following the procedure described in the previous section. Fixed effects explained only a small proportion of the variance, even for the expressive parameters for which these effects were significant or marginally significant, such as articulation and asynchrony. However, the conditional R^2_{GLMM} values were very high, with all four models explaining more than 80% of the variance. The very large proportion of variance explained by random effects for models comparing across interpretations implies that performer-related specificities could account for most of the observed differences in the mean values of the expressive parameters.

The significant effect of repetition observed in the case of asynchrony corresponded to a tendency by performers to play the second recording of each interpretation with smaller asynchronies than the first (**Figure 2B**). Similarly, a marginal tendency to play the second recording more *legato* was observed (**Figure 2A**). To further investigate the effect of repetition in the comparisons across interpretations of the *Partita*, we considered the possibility that the repetition effect was a learning effect, and that performers were still getting accustomed to each interpretation. We thus analyzed the error rates using a GLMM that models the frequency of score errors as a function of the interpretation and the repetition, using a logit (binomial) distribution. This GLMM corresponded to a repeated-measures logistic regression with interpretation and repetition as fixed effects, and random intercept as well as random effect of interpretation, and was thus analogous to the LMMs presented in **Table 3**. Although error rates were slightly lower for the second repetition (0.69% on average, vs. 0.82% for the first repetition), neither the effect of repetition, $F_{(1,\ 35)} = 0.87$, $p = 0.36$, nor the effect of interpretation, $F_{(2,\ 22)} = 0.49$, $p = 0.62$, were close to reaching significance.

Although large statistical effects were associated with the random intercepts for all expressive parameters, significant random interpretation effects were only found for articulation and timing. Random interpretation effects were generally smaller than the effects observed for the random intercepts, as indicated by the relative magnitude of the chi-square values obtained with likelihood tests (**Table 3**). In line with these results, very few random interpretation effects associated with individual performers were observed. In fact, only one such effect reached significance across all performers and expressive parameters, corresponding to performer H12 playing the "alto" interpretation with a significantly slower tempo. In contrast, a large number of significant random intercepts associated with individual performers were observed. Notably, most performers who exhibited a tendency to play significantly more detached (H7) or more *legato* (H3), or with less (H12) or more (H10) velocity than their colleagues when comparing across pieces (see **Table 2**), also displayed the same tendencies when comparing across interpretations of the *Partita*. One exception was H5 who showed a significant tendency to play more *legato* across all three pieces, but not across interpretations of the *Partita* (**Figure 2A**).

Discussion

In contrast with the LMMs comparing expressive parameters across pieces, for which important fixed effects were found for articulation and velocity, the proportion of the variance explained by fixed effects was very low for the LMMs comparing interpretations of the *Partita*. This suggests that systematic interpretation-related (or repetition-related) differences between interpretations emphasizing different melodic lines were, for the most part, relatively unimportant when comparing mean values computed on

Table 3 | Linear mixed models comparing across interpretations of the *Partita*.

Expressive parameter	Fixed effects		Random effects (performer)		R^2_{GLMM}	
	Interpretation	Repetition	Intercept (overall mean)	Interpretation (slope)	Marginal (fixed)	Conditional (fixed and random)
Articulation	$F_{(2,\ 22)} = 1.32$, $p = 0.287$	$F_{(1,\ 35)} = 3.25$, $p = 0.080$	$\chi^2(1) = 28.43$, $\boldsymbol{p < 0.001}$	$\chi^2(1) = 11.06$, $\boldsymbol{p < 0.001}$	0.016	0.823
Asynchrony	$F_{(2,\ 22)} = 2.85$, $p = 0.079$	$F_{(1,\ 35)} = 7.44$, $\boldsymbol{p = 0.010}$	$\chi^2(1) = 48.81$, $\boldsymbol{p < 0.001}$	$\chi^2(1) = 1.23$, $p = 0.267$	0.018	0.905
Timing	$F_{(2,\ 22)} = 0.06$, $p = 0.944$	$F_{(1,\ 35)} = 0.40$, $p = 0.531$	$\chi^2(1) = 68.85$, $\boldsymbol{p < 0.001}$	$\chi^2(1) = 8.24$, $\boldsymbol{p = 0.004}$	<0.001	0.970
Velocity	$F_{(2,\ 22)} = 1.27$, $p = 0.300$	$F_{(1,\ 35)} = 0.24$, $p = 0.625$	$\chi^2(1) = 58.51$, $\boldsymbol{p < 0.001}$	$\chi^2(1) = 2.41$, $p = 0.120$	0.003	0.942

The significance of the fixed effects of interpretation and repetition was assessed with Type III F-tests conducted on the final models, whereas the significance of the random intercept and slope effects was assessed with likelihood tests using REML estimation. Statistically significant p-values are indicated in bold. For each expressive parameter, the corresponding marginal and conditional R^2_{GLMM} values were computed on a random-intercept model that was equivalent to the final model but with the random slope effect excluded (see Nakagawa and Schielzeth, 2013).

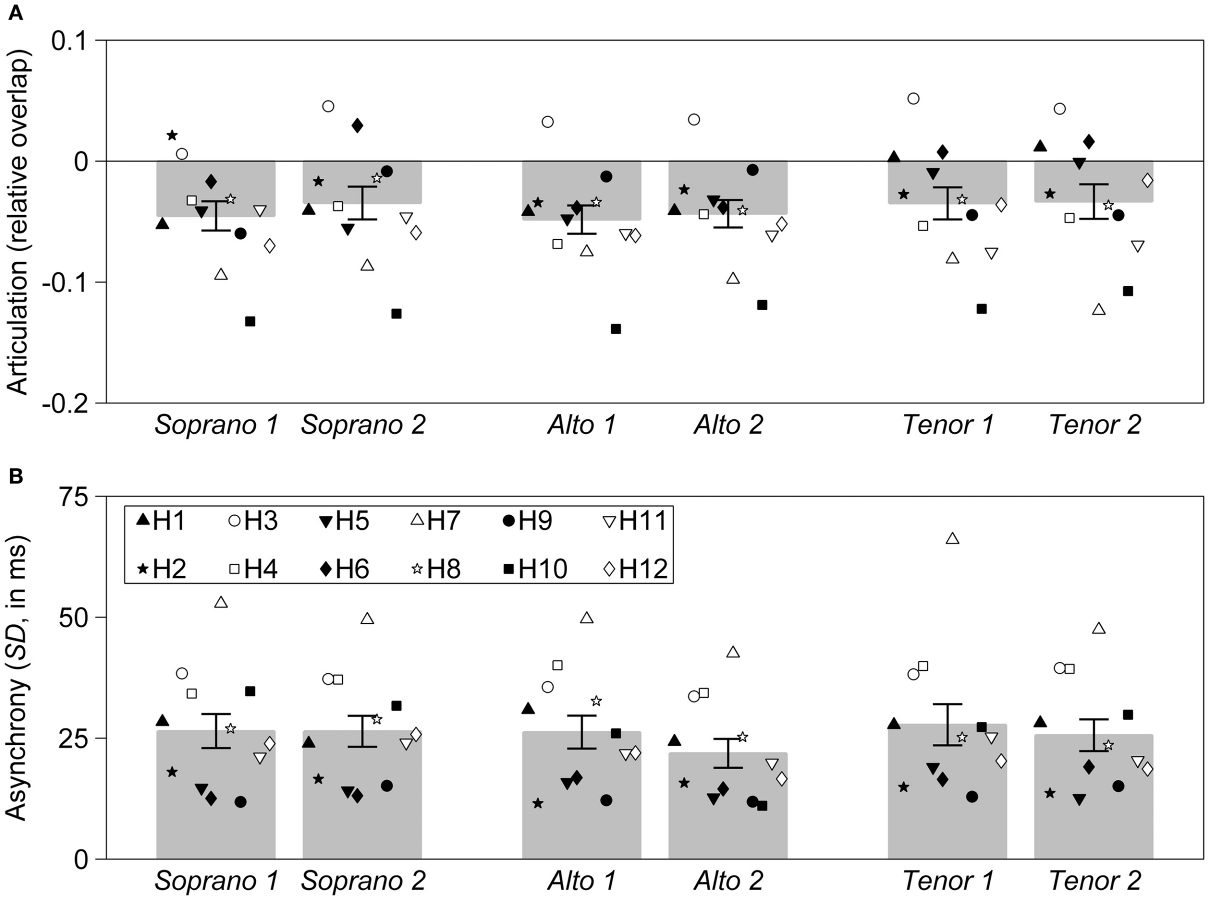


FIGURE 2 | Mean values for articulation and asynchrony, for all three interpretations of the *Partita*. Each individual harpsichordist (identified as H1, H2, ..., H12) is represented by a unique symbol. Each symbol represents a single recording. Three interpretations, each emphasizing a different melodic line (corresponding to the soprano, alto, or tenor part) were recorded. Each interpretation was recorded twice, with successive recordings indicated by the number "1" or "2." Error bars represent the standard error of the mean. **(A)** Articulation, measured as relative overlap (negative values correspond to a detached articulation and positive values to a *legato* articulation). **(B)** Asynchrony, measured as the standard deviation of onset times for nominally synchronous notes (in milliseconds).

the entire performances. To be sure, this does not imply that there were no significant differences between these interpretations, but analyzing these differences requires a finer approach which involves considering each melodic line in isolation (Gingras et al., 2009). On the other hand, random effects explained a much larger proportion of the variance for the LMMs comparing across interpretations of the *Partita* than for the LMMs comparing across pieces (even though these random effects were non-negligible when accounting for the variance in asynchrony, timing, or velocity across pieces). This result indicates that individual specificities tended to dominate when considering interpretations of the same piece, but were relatively less important when examining different pieces.

The significant effects associated with repetition (i.e., comparing the first and second recordings) in the LMMs on the interpretations of the *Partita* were somewhat unexpected, because repetition was not a significant factor in any of the LMMs modeling expressive parameters across pieces. Adding repetition as a fixed effect to these LMMs did not increase the R^2_{GLMM} values for any of the models. The overall low error rates, as well as the absence of a significant difference in error rates between repetitions or interpretations, suggest that performers were comfortable with each interpretation at the time of recording and do not argue in favor of a learning effect. Nevertheless, it is possible that changing between interpretations of the same piece during the recording session demanded more flexibility on the part of the performers than changing from one piece to another. This may explain why asynchronies were slightly but significantly smaller, and articulations slightly more *legato* (albeit with small effect sizes in both cases), on the second recording of each interpretation as performers were adjusting to the character of each interpretation.

Whereas the magnitude of the random piece effects was generally larger than that of the random intercept effects when comparing across pieces (see **Table 1**), the opposite was observed when comparing across interpretations of the *Partita* (see **Table 3**). This suggests that, whereas individual performers exhibited markedly different responses to the three pieces, individual responses to the three interpretations of the *Partita* were not as differentiated. On the other hand, performers who tended to play consistently more *legato*, or with a faster tempo, tended to do so for all three interpretations of the *Partita* (as indicated by the large random intercept effects reported in **Table 3**), whereas this performer-specific consistency was somewhat less pronounced when comparing across pieces (as indicated by the small to moderate random intercept effects shown in **Table 1**).

SIMILARITY ANALYSES ON EXPRESSIVE PROFILES

Comparisons across pieces

Kendall's tau correlation coefficients were calculated between the expressive profiles of all pairs of performers, separately for each parameter and for each piece. For the purpose of conducting comparisons across pieces, we retained only the *Partita* recordings emphasizing the highest melodic line (soprano). To avoid pseudoreplication, correlation coefficients were computed on the expressive profiles corresponding to the average of the two performances recorded by each performer for each piece (note that very similar results were obtained by averaging the correlations obtained on each of the two performances instead of computing the correlations on the averaged profiles). Correlation coefficients were computed on a note-by-note basis in the case of articulation and velocity, and on an event-by-event basis in the case of timing and asynchrony. Similarity matrices were then generated by computing all possible pairwise Kendall's taus between the 12 performers' note-by-note (or event-by-event) expressive profiles, separately for each parameter and for each piece. Eleven 12 × 12 similarity matrices were obtained, four each for the *Bergeries* and the *Partita* (one for each expressive parameter), and three for the *Prélude* for which no asynchrony patterns were extant. All correlation coefficients were positive, indicating a higher-than-chance concordance between expressive profiles (the statistical significance of each coefficient is not reported here due to the very large number of correlations, and because the aim of this analysis was not to test the significance of each pairwise correlation but to examine the global concordance between similarity matrices).

Two series of comparisons were conducted between the similarity matrices thus obtained. First, to test for within-piece profile concordance across expressive parameters, we assessed the degree of congruence between the groups of similarity matrices corresponding to all expressive parameters analyzed for a single piece. Second, to test for within-parameter profile concordance across pieces, we assessed the degree of congruence between the groups of similarity matrices corresponding to a single expressive parameter analyzed over all pieces.

To control for familywise error rates, the CADM ("Congruence among distance matrices") test (Legendre and Lapointe, 2004), which detects congruence in a group of matrices, was first applied to each group of similarity matrices that was tested separately. If the chi-square statistic obtained by the CADM test was significant (as determined by a permutation test), indicating congruence in a group of matrices, *post-hoc* tests were conducted to identify the matrix (or matrices) which explained this association, following Legendre and Lapointe (2004). The Bonferroni-Holm correction (Holm, 1979), a sequential procedure which is less conservative than the classic Bonferroni correction, was applied to the *p*-values thus obtained. Lastly, the Mantel test, a non-parametric permutation test which evaluates the degree of association between two matrices (Mantel, 1967; Legendre and Legendre, 1998) and is applicable to either distance matrices or similarity matrices (Dietz, 1983), was used to determine the pairwise rank correlation (Spearman's rho) between the similarity matrix (or matrices) identified as significantly congruent in the *post-hoc* procedure and other matrices in the group. Note that, by design, both the CADM and Mantel tests ignore the main diagonal of the matrices, meaning that all the comparisons presented here were strictly conducted between expressive profiles corresponding to different performers. The number of degrees of freedom does not affect the probability values obtained by permutation tests (McArdle and Anderson, 2001) and is not reported for the CADM and Mantel tests (see Legendre, 2000).

CADM tests were first conducted to assess the within-piece congruence between the similarity matrices corresponding to the four expressive parameters (only three in the case of the *Prélude*), separately for each piece. A significant association was detected for the *Prélude*, $\chi^2 = 94.95$, $p = 0.020$. *Post-hoc* tests revealed that the timing similarity matrix was significantly congruent with at least one other matrix in the group (Bonferroni-Holm corrected *p*-value < 0.001). Mantel tests showed a significant correlation between the matrices for timing and articulation ($r = 0.421$, $p = 0.009$), indicating that the magnitude of the pairwise correlations computed between all pairs of performers on the timing profiles was positively correlated with the magnitude of the corresponding pairwise correlations computed on the articulation profiles (**Figure 3A**; see also **Figure 3B** for a visual representation of a non-significant association between the timing and velocity pairwise correlations for the *Bergeries*). In other words, there was a significant tendency for performers with concordant timing patterns to show concordant articulation patterns. The correlation between the similarity matrices for timing and velocity also reached significance ($r = 0.347$, $p = 0.032$), corresponding to a tendency for performers with concordant timing profiles to also display concordant velocity profiles (**Figure 3C**). Furthermore, the CADM tests were also marginally significant for the *Bergeries* ($\chi^2 = 84.40$, $p = 0.076$) and the *Partita* ($\chi^2 = 85.31$, $p = 0.085$), suggesting weak or partial congruence in both cases (no *post-hoc* tests were conducted here since the tests did not reach significance).

Second, CADM tests were conducted to assess the within-parameter congruence among the similarity matrices based on a single expressive parameter across all pieces, separately for each of the four parameters. A significant association was detected for articulation ($\chi^2 = 98.33$, $p = 0.010$) and for timing ($\chi^2 = 88.29$, $p = 0.041$), but not for asynchrony or velocity (both *p*-values > 0.27). For articulation, *post-hoc* tests revealed that the articulation similarity matrix for the *Partita* was congruent with at least one other matrix (Bonferroni-Holm corrected *p*-value $= 0.040$). However, the corrected *p*-values for the matrices corresponding to the *Bergeries* and the *Prélude* were both marginally significant, suggesting that the articulation similarity matrices for all three pieces were at least partially congruent. Mantel tests showed a significant correlation between the matrices for the *Bergeries* and the *Partita* ($r = 0.314$, $p = 0.038$), indicating that the magnitude of

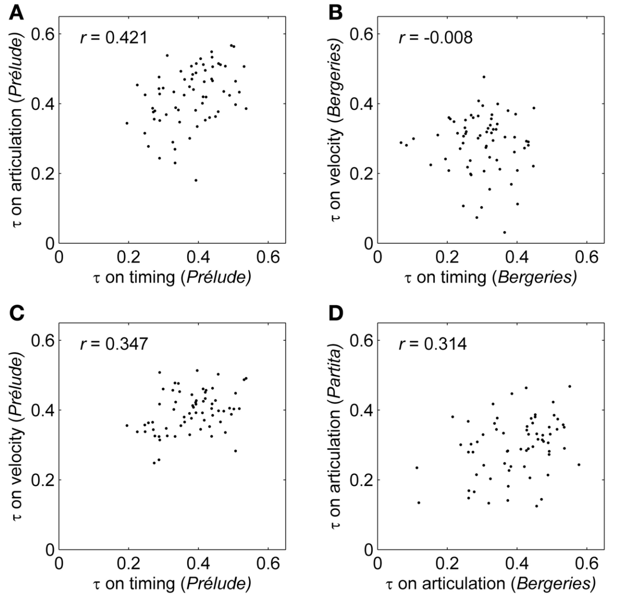


FIGURE 3 | Concordance between the pairwise correlations computed on expressive profiles. τ: Kendall's tau correlation coefficient computed on expressive profiles between pairs of performers. *r*: Mantel correlation coefficient between similarity matrices. Each dot corresponds to the pairwise correlation between two performers. **(A)** Timing and articulation pairwise correlations for the *Prélude*. **(B)** An example of a non-significant association between the timing and velocity pairwise correlations for the *Bergeries*. **(C)** Timing and velocity pairwise correlations for the *Prélude*. **(D)** Pairwise correlations on the articulation profiles for the *Bergeries* and the *Partita*.

the pairwise correlations computed between the *Bergeries* articulation profiles for all pairs of performers was correlated with the magnitude of the corresponding pairwise correlations computed on articulation profiles for the *Partita* (**Figure 3D**). A marginally significant correlation was also observed between the articulation similarity matrices for the *Partita* and the *Prélude* ($r = 0.241$, $p = 0.072$). For timing, *post-hoc* tests revealed that the timing similarity matrix for the *Prélude* was congruent with at least one other matrix (Bonferroni-Holm corrected *p*-value = 0.045), and Mantel tests showed a significant correlation between the timing matrices for the *Prélude* and the *Partita* ($r = 0.376$, $p = 0.033$).

Finally, to control for the fact that the comparisons across pieces employed the interpretation of the *Partita* emphasizing the highest melodic line (soprano), we repeated these analyses using the interpretations of the *Partita* which emphasized the alto and tenor parts, respectively. We obtained similar results to those described in the previous paragraph, with identical outcomes for the CADM tests in practically all cases. Exceptions were the CADM test on asynchrony, which was marginally significant with the alto interpretation ($\chi^2 = 82.22, p = 0.092$) but not with other interpretations (all other *p*-values > 0.27), and the CADM test on timing, which reached significance with either the soprano or tenor interpretations (both *p*-values < 0.05), but was only marginally significant with the alto interpretation ($\chi^2 = 83.28$, $p = 0.077$). This suggests that the choice of the interpretation of the *Partita* for the purpose of conducting comparisons across pieces had only a minor influence on the outcome of the similarity analyses.

Comparisons across interpretations of the Partita

Following the procedure described above, similarity matrices were generated by computing all possible pairwise Kendall's taus between the 12 performers' note-by-note (or event-by-event) expressive profiles, separately for each parameter and for each interpretation of the *Partita*. With very few exceptions for which slightly negative values were obtained (corresponding to 3 out of 792 pairwise correlations), all Kendall's taus were positive, indicating a higher-than-chance concordance between expressive profiles. Twelve 12 × 12 similarity matrices were obtained, for each of the four expressive parameters and each of the three interpretations.

CADM tests were first conducted to assess the within-interpretation congruence among the similarity matrices corresponding to the four expressive parameters, separately for each interpretation. As reported in the previous section, a marginal tendency was found for the soprano interpretation ($\chi^2 = 85.31$, $p = 0.085$). Additionally, a significant association was detected for the alto ($\chi^2 = 97.22$, $p = 0.026$) and tenor ($\chi^2 = 102.01$, $p = 0.005$) interpretations. In the case of the alto interpretation, *post-hoc* tests revealed that the asynchrony and timing similarity matrices were congruent with at least one other matrix (both Bonferroni-Holm corrected *p*-values < 0.01). Mantel tests showed a significant correlation between the asynchrony and timing matrices ($r = 0.558, p = 0.001$) and between the velocity and timing matrices ($r = 0.373, p = 0.009$). For the tenor interpretation, *post-hoc* tests indicated that the timing matrix was congruent with at least one other matrix (Bonferroni-Holm corrected *p*-value < 0.001). Mantel tests showed that the timing matrix was significantly correlated with the articulation ($r = 0.409$, $p = 0.022$), asynchrony ($r = 0.391$, $p = 0.007$), and velocity ($r = 0.283, p = 0.022$) matrices.

CADM tests were then conducted to assess the within-parameter congruence among the similarity matrices based on a single expressive parameter across all interpretations, separately for each of the four parameters. The CADM tests were highly significant for all parameters (all $\chi^2 > 125$, all *p*-values < 0.001). *Post-hoc* tests revealed that all matrices corresponding to the same expressive parameter were congruent with each other (all Bonferroni-Holm corrected *p*-values < 0.01). Similarly, Mantel tests indicated that all pairwise correlations conducted between similarity matrices corresponding to the same parameter were significant (all $r > 0.39$, all *p*-values < 0.01).

Comparison between Kendall's tau and Spearman's rho

In order to evaluate whether the choice of non-parametric correlation coefficient affected the outcome of the similarity analyses reported in the preceding sections, we repeated all analyses using Spearman's rho correlation coefficient instead of Kendall's tau. The CADM tests conducted on the similarity matrices generated using Spearman's rho coefficients yielded chi-square and *p*-values very similar to those obtained on the corresponding matrices generated using Kendall's tau, with identical outcomes for the significance tests in all cases except for the within-piece, across-parameters congruence for the *Prélude* which was marginally significant with Spearman's rho ($\chi^2 = 85.35$, $p = 0.074$) but reached significance with Kendall's tau ($\chi^2 = 94.95$, $p = 0.020$). Overall, the comparable results obtained with both correlation coefficients suggest that our approach is robust to the type of non-parametric correlation used in the similarity analysis.

Discussion

The CADM tests conducted to assess within-piece, or within-interpretation, congruence across all expressive parameters were at least marginally significant for all pieces and interpretations analyzed here, suggesting that this type of concordance across different parameters is not uncommon. For the CADM tests that reached significance (for the *Prélude* and for the alto and tenor interpretations of the *Partita*), *post-hoc* tests showed that, in all cases, the timing similarity matrix was shown to be significantly congruent with at least one other matrix. This indicates that the magnitude of the correlations between the timing profiles of different performers tended to be positively associated with the magnitude of the correlations between the expressive profiles computed on at least one other expressive parameter, suggesting that timing profiles seem to play a central role in the within-piece or within-interpretation congruences observed here.

CADM tests conducted to assess within-parameter congruence across pieces or interpretations revealed much higher congruences across interpretations of the *Partita* than across pieces, a result which is probably expected given that different interpretations of the same piece are likely to be much more similar to each other than performances of different pieces. Indeed, all Mantel correlations between the similarity matrices corresponding to the same expressive parameter were highly significant for all four parameters when comparing across interpretations of the *Partita* (note that Mantel *r* values are often comparatively small even when significant, see Dutilleul et al., 2000). On the other hand, CADM tests revealed significant congruences across pieces only for timing and articulation. These findings suggest that while it is very likely that performers who display concordant profiles for one interpretation of a piece will also display concordant profiles for a different interpretation of the same piece when considering the same expressive parameter, this is not as likely when comparing across pieces, and was only observed on some expressive parameters.

Finally, comparable results were obtained with either Kendall's tau or Spearman's rho as a measure of pairwise similarity between expressive profiles, indicating that the approach presented here is not dependent on a particular type of non-parametric correlation coefficient. Nevertheless, Kendall's tau remains a more general measure of concordance in our view, for the reasons detailed in the Introduction, and is therefore more suitable than Spearman's rho for comparing expressive profiles across pieces (or interpretations) and parameters.

GENERAL DISCUSSION

In this article, we sought to disentangle performer- and piece-specific influences on expressive strategies in harpsichord performance, by pursuing two lines of inquiry, one based on an analysis of the proportion of variance in the mean values for each expressive parameter explained by performer and piece (or interpretation) effects, and a second one on a similarity analysis of note-by-note expressive profiles. These analyses were conducted on a dataset of recordings of three pieces representative of the harpsichord repertoire made by 12 professional performers and focused on four expressive parameters: articulation, asynchrony, timing, and velocity.

The first approach used LMMs to show that piece-specific influences explained a large proportion of the variance in the mean values for some expressive parameters such as articulation (and to a lesser extent velocity), whereas analogous analyses on the different interpretations of the *Partita* showed only negligible interpretation-specific effects. On the other hand, individual differences explained a much larger proportion of the variance for the LMMs comparing across interpretations of the *Partita* than for those comparing across pieces, indicating that performer-specific influences were prevalent in the former case. These individual differences can be sorted into two categories: (1) piece- or interpretation-related differences between performers (corresponding to a significant random slope in a linear model) and (2) global differences between performers across all pieces or interpretations (corresponding to a significant random intercept). The former were observed on all expressive parameters when comparing across pieces but were generally less important when comparing across different interpretations of the same piece, whereas the latter only reached significance for articulation and velocity when comparing across pieces but amounted to large effects when comparing across interpretations. The fact that some individual performers consistently deviated from the mean values for some expressive parameters, both across different pieces and across interpretations of the same piece, suggests that global, undifferentiated expressivity measures computed over an entire performance, such as the mean overlap or the average key velocity, may in some cases constitute valid markers of artistic individuality that reliably characterize a performer's playing style. Finally, these analyses also revealed important differences in the degree of individuality expressed by performers, with some harpsichordists exhibiting statistically significant individual random intercept or slope effects for several expressive parameters, and others only for a few or none.

The second approach used permutation tests on similarity matrices generated from pairwise Kendall's tau correlations between note-by-note (or event-by-event) expressive profiles to show that, when examining profiles associated with different expressive parameters but all corresponding to the same piece (or same interpretation), we observed in all cases a significant effect, or at least marginally significant tendency, for the degree of concordance between the profiles of different performers for one expressive parameter to be positively correlated with the degree of concordance between their profiles on at least one other expressive parameter. Moreover, we observed that, when comparing profiles associated with the same expressive parameter but corresponding to different interpretations of the *Partita*, there was a significant tendency, for all four expressive parameters, for performers with concordant profiles in one interpretation to also display concordant profiles in another interpretation. This was the case only for articulation and timing when comparing across pieces. These findings have several implications. First, the fact that concordance between different performers in one expressive parameter also tends to be associated with concordance in a different parameter suggests that parameters cannot always be considered in isolation, and that a type of interpretive concinnity can manifest itself across expressive parameters. Although this type of interaction between expressive devices has been reported previously, for instance between tempo and loudness curves (Widmer and Goebl, 2004) or between asynchrony (melody lead) and velocity (Repp, 1996a; but see Goebl, 2001), it had not, to our knowledge, been demonstrated across different pieces or between different performers. Second, the within-parameter, between-pieces congruence observed for articulation and timing indicates that performers who use similar articulation (or timing) patterns in one piece also tend to display similar articulation (or timing) profiles in another piece. This suggests that, at least for these expressive parameters, interpretive choices can transcend pieces and maybe even compositional genres. Lastly, and more generally, these observations point to the existence of expressive meta-strategies encompassing several expressive parameters and that can manifest themselves either across interpretations or even different pieces.

The fact that a significant within-parameter, between-pieces concordance was only observed for articulation and timing may be related to the important role played by these expressive devices in harpsichord performance (Gingras et al., 2009). In particular, the importance of articulation in Baroque music has been noted elsewhere (Rosenblum, 1997; Lawson and Stowell, 1999). In contrast, the relevance of note-by-note velocity patterns is presumably de-emphasized, at least to some extent, due to the limited dynamic differentiation available on the harpsichord (Penttinen, 2006), and it is not clear that note-by-note variations in velocity are intentionally employed for expressive purposes by harpsichordists. Although local tempo variations are generally considered to be intimately related to the formal structure of the pieces under consideration (Todd, 1985; Repp, 1992), individual performers may interpret formal structures in idiosyncratic ways that are consistent across different pieces, so that performers who agree in their timing profiles for one piece tend to agree for a different piece. Hence, the significant between-pieces congruence observed for timing suggests that, for instance, performers using *rallentando* (a gradual slowing down) in similar places in the *Partita* might be expected to also have comparable tempo variation profiles in the *Bergeries*, whereas performers with different profiles for the former piece would also be expected to exhibit less concordance in their profiles for the latter piece. Note-by-note articulation patterns are, presumably, not tightly linked with specific formal structures, but likely correspond to piece-independent patterns that are characteristic of an individual performer's playing style. Perhaps paradoxically, the between-pieces concordance in note-by-note articulation patterns, which suggests the existence of a performer-related specificity that transcends pieces, is contrasted with a very strong piece-specific effect on the global articulation trends (see **Table 1**). Clearly, more research, using a greater number of pieces and possibly involving comparisons across instruments and repertoires, is necessary not only to elucidate the issues outlined here, but also to investigate more comprehensively the nature and prevalence of the expressive meta-strategies that the present research has uncovered.

A particularity of the experimental design used here is that performers were asked to emphasize a specific melodic line in the *Partita* (although they were free to choose whichever expressive strategy they desired to achieve that aim), whereas they were simply invited to play as if in a recital setting for the other pieces. Although the instruction to emphasize a melodic line would likely have affected the performers' expressive choices, thus potentially biasing our results, we do not believe this to be a major concern here. First, we retained only the *Partita* recordings emphasizing the highest melodic line (soprano), which is probably closest to a natural interpretation (Palmer and Holleran, 1994; Palmer, 1996; Goebl, 2001), for the comparisons across pieces. In any case, very similar results were obtained when substituting the soprano interpretation with the alto or tenor interpretations, both for the LMMs and for the similarity analyses, suggesting that the choice of interpretation had only a minimal impact on our results. Second, the results do not show any evidence that the performers were unable to fully express their individuality in performances of the *Partita*. In that regard, it is especially relevant to note that significant congruences were observed between the timing profiles for the *Prélude* and the *Partita*, as well as for the note-by-note articulation patterns between the *Partita* and the *Bergeries* (with a marginally significant congruence between the *Partita* and the *Prélude*). This indicates that the concordance in timing or articulation patterns between individual performers was preserved between the *Partita* and other pieces, even though the performance instructions were different, which is a noteworthy result in itself.

Besides the dichotomy between measured and unmeasured pieces in our sample, it is likely that other stylistic features, related for instance to the date of composition, the compositional genre, the meter (the *Bergeries* is written in 6/8, whereas the *Partita* follows a 4/4 meter) or the texture (the *Partita* is written in a much more polyphonic style than the other pieces) played a role in the performers' choice of expressive strategies. Our analysis did not account for these aspects, but further research may address them more directly by examining a greater number of pieces and perhaps more explicitly adopting a musicological perspective. Other expressive strategies relating to tempo and meter, such as the use of *notes inégales* (in which some notes with equal written time values are performed with unequal durations, usually as alternating long and short) and other types of durational contrasts (Fabian and Schubert, 2010; Moelants, 2011), or metrical emphasis (strong vs. weak beats), could also be explored in greater depth.

In conclusion, this study highlighted the usefulness of LMMs, and more generally mixed models employing likelihood estimation, in quantifying piece-, interpretation-, and performer-specific influences on expressive choices, as well as the relevance of similarity analyses, based on methods found more commonly in biological sciences (such as the CADM and Mantel tests), to the comparison of expressive profiles across pieces (or interpretations) and expressive parameters. Notably, because our methodology for the similarity analysis relies entirely on non-parametric tests, it is potentially broadly applicable and could likely be generalized to the study of other expressive parameters or even perceptual response profiles. Our findings constitute a significant addition to the literature on individuality in music performance, especially given that very few studies compared the same performers across different pieces or interpretations of the same piece. Finally, the combination of the type of analysis proposed here with empirical studies examining the perception of individuality in music performance (Gingras et al., 2011; Koren and Gingras, 2011) could conceivably prove to be a very fruitful and synergistic endeavor in the investigation of artistic individuality.

MATERIALS AND METHODS

PARTICIPANTS

Twelve professional harpsichordists, five female and seven male, from the Montreal (Canada) area were invited to participate in the experiment. Their average age was 39 years (range: 21–61 years). They had played the harpsichord for a mean duration of 22 years (range: 6–40). Seven of them had previously won prizes in regional, national, or international harpsichord competitions. Ten reported being right-handed, one left-handed, and one ambidextrous. All harpsichordists signed a consent form and received financial compensation for their participation in the study, which was approved and reviewed by the Research Ethics Board of McGill University (Montreal, Canada).

PROCEDURE

For the *Prélude* and the *Bergeries*, performers received no instructions besides playing the pieces as if in a "recital setting." Each piece was recorded twice. In the case of the *Partita*, performers were instructed to play three versions, each emphasizing a different voice (respectively, the soprano, alto, and tenor parts). Each interpretation was recorded twice, for a total of six recordings per performer. The order of the instructions was randomized according to a Latin square design. Performers were given 20 min to practice before recording the pieces (the scores of the *Prélude* and of the *Bergeries* were given to the performers 4–6 weeks before the recording session). The entire recording session lasted ~1 h.

Performances took place in an acoustically treated studio, on an Italian-style Bigaud harpsichord (Heugel, Paris, France) with two 8-foot stops. Only the back stop was used for the experiment. This harpsichord was equipped with a MIDI console, allowing precise measurement of performance parameters. MIDI velocities were estimated by a mechanical double contact located underneath the keys and from which the travel time of the keys was measured, with a high velocity corresponding to a shorter travel time (faster attack). MIDI velocity values for each note event were coded in a range between 16 (slowest) and 100 (fastest). The measured velocities were calibrated separately for each key by authors Bruno Gingras and Pierre-Yves Asselin.

The audio signal was recorded through two omnidirectional microphones MKH 8020 (Sennheiser GmbH, Wedemark Wennebostel, Germany). The microphones were located 1 m above the resonance board and were placed 25 cm apart. The audio and MIDI signals were sent to a PC computer through an RME Fireface audio interface (Audio AG, Haimhausen, Germany). Audio and MIDI data were then recorded using Cakewalk's SONAR software (Cakewalk, Inc., Boston, MA, USA) and stored on a hard disk.

PERFORMANCE DATA ANALYSIS

Performances were matched to the scores of the pieces using an algorithm developed by the authors, which has been shown to be suitable for ornamented harpsichord pieces (Gingras and McAdams, 2011). To ensure that the excerpts from all three pieces were of comparable duration, only the first part of the rondo from the *Bergeries* was used in all subsequent analyses. The excerpt from the *Bergeries* comprised 281 notes, whereas the *Partita* contained 153 notes, and the *Prélude* 140 notes. The average duration (from first to last onset) was 54.2 s for the *Bergeries* excerpt (range: 47.1–61.6 s), 36.8 s for the *Partita* (range: 28.3–47.5 s), and 59.9 s for the *Prélude* (range: 39.1–84.2 s). These durations corresponded to the following tempi: for the *Bergeries*, the mean tempo was 107.0 beats per minute (bpm), ranging from 93.5 to 122.3 bpm, with the beat corresponding to an eighth note (6/8 meter); for the *Partita*, the mean tempo was 52.9 bpm (range: 40.4–67.7 bpm), with the beat corresponding to a quarter note (4/4 meter); for the *Prélude*, the mean tempo was 149.2 bpm (range: 99.8–214.7 bpm), with each note onset counted as a "beat" in the absence of a notated rhythmic structure.

The mean error rates per performance, defined as the proportion of wrong notes or missing notes relative to the total number of score notes, were as follows: for the *Bergeries*, 0.37% (range: 0–1.42%); for the *Partita*, 0.82% (range: 0–2.61%); and for the *Prélude*, 0.54% (range: 0–2.14%). These low error rates are comparable to the rates reported by Repp (1996b) and Goebl (2001) in studies on professional piano performance, suggesting that the performance data collected for the current study were of suitable quality for assessing individual expressive profiles in professional harpsichord performance.

Four expressive parameters were analyzed for each performance: articulation, note onset asynchrony, timing, and velocity. Articulation refers to the amount of overlap between two consecutive note events n_i and n_j belonging to the same melodic line or voice. A *legato* articulation corresponds to a positive overlap (when the offset of note n_i occurs after the onset of note n_j), whereas a detached or staccato articulation corresponds to a negative overlap. Here, the onset of a note is defined as the time at which the corresponding key is pressed (as measured by the MIDI system) and its offset corresponds to the time at which the key is released. Because the amount of overlap varies with tempo (Repp, 1995), we chose to use the overlap ratio, defined as the ratio of the overlap between two consecutive note events and the inter-onset interval between these notes, as a measure of articulation (Bresin and Battel, 2000).

Note onset asynchrony is defined as the difference in onset time between note onsets that are notated in the musical score as synchronous (Palmer, 1989). Several measures of onset asynchrony have been constructed. Rasch (1979) proposed to use the root mean square, or standard deviation of the onset times of nominally simultaneous notes. We chose to use this measure here. Onset asynchrony values were not computed in the case of the *Prélude*, whose score does not include nominally synchronous notes.

To analyze expressive timing, tempo values were computed from the inter-onset interval between consecutive note onset events. To allow for meaningful comparisons across pieces for the LMM analysis, the logarithm in base two of the total duration of the piece (defined as the time interval between the first onset and the last onset of the piece, see Moelants, 2000) divided by the geometric mean of the total duration across all performances, was used (Wagner, 1974; see also Repp, 1992). This procedure yields a tempo valuation that is centered and scaled for each piece, with a value of 1 corresponding to a tempo that is twice as slow (duration twice as long) as the mean tempo for all performances of the piece, and a value of −1 corresponding to a tempo that is twice as fast (duration twice as short). Untransformed durations were used for the LMMs analyzing the different interpretations of the *Partita*.

For the event-by-event timing profiles, local tempo values were obtained for each note onset event *e* by computing the logarithm of the ratio of the duration (inter-onset interval) of *e* to its expected duration (i.e., the duration obtained by dividing the notated duration of *e* by the notated duration of the entire piece, corresponding to a "deadpan" or mechanical performance with an invariant tempo) was used. In this case, a local tempo value of 0 for *e* corresponds to the mean tempo of the performance (indicating no local deviation from the mean tempo), whereas a value of 1 corresponds to a local tempo that is twice as slow as the mean tempo, and a value of −1 to a tempo that is twice as fast.

Lastly, in the case of velocity, the raw MIDI velocity values associated with the key press corresponding to each note onset (see Procedure) were used for the analysis.

STATISTICAL ANALYSIS

LMMs were fitted using the PROC MIXED function in SAS 9.0 (SAS Institute, Cary, NC, USA) (Singer, 1998; Littell et al., 2006). All models were fitted using a variance components (VC) covariance matrix, which is the default covariance matrix structure for LMMs both in PROC MIXED and in the analogous MIXED procedure in SPSS 19.0 (SPSS Inc., Chicago, IL, USA). Analogous models were also fitted with the unstructured and compound symmetry covariance structures, but these models either did not converge or yielded worse fits than equivalent LMMs fitted with the default VC structure. Although all reported analyses were conducted in SAS, we verified that equivalent models yielded identical results with the MIXED procedure in SPSS. Additionally, the GLMM analysis on error rates was conducted using the PROC GLIMMIX function in SAS.

The CADM ("Congruence among distance matrices") (Legendre and Lapointe, 2004) and Mantel tests (Mantel, 1967; Mantel and Valand, 1970) were conducted on the similarity matrices using the routines CADM.global and CADM.post in package ape (Paradis et al., 2004) in R (R Core Team, 2013). One-tailed significance tests, corresponding to a positive association, were conducted for both the CADM and Mantel tests, following the procedure described in Legendre and Lapointe (2004). 99,999 permutations were conducted to assess significance for both tests.

ACKNOWLEDGMENTS

This research was supported by a grant from the Canadian Natural Sciences and Engineering Research Council and a Canada Research Chair (Stephen McAdams). We wish to thank Alain Poirier, former director of the Conservatoire National Supérieur de Musique et de Danse de Paris, for the loan of the MIDI harpsichord. We also thank Bennett Smith, Meghan Goodchild, Julien Boissinot, and Harold Kilianski for their technical assistance, McGill Sound Recording for permission to use their equipment, and Martha de Francisco for her assistance with sound recording.

REFERENCES

Arndt, S., Turvey, C., and Andreasen, N. C. (1999). Correlating and predicting psychiatric symptom ratings: Spearman's r versus Kendall's tau correlation. *J. Psychiatr. Res.* 33, 97–104. doi: 10.1016/S0022-3956(98)90046-2

Bresin, R., and Battel, G. U. (2000). Articulation strategies in expressive piano performance: analysis of legato, staccato, and repeated notes in performances of the Andante movement of Mozart's Sonata in G Major (K 545). *J. New Music Res.* 29, 211–224. doi: 10.1076/jnmr.29.3.211.3092

Cheng, J., Edwards, L. J., Maldonado-Molina, M. M., Komro, K. A., and Muller, K. E. (2009). Real longitudinal data analysis for real people: building a good enough mixed model. *Stat. Med.* 29, 504–520. doi: 10.1002/sim.3775

Dalla Bella, S., and Palmer, C. (2011). Rate effects on timing, key velocity, and finger kinematics in piano performance. *PLoS ONE* 6:e20518. doi: 10.1371/journal.pone.0020518

Dietz, E. J. (1983). Permutation tests for association between two distance matrices. *Syst. Biol.* 32, 21–26. doi: 10.1093/sysbio/32.1.21

Dutilleul, P., Stockwell, J., Frigon, D., and Legendre, P. (2000). The Mantel test versus Pearson's correlation analysis: assessment of the differences for biological and environmental studies. *J. Agric. Biol. Environ. Stat.* 5, 131–150. doi: 10.2307/1400528

Fabian, D., and Schubert, E. (2010). A new perspective on the performance of dotted rhythms. *Early Music* 38, 585–588. doi: 10.1093/em/caq079

Gabrielsson, A. (2003). Music performance research at the millenium. *Psychol. Music* 31, 221–272. doi: 10.1177/03057356030313002

Gingras, B. (2008). *Expressive Strategies and Performer-Listener Communication in Organ Performance*. Unpublished Ph.D. dissertation, McGill University (Montreal, QC).

Gingras, B., Asselin, P.-Y., Goodchild, M., and McAdams, S. (2009). "Communicating voice emphasis in harpsichord performance," in *Proceedings of the 7th Triennial Conference of European Society for the Cognitive Sciences of Music (ESCOM 2009)*, eds J. Louhivuori, T. Eerola, S. Saarikallio, T. Himberg, and P. S. Eerola (Jyväskylä), 154–158.

Gingras, B., Lagrandeur-Ponce, T., Giordano, B. L., and McAdams, S. (2011). Perceiving musical individuality: performer identification is dependent on performer expertise and expressiveness, but not on listener expertise. *Perception* 40, 1206–1220. doi: 10.1068/p6891

Gingras, B., and McAdams, S. (2011). Improved score-performance matching using both structural and temporal information from MIDI recordings. *J. New Music Res.* 40, 43–57. doi: 10.1080/09298215.2010.545422

Gingras, B., McAdams, S., Schubert, P., and Utz, C. (2010). "The performer as analyst: a case study of JS Bach's "Dorian"Fugue (BWV 538)," in *Music Theory and Interdisciplinarity—8th Congress of the Gesellschaft für Musiktheorie Graz 2008* (Saarbrücken: Pfau-Verlag), 305–318.

Goebl, W. (2001). Melody lead in piano performance: expressive device or artifact? *J. Acoust. Soc. Am.* 110, 563–572. doi: 10.1121/1.1376133

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scand. J. Stat.* 6, 65–70.

Hubert, L. J. (1979). Comparison of sequences. *Psychol. Bull.* 86, 1098–1106. doi: 10.1037/0033-2909.86.5.1098

Huhta, V. (1979). Evaluation of different similarity indices as measures of succession in arthropod communities of the forest floor after clear-cutting. *Oecologia* 41, 11–23. doi: 10.1007/BF00344834

Koren, R., and Gingras, B. (2011). "Perceiving individuality in musical performance: recognizing harpsichordists playing different pieces," in *Proceedings of the International Symposium on Performance Science 2011*, eds A. Williamon, D. Edwards, and L. Bartel (Utrecht: European Association of Conservatoires), 473–478.

Kumari, S., Nie, J., Chen, H.-S., Ma, H., Stewart, R., Li, X., et al. (2012). Evaluation of gene association methods for coexpression network construction and biological knowledge discovery. *PLoS ONE* 7:e50411. doi: 10.1371/journal.pone.0050411

Laird, N., Lange, N., and Stram, D. (1987). Maximum likelihood computations with repeated measures: application of the EM algorithm. *J. Am. Stat. Assoc.* 82, 97–105. doi: 10.1080/01621459.1987.10478395

Laird, N., and Ware, J. (1982). Random-effects models for longitudinal data. *Biometrics* 38, 963–974. doi: 10.2307/2529876

Lawson, C., and Stowell, R. (1999). *The Historical Performance of Music : An Introduction*. Cambridge, UK. ; New York, NY: Cambridge University Press. doi: 10.1017/CBO9780511481710

Legendre, P. (2000). Comparison of permutation methods for the partial correlation and partial Mantel tests. *J. Stat. Comput. Simul.* 67, 37–73. doi: 10.1080/00949650008812035

Legendre, P., and Lapointe, F.-J. (2004). Assessing congruence among distance matrices: single-malt scotch whiskies revisited. *Aust. N.Z. J. Stat.* 46, 615–629. doi: 10.1111/j.1467-842X.2004.00357.x

Legendre, P., and Legendre, L. (1998). *Numerical Ecology*. 2nd Edn. Amsterdam: Elsevier.

Lindstrom, M. J., and Bates, D. M. (1988). Newton-Raphson and EM algorithms for linear mixed-effects models for repeated-measures data. *J. Am. Stat. Assoc.* 83, 1014–1022.

Littell, R. C., Milliken, G. A., Stroup, W. W., Wolfinger, R. D., and Schabenberger, O. (2006). *SAS for Mixed Models*. 2nd Edn. Cary, NC: SAS Institute Inc.

Mantel, N. (1967). The detection of disease clustering and a generalized regression approach. *Cancer Res.* 27, 209–220.

Mantel, N., and Valand, R. S. (1970). A technique of nonparametric multivariate analysis. *Biometrics* 26, 547–558. doi: 10.2307/2529108

McArdle, B. H., and Anderson, M. J. (2001). Fitting multivariate models to community data: a comment on distance-based redundancy analysis. *Ecology* 82, 290–297. doi: 10.1890/0012-9658(2001)082[0290:FMMTCD]2.0.CO;2

Moelants, D. (2000). Statistical analysis of written and performed music. A study of compositional principles and problems of coordination and expression in "punctual" serial music. *J. New Music Res.* 29, 37–60. doi: 10.1076/0929-8215(200003)29:01;1-P;FT037

Moelants, D. (2011). The performance of notes inégales: the influence of tempo, musical structure, and individual performance style on expressive timing. *Music Percept.* 28, 449–460. doi: 10.1525/mp.2011.28.5.449

Morrell, C. H. (1998). Likelihood ratio testing of variance components in the linear mixed-effects model using restricted maximum likelihood. *Biometrics* 54, 1560–1568. doi: 10.2307/2533680

Nakagawa, S., and Schielzeth, H. (2013). A general and simple method for obtaining R^2 from generalized linear mixed-effects models. *Methods Ecol. Evol.* 4, 133–142. doi: 10.1111/j.2041-210x.2012.00261.x

Newson, R. (2002). Parameters behind "nonparametric" statistics: Kendall's tau, Somers' D and median differences. *Stata J.* 2, 45–64.

Palmer, C. (1989). Mapping musical thought to musical performance. *J. Exp. Psychol. Hum. Percept. Perform.* 15, 331–346. doi: 10.1037/0096-1523.15.2.331

Palmer, C. (1996). On the assignment of structure in music performance. *Music Percept.* 14, 23–56. doi: 10.2307/40285708

Palmer, C., and Holleran, S. (1994). Harmonic, melodic, and frequency height influences in the perception of multivoiced music. *Percept. Psychophys.* 56, 301–312. doi: 10.3758/BF03209764

Paradis, E., Claude, J., and Strimmer, K. (2004). APE: analyses of phylogenetics and evolution in R language. *Bioinformatics* 20, 289–290. doi: 10.1093/bioinformatics/btg412

Penttinen, H. (2006). "On the dynamics of the harpsichord and its synthesis," in *Proceedings of the 9th International Conference on Digital Audio Effects (DAFx-06)*, ed V. Verfaille (Montreal, QC), 115–120.

Popescu, M., and Dinu, L. P. (2009). "Comparing statistical similarity measures for stylistic multivariate analysis," in *Proceedings of the 2009 International Conference on Recent Advances in Natural Language Processing (RANLP 2009)*, eds G. Angelova, K. Bontcheva, R. Mitkov, N. Nicolov, and N. Nikolov (Borovets: Association for Computational Linguistics), 349–354.

Ramirez, R., Maestre, E., Perez, A., and Serra, X. (2011). Automatic performer identification in celtic violin audio recordings. *J. New Music Res.* 40, 165–174. doi: 10.1080/09298215.2011.572171

Ramirez, R., Maestre, E., and Serra, X. (2010). Automatic performer identification in commercial monophonic jazz performances. *Pattern Recognit. Lett.* 31, 1514–1523. doi: 10.1016/j.patrec.2009.12.032

Rasch, R. A. (1979). Synchronization in performed ensemble music. *Acustica* 43, 121–131.

Raudenbush, S., and Bryk, A. (2002). *Hierarchical Linear Models: Applications and Data Analysis Methods (Advanced Quantitative Techniques in the Social Sciences)*. 2nd Edn. Thousand Oaks, CA: Sage Publications, Inc.

R Core Team. (2013). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing. Available online at: http://www.R-project.org/.

Repp, B. H. (1992). Diversity and commonality in music performance - an analysis of timing microstructure in Schumann's "Träumerei." *J. Acoust. Soc. Am.* 92, 2546–2568. doi: 10.1121/1.404425

Repp, B. H. (1995). Acoustics, perception, and production of legato articulation on a digital piaNo. *J. Acoust. Soc. Am.* 97, 3862–3874. doi: 10.1121/1.413065

Repp, B. H. (1996a). Patterns of note onset asynchronies in expressive piano performance. *J. Acoust. Soc. Am.* 100, 3917–3931. doi: 10.1121/1.417245

Repp, B. H. (1996b). The art of inaccuracy: why pianists' errors are difficult to hear. *Music Percept.* 14, 161–183. doi: 10.2307/40285716

Rosenblum, S. P. (1997). Concerning articulation on keyboard instruments: aspects from the renaissance to the present. *Perform. Pract. Rev.* 10, 31–40. doi: 10.5642/perfpr.199710.01.04

Saunders, C., Hardoon, D. R., Shawe-Taylor, J., and Widmer, G. (2008). Using string kernels to identify famous performers from their playing style. *Intell. Data Anal.* 12, 425–440.

Schielzeth, H., and Forstmeier, W. (2009). Conclusions beyond support: overconfident estimates in mixed models. *Behav. Ecol.* 20, 416–420. doi: 10.1093/beheco/arn145

Singer, J. D. (1998). Using SAS PROC MIXED to fit multilevel models, hierarchical models, and individual growth models. *J. Educ. Behav. Stat.* 23, 323–355. doi: 10.3102/10769986023004323

Sloboda, J. A. (2000). Individual differences in music performance. *Trends Cogn. Sci.* 4, 397–403. doi: 10.1016/S1364-6613(00)01531-X

Snijders, T., and Bosker, R. (1999). *Multilevel Modeling: An Introduction to Basic and Advanced Multilevel Modeling*. London: Sage Publications.

Stamatatos, E., and Widmer, G. (2005). Automatic identification of music performers with learning ensembles. *Artif. Intell.* 165, 37–56. doi: 10.1016/j.artint.2005.01.007

Todd, N. (1985). A model of expressive timing in tonal music. *Music Percept.* 3, 33–58. doi: 10.2307/40285321

Van Vugt, F. T., Jabusch, H.-C., and Altenmüller, E. (2013). Individuality that is unheard of: systematic temporal deviations in scale playing leave an inaudible pianistic fingerprint. *Front. Psychol.* 4:134. doi: 10.3389/fpsyg.2013.00134

Verbeke, G., and Molenberghs, G. (2000). *Linear Mixed Models for Longitudinal Data*. New York, NY: Springer.

Wagner, C. (1974). Experimentelle Untersuchungen über das Tempo [Experimental investigations concerning tempo]. *Österreichische Musikz* 29, 589–604.

West, B. T., Welch, K. B., and Galecki, A. T. (2007). *Linear Mixed Models: Practical Guide Using Statistical Software*. Boca Raton, FL: Chapman and Hall/CRC Press.

Widmer, G., and Goebl, W. (2004). Computational models of expressive music performance: the state of the art. *J. New Music Res.* 33, 203–216. doi: 10.1080/0929821042000317804

Yona, G., Dirks, W., Rahman, S., and Lin, D. M. (2006). Effective similarity measures for expression profiles. *Bioinformatics* 22, 1616–1622. doi: 10.1093/bioinformatics/btl127

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 30 June 2013; accepted: 12 November 2013; published online: 28 November 2013.

Citation: Gingras B, Asselin P-Y and McAdams S (2013) Individuality in harpsichord performance: disentangling performer- and piece-specific influences on interpretive choices. Front. Psychol. ***4****:895. doi: 10.3389/fpsyg.2013.00895*

This article was submitted to Cognitive Science, a section of the journal Frontiers in Psychology.

frontiers in
PSYCHOLOGY

ORIGINAL RESEARCH ARTICLE
published: 24 February 2014
doi: 10.3389/fpsyg.2014.00141

Perceiving individuality in harpsichord performance

***Réka Koren*[1] *and Bruno Gingras*[2] ***

[1] *Goldsmiths College, University of London, London, UK*
[2] *Department of Cognitive Biology, University of Vienna, Vienna, Austria*

Edited by:
Eddy J. Davelaar, Birkbeck College, UK

Reviewed by:
Eddy J. Davelaar, Birkbeck College, UK
Sarah Creel, University of California at San Diego, USA

****Correspondence:***
Bruno Gingras, Department of Cognitive Biology, University of Vienna, Althanstrasse 14, Vienna A-1090, Austria
e-mail: brunogingras@gmail.com

Can listeners recognize the individual characteristics of unfamiliar performers playing two different musical pieces on the harpsichord? Six professional harpsichordists, three prize-winners and three non prize-winners, made two recordings of two pieces from the Baroque period (a variation on a *Partita* by Frescobaldi and a rondo by François Couperin) on an instrument equipped with a MIDI console. Short (8 to 15 s) excerpts from these 24 recordings were subsequently used in a sorting task in which 20 musicians and 20 non-musicians, balanced for gender, listened to these excerpts and grouped together those that they thought had been played by the same performer. Twenty-six participants, including 17 musicians and nine non-musicians, performed significantly better than chance, demonstrating that the excerpts contained sufficient information to enable listeners to recognize the individual characteristics of the performers. The grouping accuracy of musicians was significantly higher than that observed for non-musicians. No significant difference in grouping accuracy was found between prize-winning performers and non-winners or between genders. However, the grouping accuracy was significantly higher for the rondo than for the variation, suggesting that the features of the two pieces differed in a way that affected the listeners' ability to sort them accurately. Furthermore, only musicians performed above chance level when matching variation excerpts with rondo excerpts, suggesting that accurately assigning recordings of different pieces to their performer may require musical training. Comparisons between the MIDI performance data and the results of the sorting task revealed that tempo and, to a lesser extent, note onset asynchrony were the most important predictors of the perceived distance between performers, and that listeners appeared to rely mostly on a holistic percept of the excerpts rather than on a comparison of note-by-note expressive patterns.

Keywords: music performance, individuality, harpsichord, categorization, musical expertise

INTRODUCTION

Identification and categorization are essential features of perception without which it is impossible to properly interpret sensory information (Riesenhuber and Poggio, 2000). Thus, they constitute vital abilities that are crucial for day-to-day survival (Ashby and Maddox, 2005). Generally, the probability that two objects or stimuli will be assigned to the same category or misidentified increases as the similarity between them increases (Ashby and Perrin, 1988). Indeed, there is a tight empirical (Ashby and Lee, 1991) and theoretical (Riesenhuber and Poggio, 2000) link between similarity, categorization, and identification.

The ability to identify individuals is particularly important in species, such as humans, that value kin recognition (Tang-Martinez, 2001) and social interaction (Thompson and Hardee, 2008). The perception of individuality, which is closely related to identification, is in all likelihood also based on a general process of similarity estimation (Ashby and Lee, 1991). Humans are able to identify individuals on the basis of relatively static cues such as facial features (Carey, 1992; Haxby et al., 2000; Andrews and Ewbank, 2004), or by using dynamic displays such as gait and walking (Johansson, 1973; Cutting and Kozlowski, 1977; Loula et al., 2005; Blake and Shiffrar, 2007). This ability to recognize identity cues is not confined to visual perception, but also extends to acoustic cues. Thus, individuals can be recognized and differentiated on the basis of acoustic stimuli such as voices (Belin et al., 2004) – whether those of famous (Van Lancker et al., 1985), familiar (Blatchford and Foulkes, 2006) or unfamiliar people (Sheffert et al., 2002), clapping patterns (Repp, 1987), or even tones which follow similar temporal patterns to clapping (Flach et al., 2004).

Identity cues can also be conveyed efficiently through music performance. Skilled music performance comprises two major components: a technical component and an expressive one (Sloboda, 2000). The former refers mostly to the biomechanical aspects that play a role in producing a fluent performance, while the latter corresponds to intentional variations in performance parameters with the aim of influencing cognitive and esthetic outcomes for the listener. The main performance parameters that can be expressively varied by performers include timbre, pitch, rhythm, tempo, dynamics, and articulation (Sloboda, 2000; Juslin and Sloboda, 2001). Some of these parameters may not be available depending on the musical instrument, and may not be appropriate depending on the musical genre of the piece being performed. Through expressive variations in these parameters, performers not only showcase their musical creativity and personality, but also display their individuality in a manner that may be uniquely identifiable, for instance by playing a well-known repertoire piece in a manner that is recognizably different from typical performances of that piece (Repp, 1992, 1997; Lehmann et al., 2007, p. 85). Hence, famous musicians such as John Coltrane or Sonny Rollins can be recognized after playing only a few notes (Benadon, 2003), and pianists have been shown to be able to recognize their own performances from modified recordings in which only temporal information was available (Repp and Knoblich, 2004; Repp and Keller, 2010).

However, few studies have investigated whether listeners could accurately distinguish between unfamiliar performers playing different interpretations of the same piece on the same instrument. We have recently shown that both non-musicians and musicians perform significantly above chance in such a task, even in the absence of timbral or dynamic differentiation, supporting previous findings which suggested that timing cues can be sufficient to enable performer recognition (Gingras et al., 2011). Our results also showed that expressive interpretations were sorted more accurately than inexpressive ones. Although these findings indicate that listeners can distinguish between unfamiliar performers playing two different interpretations of the same piece, it remained to be seen whether similar results would be observed in a study using performances from two different pieces. Using a machine-learning approach, Stamatatos and Widmer (2005) had previously shown that a set of classifiers trained on a database of piano performances of 22 pianists playing one piece by Fryderyk Chopin could reliably recognize these same pianists performing a different piece by the same composer. Indeed, the authors noted that the 70% recognition rate achieved by their learning ensemble represented "a level of accuracy unlikely to be matched by human listeners" (Stamatatos and Widmer, 2005, p. 54), a claim that has not been empirically verified so far to our knowledge and that stands in contrast to earlier observations suggesting that humans are generally more accurate than machines in such categorization tasks (Ashby and Maddox, 1992).

The main objective of the present study was to test whether human listeners are indeed able to recognize unfamiliar performers playing two different pieces, by asking listeners to group together excerpts from two different harpsichord pieces which they think have been played by the same performer. Although our study does not use the same stimuli and design as Stamatatos and Widmer (2005), it can be considered a general test of their prediction that human listeners cannot match the accuracy of a learning ensemble in such a task. Thus, the current study differs from our earlier study (Gingras et al., 2011) in its use of two different pieces instead of one piece recorded with two different interpretations, in addition to its focus on harpsichord instead of organ music. As indicated by previous research (Repp and Knoblich, 2004; Gingras et al., 2011), excerpts from the same piece can be matched to the same performer by relying mostly on expressive timing and articulation patterns. However, successfully matching excerpts from two different pieces would presumably require the listener to detect identity cues of a more general nature that transcend specific pieces, such as performer-specific timing or articulation patterns that can be found across different pieces (see Gingras et al., 2013). We therefore hypothesized that matching excerpts from the same piece should be perceptually easier than matching excerpts from different pieces, and that this would be reflected in a higher level of accuracy. We also considered the possibility that some pieces would be easier to sort accurately than others, either because they afford performers more possibilities for conveying their artistic individuality, or because they can be processed more easily by listeners.

Our rationale for using harpsichord performances in this study was to extend performance research on other keyboard instruments besides the piano, as there is a large body of research on piano performance (see Gabrielsson, 2003 for a review), but few published empirical studies on harpsichord performance. Moreover, the harpsichord is not widely known in the general public, and thus represents an ideal medium for a study on the recognition of unfamiliar performers. Finally, its mechanism affords no or very little timbre differentiation (excluding registration changes), and only limited dynamic variation (Penttinen, 2006). Consequently, as with organ performance (Gingras, 2008; Gingras et al., 2010), expressivity in harpsichord performance is confined mostly to timing and articulation. Because all recordings were realized on the same instrument and with the same registration (configuration of stops controlling the timbre), listeners had to rely almost exclusively on temporal parameters to discriminate between performers, as in our earlier study (Gingras et al., 2011).

Another aim of the study was to investigate the effects of the performer's level of expertise and the effects of musical training on categorization accuracy. Repp (1997) showed that recordings from world-famous pianists tend to be perceived by listeners as exhibiting more individuality than those of graduate students in piano performance. This link between the performers' level of expertise and their perceived individuality was also observed in the study by Gingras et al. (2011), which showed that listeners sorted performances by prize-winning performers more accurately than those by non-prize-winners.

Several studies have demonstrated a timbre- and pitch-processing advantage for musicians versus non-musicians (for a review, see Chartrand et al., 2008). However, the effect of musical expertise appears to be task-dependent, and a number of responses to musical stimuli are largely unaffected by musical training (Bigand and Poulin-Charronnat, 2006). Both musicians and non-musicians can reliably distinguish among different levels of expressiveness in performances of the same piece (Kendall and Carterette, 1990), and discriminate between familiar and novel performances of the same piece (Palmer et al., 2001). Although musicians discriminated between performers more accurately than non-musicians in the sorting task described in Gingras et al. (2011), the difference was not statistically significant. Here, we also compared the sorting accuracy of musicians and non-musicians by using an experimental design similar to our earlier study. Additionally, we also controlled for possible gender effects by balancing the number of male and female participants for both musicians and non-musicians.

As in Gingras et al. (2011), we used a constrained sorting task in which participants are given information about the underlying category structure (in this case, the number of performers and the number of pieces) prior to the experiment. Whereas unconstrained sorting tasks tend to focus on the processes leading to category construction, constrained tasks focus on the types of category structures that can be learned by participants in the absence of trial-by-trial feedback (Ell and Ashby, 2012) and are thus appropriate for investigating listeners' ability to discriminate between individual performers.

METHODS

RECORDING HARPSICHORD PERFORMANCES

Participants

Twelve professional harpsichordists, five female and seven male, from the Montreal (Canada) area recorded the pieces that were used as stimuli in the listening experiment. Their average age was 39 years (range: 21–61 years). They had played the harpsichord for a mean duration of 22 years (range: 6–40). Seven of them had previously won prizes in regional, national, or international harpsichord competitions. Ten reported being right-handed, one left-handed, and one ambidextrous. All harpsichordists signed a consent form and received financial compensation for their participation in the study, which was approved and reviewed by the Research Ethics Board of McGill University (Montreal, QC, Canada).

Materials

Two pieces were selected for this study: the third variation from the *Partita No. 12 sopra l'aria di Ruggiero* by Girolamo Frescobaldi (1583–1643), and *Les Bergeries*, a rondo by François Couperin (1668–1733). Both pieces are representative of the Baroque harpsichord repertoire.

Procedure

For the *Bergeries*, performers received no instructions besides playing the pieces as if in a "recital setting." Each piece was recorded twice. In the case of the *Partita*, performers were instructed to play three versions, each emphasizing a different voice (respectively, the soprano, alto, and tenor parts). Each of the three versions was recorded twice, for a total of six recordings per performer. The order of the instructions for the *Partita* was randomized according to a Latin square design. The entire recording session lasted approximately one hour. Only the recordings emphasizing the highest voice were used here.

Performances took place in an acoustically treated studio, on an Italian-style Bigaud harpsichord (Heugel, Paris, France) with two 8-foot stops. Only the back stop was used for the experiment. This harpsichord was equipped with a MIDI console, allowing precise measurement of performance parameters. MIDI velocities were estimated by a mechanical double contact located underneath the keys and from which the travel time of the keys was measured, with a high velocity corresponding to a shorter travel time (faster attack). MIDI velocity values for each note event were coded in a range between 16 (slowest) and 100 (fastest). The measured velocities were calibrated separately for each key prior to the recording sessions.

The audio signal was recorded through two omnidirectional microphones MKH 8020 (Sennheiser GmbH, Wedemark Wennebostel, Germany). The microphones were located 1 m above the resonance board and were placed 25 cm apart. The audio and MIDI signals were sent to a PC computer through an RME Fireface audio interface (Audio AG, Haimhausen, Germany). Audio and MIDI data were then recorded using Cakewalk's SONAR software (Cakewalk, Inc., Boston, MA, USA) and stored on a hard disk.

SORTING TASK

Participants

Twenty participants with two or fewer years of musical training (mean age = 27.4 years, SD = 7.2 years), henceforth referred to as non-musicians, and 20 participants having completed at least 1 year in an undergraduate university in music performance or musicology (mean age = 30.2 years, SD = 9.6 years), henceforth referred to as musicians, participated in the experiment. Both groups of participants were balanced for gender. All participants signed a consent form and received a small gift for their participation in the study and the chance to win one out of four prizes worth 20 pounds each. The study was approved by the Research Ethics Board of Goldsmiths College, University of London (London, UK).

Materials

In order to reduce the number of excerpts to a manageable number, four performances (two for each piece) from three prize-winners and three non-prize-winners were used for the sorting task, for a total of 24 excerpts. The six performers were selected according to the following criteria: small error rate (few wrong or missing notes), small tempo differences between both performances of the same piece, and overall quality of the recordings. From the *Bergeries*, an excerpt corresponding to the end of the rondo section was chosen (**Figure 1**), whereas the beginning of the *Partita* was retained (**Figure 2**). Both excerpts were chosen to be syntactically coherent musical units with a clear harmonic closure (perfect authentic cadence) at the end. The duration of the excerpts was 11– 15 s for the *Bergeries* excerpts, and 8–11 s for the *Partita* excerpts. Audacity was used for cutting out and editing (fading in and fading out) the excerpts. The task was practiced with a reduced training stimulus set using four excerpts (two for each piece) from two different performers whose recordings were not included in the main experiment, for a total of eight excerpts.

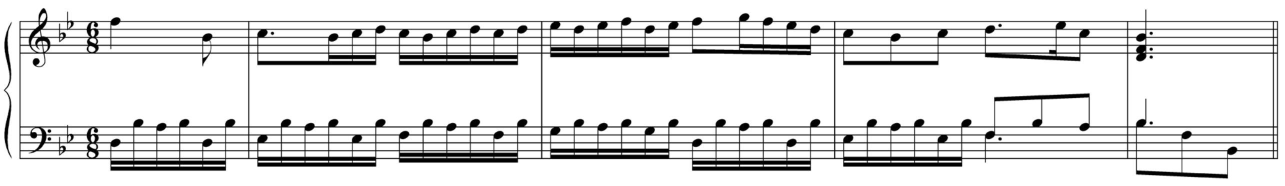


FIGURE 1 | Excerpt from Couperin's *Bergeries* used in the sorting task.

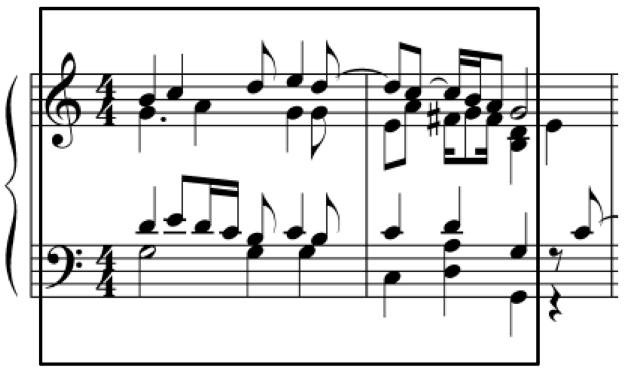


FIGURE 2 | Excerpt from Frescobaldi's *Partita* used in the sorting task. The boxed section corresponds to the fragment heard by participants.

Procedure

The experimental interface, programmed in MATLAB (Gingras et al., 2011), consisted of a computer monitor on which the musical excerpts were represented by 24 randomly numbered icons which when clicked on played the corresponding audio recordings. Participants could listen to the excerpts in any order and as many times they wished. They were asked to group together the excerpts that they believed to have been played by the same performer by moving them into one of six boxes that represented the six performers. The icons corresponding to each of the two pieces had different colors and participants were told that each performer had played each piece twice. Participants had to listen to an excerpt at least once before being able to drag its corresponding icon into a box representing a performer. At the end of the experiment participants were required to listen to the content of each box before finishing the experiment, to ensure that they had listened to each excerpt at least twice. For the experiment to be completed, each box had to contain exactly two icons corresponding to each piece. There was no time limit for the categorization but the time taken to arrange the selections was recorded. Prior to the sorting task, participants practiced the task in a familiarization phase using eight alternate excerpts played by two performers whose recordings were not included in the main task. No feedback was provided on the participants' performance in the familiarization phase. The experiment took place in a sound-attenuated booth. Participants wore Sennheiser HD202 headphones and were screened for peripheral hearing problems by a standard audiometric procedure, using an Amplivox 2160 pure-tone diagnostic audiometer prior to testing. After finishing the computer-based sorting task, participants were asked to complete a questionnaire about their musical background and the strategies they used to complete the sorting task.

PERFORMANCE DATA ANALYSIS

Performances were matched to the scores of the pieces using an algorithm developed by Gingras and McAdams (2011). Four expressive parameters were analyzed for each excerpt: articulation, note onset asynchrony, timing, and velocity. Articulation refers to the amount of overlap between two consecutive note events n_i and n_j belonging to the same melodic line or voice. A legato articulation corresponds to a positive overlap (when the offset of note n_i occurs after the onset of note n_j), whereas a detached or staccato articulation corresponds to a negative overlap. Here, the onset of a note is defined as the time at which the corresponding key is pressed (as measured by the MIDI system) and its offset corresponds to the time at which the key is released. Note onset asynchrony is defined as the difference in onset time between note onsets that are notated in the musical score as synchronous (Palmer, 1989). To analyze expressive timing, tempo values were computed from the inter-onset interval (IOI) between consecutive note onset events. In the case of velocity, the raw MIDI velocity values associated with the key press corresponding to each note onset were used for the analysis. In addition, performance errors (wrong notes and missing notes) were also identified using the score-performance matcher described in Gingras and McAdams (2011).

RESULTS

ANALYSIS OF THE EXPRESSIVE PARAMETERS OF THE PERFORMANCES

In order to compare the excerpts on the basis of their expressive parameters, an analysis was conducted on the following parameters: mean tempo (expressed as mean dotted quarter-note duration in the case of the *Bergeries* and mean quarter-note duration in the case of the *Partita*), tempo variability [expressed as the coefficient of variation of the tempo, which corresponds to the standard deviation of the (dotted) quarter-note duration normalized by the mean duration], articulation (expressed as the degree of overlap between successive notes), and onset asynchrony (referring to the difference in onset times between notes that are attacked simultaneously in the score, such as notes belonging to the same chord). These parameters essentially comprise the range of expressive factors that are controlled by the performer in harpsichord music (excluding registration effects, which were controlled for in this experiment). **Table 1** lists the mean values for these parameters averaged over both recordings of each piece for each performer (identified by the letters A to F). Because performance errors could also potentially contribute to identifying individual performers, **Table 1** also includes the total number of performance errors for each recording.

Because the purpose of this analysis was to investigate whether there were significant differences between performers, mixed-model analyses of variance were conducted for each of the five aforementioned expressive parameters, with performer as a random factor and order of recording as a fixed factor, on the 12 excerpts from each piece that were used in the sorting task (**Table 2**). No statistical analyses were conducted on the performance errors, given that the majority of recordings did not contain a single error. The order of recording was not a significant factor for any of the expressive parameters (all p-values > 0.2). The significance of the random effects associated with each performer was assessed by comparing a model that incorporated only the fixed effect of the order of the recording to a model that also included a random intercept associated with each performer.

The IOI was computed for each dotted quarter note for all *Bergeries* excerpts, and for each quarter note for all *Partita* excerpts. The mean value obtained for each excerpt was used as a measure of tempo. Significant differences in mean quarter-note duration were observed between performers, with performers B being

Table 1 | Mean values for the expressive parameters and total performance errors for each performer.

Performer	A*	B	C	D*	E	F*
Bergeries						
Mean dotted quarter-note duration (ms)	1361 (3)	1701 (40)	1293 (39)	1513 (16)	1572 (28)	1630 (9)
Mean coefficient of variation of dotted quarter-note duration (%)	6.7 (2.8)	14.8 (0.1)	11.1 (1.3)	20.9 (0.3)	10.0 (1.5)	12.1 (0.1)
Mean overlap (% of note duration)	29.3 (3.0)	23.3 (1.5)	31.4 (6.8)	4.0 (0.7)	8.1 (1.0)	10.0 (1.3)
Mean root-mean-square asynchrony (ms)	15 (4)	36 (1)	31 (6)	23 (1)	6 (3)	15 (1)
Mean velocity (MIDI units)	59 (0)	51 (0)	58 (0)	58 (1)	63 (0)	60 (1)
Total performance errors (for each recording)	0; 0	0; 1	4; 2	0; 1	0; 0	0; 0
Partita						
Mean quarter-note duration (ms)	1272 (19)	1287 (22)	1236 (50)	1474 (13)	1144 (28)	1218 (0)
Mean coefficient of variation of quarter-note duration (%)	6.1 (1.6)	11.5 (1.1)	7.0 (0.1)	12.0 (0.1)	10.4 (3.9)	7.0 (0.3)
Mean overlap (% of note duration)	−2.0 (2.6)	2.3 (1.7)	−12.1 (3.2)	−14.4 (1.7)	−14.6 (1.6)	−7.8 (1.2)
Mean root-mean-square asynchrony (ms)	29 (1)	55 (5)	23 (3)	66 (8)	40 (13)	33 (1)
Mean velocity (MIDI units)	48 (0)	60 (2)	55 (1)	57 (0)	63 (4)	54 (2)
Total performance errors (for each recording)	1; 2	0; 0	1; 1	0; 0	1; 0	0; 0

Prize-winners are indicated with an asterisk. Standard deviations of the mean values obtained for each of the two recordings per performer are given in parentheses.

significantly slower and C significantly faster for the *Bergeries*, and D being significantly slower for the *Partita*.

The coefficient of variation of the tempo, obtained by dividing the standard deviation of the tempo by the mean tempo and expressing the result as a percentage of the mean (dotted) quarter-note duration, was used as a measure of the degree of tempo variability. Whereas significant differences between performers were observed for the *Bergeries*, with performer A displaying a smaller amount of variation and performer D a larger one, no significant differences were found for the *Partita*.

Articulation refers to the amount of overlap between two consecutive note events n_i and n_j belonging to the same melodic line or voice. A positive overlap indicates a *legato* articulation, while a negative value represents a detached or *staccato* articulation. Because the amount of overlap varies with tempo (Repp, 1995), we chose to use the overlap ratio, defined as the ratio of the overlap between two consecutive note events and the IOI between these notes, as a measure of articulation (Bresin and Battel, 2000). Significant differences in the amount of overlap were found between performers for the *Bergeries*, with performer C playing more *legato* and D playing more detached, and for the *Partita*, with performer B playing more *legato*.

Note onset asynchrony is defined as the difference in onset time between note onsets that are notated in the musical score as synchronous (Palmer, 1989). Several measures of onset asynchrony have been constructed. Rasch (1979) proposed to use the root mean square, or standard deviation of the onset times of nominally simultaneous notes. We chose to use this measure here. Significant differences were observed between performers for the *Bergeries*, with performer B using larger asynchronies and performer E using smaller ones, and for the *Partita*, with performer D using larger asynchronies. Asynchronies were generally larger than those observed in organ performance (Gingras et al., 2011), and comparable to or even larger than the asynchronies of 15–20 ms which are typically observed in piano performance (Palmer, 1989). Given that the reported threshold for detecting onset asynchronies is around 20 ms (Hirsh, 1959), listeners could conceivably differentiate between performers on the basis of the amount of onset asynchrony.

Table 2 | Mixed-models analyses of variance on the expressive parameters.

Piece	*Bergeries*	*Partita*
Mean tempo	$\chi^2(1) = 14.10$, $p < 0.001$ B*, C*	$\chi^2(1) = 10.31$, $p = 0.001$ D**
Coefficient of variation of the tempo (%)	$\chi^2(1) = 8.93$, $p = 0.003$ A*, D*	$\chi^2(1) = 2.46$, $p = 0.117$ n.s.
Mean overlap (% of note duration)	$\chi^2(1) = 10.55$, $p = 0.001$ C*, D*	$\chi^2(1) = 8.40$, $p = 0.004$ B*
Mean root-mean-square asynchrony (ms)	$\chi^2(1) = 9.00$, $p = 0.003$ B*, E*	$\chi^2(1) = 5.80$, $p = 0.016$ D*
Mean velocity (MIDI units)	$\chi^2(1) = 17.55$, $p < 0.001$ B**, E*	$\chi^2(1) = 7.16$, $p = 0.008$ A*, E*

*Individual performers are identified by codes A to F. The significance of the random intercept effects predicted for each individual performer was assessed using two-tailed t-tests. *p < 0.05; **p < 0.01; n.s.: no significant effect.*

Finally, the mean MIDI velocity associated with the keypress corresponding to each note onset was computed for both excerpts. Significant differences were observed between performers for the *Bergeries*, with performer B using lower velocities and performer E using higher ones, and for the *Partita*, with performer A using lower velocities and performer E again using higher ones.

From these analyses, we may conclude that performers could be statistically differentiated on the basis of mean tempo, mean overlap, amount of onset asynchrony, and velocity for both pieces, and additionally on the basis of the amount of variation of the tempo in the case of the *Bergeries*.

GENERAL ASSESSMENT OF THE CATEGORIZATION ACCURACY

To assess the categorization accuracy for each participant, we compared their partitioning of the excerpts with the correct categorization solution, which corresponds to a grouping of the 24 excerpts in which all excerpts played by the same performer are grouped together and no excerpts played by different performers are grouped together. Categorization accuracy was evaluated using the adjusted Rand index (Hubert and Arabie, 1985), a chance-corrected measure of the agreement between the correct categorization solution and the grouping proposed by the participant. A positive adjusted Rand index indicates that a greater number of excerpts were grouped correctly than would be expected by chance ("chance" corresponding here to a randomly generated partition of the excerpts), whereas a negative adjusted Rand index indicates that fewer excerpts were grouped correctly than would be expected by chance, and a value of zero corresponds to chance performance. 39 participants (out of 40) performed better than chance (corresponding to a positive adjusted Rand index), with only one non-musician performing worse than chance (corresponding to a negative adjusted Rand index). Furthermore, for 17 musicians (85%) and nine non-musicians (45%), the adjusted Rand index was significantly above zero (indicating a performance significantly better than chance), one-tailed $p < 0.05$ estimated using a bootstrapped (Efron and Tibshirani, 1993) null distribution of 1,000,000 permutations with replacement with a mean adjusted Rand index of -0.0001, 95% CI $[-0.101, 0.133]$. The difference between the proportion of musicians and non-musicians who performed significantly better than chance was significant, as determined by a chi-square test, $\chi^2(1) = 7.03$, $p = 0.008$. The same proportion of male and female participants, 65% (corresponding to 13 participants for each gender) performed significantly better than chance.

To evaluate the effect of musical training and gender on sorting accuracy, we conducted an analysis of variance with musical training and gender as between-subject factors. Variances did not deviate significantly from homogeneity, as indicated by a Levene test, and the distribution of the adjusted Rand indices did not deviate from normality across factorial combinations. Musicians performed significantly better than non-musicians, $F(1,36) = 15.527$, $p < 0.001$, $\eta^2 = 0.301$. The effect of gender was not significant, $F(1,36) = 0.110$, $p = 0.742$, $\eta^2 = 0.002$, and no significant interaction was found between musical training and gender, $F(1,36) = 0.002$; $p = 0.962$, $\eta^2 = 0.000$.

Listening activity, defined by the total number of times a participant listened to the excerpts, was found to be significantly correlated with categorization accuracy, $r(38) = 0.480$, $p = 0.002$. The correlation was stronger for musicians, $r(18) = 0.458$, $p = 0.043$, than for non-musicians, $r(18) = 0.332$, $p = 0.152$, but this difference was not significant as determined by a Z-test on the Fisher-transformed correlation coefficients ($z = 0.43$, $p = 0.664$). Very similar results were obtained when correlating the total amount of time spent on the sorting task with the categorization accuracy (the total amount of time spent on the task was highly correlated with listening activity, $r(38) = 0.988$, $p < 0.001$). Although musicians listened to more excerpts than non-musicians on average, the difference was not statistically significant, as shown by a two-tailed Mann-Whitney test (the data was not normally distributed), $U = 168.0$, $p = 0.394$.

EFFECT OF PIECE AND PERFORMER EXPERTISE ON CATEGORIZATION ACCURACY

To assess whether the ability of participants to correctly sort excerpts varied according to the piece, and to compare the participants' ability to correctly group together excerpts from the same performer and the same piece versus same performer/different pieces, the participants' partitions were decomposed by taking into account the performers and the pieces corresponding to the excerpts that were grouped together. Such analyses involve comparisons of pairs of excerpts (Miller, 1969; Daws, 1996). The proportion of pairs of excerpts correctly grouped together (pairs played by the same performer and identified as such) out of the total number of pairs was then computed for the following types of pairs (**Figure 3**):

(a) One *Bergeries* and one *Partita* excerpt from the same performer (B/P).
(b) Two *Partita* excerpts from the same performer (P/P).
(c) Two *Bergeries* excerpts from the same performer (B/B).

Combinatorial probabilities are used to determine the chance performance level. In the present case, a partition of 24 excerpts into six groups of four excerpts yields 36 pairs of excerpts, given by $6 \times [4!/(2! \times 2!)]$. The total number of possible pairs is given by $24!/(22! \times 2!)$, yielding 276 pairs. These 276 pairs can be further decomposed in 66 possible pairs comprising two *Bergeries* excerpts, given by $12!/(10! \times 2!)$, 66 possible pairs comprising two *Partita* excerpts, and the remaining 144 pairs which contain one excerpt from each piece.

To compute the chance performance level, we need to estimate the probability of randomly assigning a pair to a given partition. Here, the chance performance level, which corresponds to the probability of randomly assigning a pair to a given partition, differs between pairs containing one excerpt from each piece and pairs comprising two excerpts from the same piece. Because participants were constrained to assign exactly two *Bergeries* excerpts and two *Partita* excerpts to each performer, each partition always included exactly six pairs comprising two excerpts from the *Bergeries*, six pairs comprising two excerpts from the *Partita*, and 24 pairs comprising one excerpt from each piece (note that four such combinations are formed when assigning two *Bergeries* excerpts and two *Partita* excerpts to a performer, hence the 24 pairs obtained for the six performers). Therefore, the chance performance level for same-piece pairs is equivalent to 6/66, or $p = 0.091$, whereas the

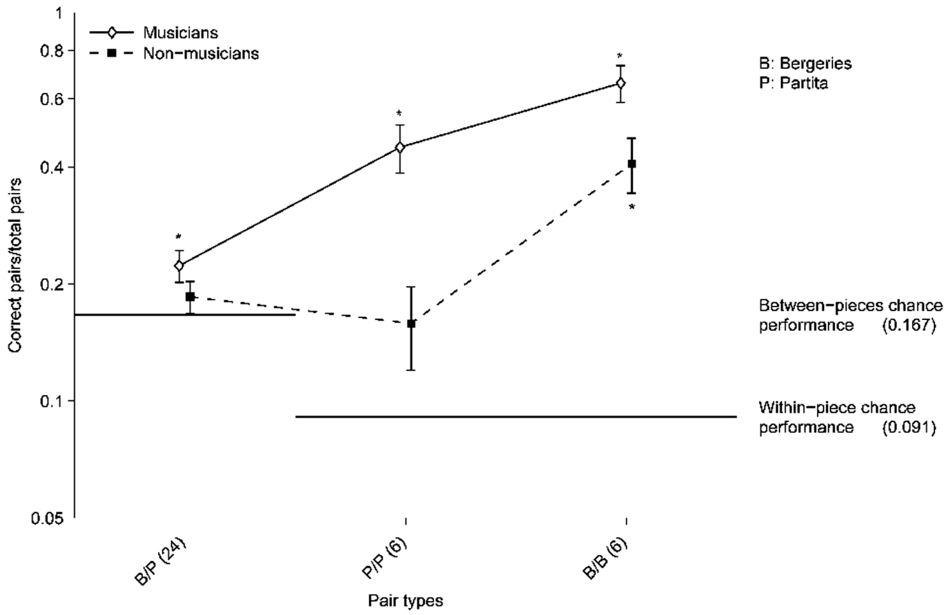


FIGURE 3 | Proportion of correct pairs compared to the total number of pairs for all pair types. The total number of pairs is indicated in parentheses for each pair type. B/P: pairs corresponding to one *Bergeries* and one *Partita* excerpt from the same performer. P/P: pairs corresponding to two *Partita* excerpts from the same performer. B/B: pairs corresponding to two *Bergeries* excerpts from the same performer (B/B). Error bars indicate standard errors of the mean. Asterisks indicate values that are significantly different from chance performance (for both musicians and non-musicians) as determined by one-tailed *t*-tests with Bonferroni correction. Points are connected to help distinguish the two participants group visually.

chance performance level for *Bergeries–Partita* pairs is equivalent to 24/144, or $p = 0.167$.

The proportion of correct pairs was significantly above chance for *Bergeries–Bergeries* pairs, for both musicians and non-musicians, as determined by one-tailed *t*-tests with Bonferroni correction (the distributions did not deviate significantly from normality). However, only musicians performed significantly above chance in the case of the *Partita–Partita* pairs and in the case of the *Bergeries–Partita* pairs. Note that the *t*-test for the *Partita–Partita* pairs for non-musicians barely reached significance ($p = 0.043$) prior to the Bonferroni correction.

Because chance performance levels depend on the pair composition, further analyses comparing across pair types were conducted on the chance-corrected proportions of correct pairs. Additionally, in order to estimate the effect of the performers' expertise (prize-winners versus non-prize-winners) on the participants' performance in the sorting task, the categorization accuracy for the prize-winners was compared with that observed for the non-prize-winners. Furthermore, because listening activity was significantly correlated with categorization accuracy, it was included as a covariate in subsequent analyses. Gender was excluded because previous analyses had indicated that it was not a significant factor. Thus, a repeated-measures logistic regression analysis on the chance-corrected proportion of correct pairs was conducted, with participants' musical training as a between-subjects factor, performer expertise and pair composition (*Bergeries–Bergeries*, *Partita–Partita*, and *Bergeries–Partita*) as within-subject factors, and listening activity as a continuous covariate associated with each participant. Chance performance for each pair type was added to the model as an offset, following the procedure described in Lipsitz et al. (2003). We verified that the only parameter estimate that was affected by adding this offset to the logistic regression model was the main effect of pair composition. In line with prior analyses, significant effects were observed for musical training, $\chi^2(1) = 5.72$, $p = 0.017$, pair composition, $\chi^2(2) = 483.02$, $p < 0.001$, and listening activity, $\chi^2(1) = 7.45$, $p = 0.006$. In addition, a significant interaction between musical training and pair composition was found, $\chi^2(2) = 16.28$, $p < 0.001$, which corresponds to the larger effect of musical training observed on the *Partita–Partita* pairs compared to the *Bergeries–Bergeries* pairs (**Figure 3**). However, performer expertise was not a significant predictor of the chance-corrected proportion of correct pairs, $\chi^2(1) = 0.09$, $p = 0.763$. No other interactions reached significance (all *p*-values > 0.25).

Post hoc tests using the Bonferroni correction procedure were used to compare the chance-corrected proportion of correct pairs for the three different types of pairs, collapsing over levels of musical training. All pairwise comparisons were significant, indicating that the chance-corrected accuracy for *Bergeries–Bergeries* pairs was significantly higher than for the other two pair types, and that the accuracy for *Partita–Partita* pairs was significantly higher than for *Bergeries–Partita* pairs, all Bonferroni-corrected *p*-values < 0.001.

PERCEPTUAL DISTANCE BETWEEN THE EXCERPTS AND PERFORMANCE PARAMETERS

In order to assess whether performance parameters such as articulation, asynchrony, tempo, or velocity could explain the perceived distance between excerpts, with perceptually distant excerpts corresponding to excerpts rarely or never grouped together and "close" excerpts (excerpts thought to have been played by the same performer) representing excerpts always grouped together, we conducted a distance-based multivariate regression using forward selection (McArdle and Anderson, 2001). This analysis seeks to model the proportion of variance in the perceptual distance between excerpts (obtained from the co-occurrence matrix) that is explained by the mean values of the performance parameters obtained for each excerpt (see **Table 1**). Because significance testing for distance-based multivariate regression is done through a permutation procedure, degrees of freedom are not reported. Furthermore, the *F*-ratios yielded by this procedure do not exactly correspond to the *F*-ratios obtained in a traditional analysis of variance and are thus labeled "pseudo-*F* ratios." We used the DISTLM-forward program (Anderson, 2003) to conduct these analyses. The forward selection procedure used the proportion of the variance explained by each expressive parameter as criterion for selection. For all analyses, 99,999 permutations were conducted to test for statistical significance. **Table 3** reports the results of the distance-based multivariate regression analyses conducted separately for each piece and between both pieces.

Although the influence of multicollinearity on the model results cannot be evaluated directly when using the DISTLM procedure (Link et al., 2013), highly correlated independent variables can lead to spurious results about their relationships with the dependent variable (Zar, 1999). Thus, performance parameters that were highly correlated ($|r| > 0.7$) were not included in the forward-selection model (Zar, 1999), although they were included in the marginal tests because we cannot exclude the possibility that these parameters play a role in the listeners' evaluation of perceptual distance simply due to the presence of multicollinearity (see **Table 4** for the correlation matrices). For each pair of highly correlated variables, the variable with the highest correlations with the remaining independent variables was excluded from the forward-selection procedure. In the case of the *Bergeries*, velocity was strongly inversely correlated with asynchrony, $r(10) = -0.86$, $p < 0.001$, and was excluded. For the *Partita*, the coefficient of variation of the tempo was highly correlated with both asynchrony, $r(10) = 0.67$, $p = 0.017$, and velocity, $r(10) = 0.77$, $p = 0.003$, and

Table 3 | Distance-based multivariate regression.

	Marginal tests			Sequential tests		
	Pseudo-*F*	**Variance**	***p***	**Pseudo-*F***	**Variance**	***p***
Bergeries						
Mean tempo	2.53	0.202	<0.001	2.53	0.202 (1)	<0.001
Coefficient of variation of the tempo (%)	2.01	0.167	0.007	2.12	0.152 (2)	0.023
Mean overlap (% of note duration)	2.04	0.170	0.007	0.78	0.051 (4)	0.586
Mean root-mean-square asynchrony (ms)	1.66	0.142	0.061	2.23	0.141 (3)	0.036
Mean velocity (MIDI units)*	1.74	0.149	0.052	N/A	N/A	N/A
Total variance explained by significant predictors in the forward-stepwise model: 0.495						
Partita						
Mean tempo	2.25	0.184	<0.001	2.14	0.140 (3)	0.004
Coefficient of variation of the tempo (%)*	1.95	0.163	0.006	N/A	N/A	N/A
Mean overlap (% of note duration)	1.67	0.143	0.024	1.95	0.144 (2)	0.003
Mean root-mean-square asynchrony (ms)	2.36	0.191	<0.001	2.36	0.191 (1)	<0.001
Mean velocity (MIDI units)	1.88	0.158	0.004	1.41	0.088 (4)	0.150
Total variance explained by significant predictors in the forward-stepwise model: 0.476						
Between pieces						
Mean tempo	2.12	0.088	<0.001	2.12	0.088 (1)	<0.001
Coefficient of variation of the tempo (%)	1.22	0.052	0.158	1.02	0.040 (4)	0.494
Mean overlap (% of note duration)	1.09	0.047	0.341	1.33	0.052 (3)	0.074
Mean root-mean-square asynchrony (ms)	2.03	0.085	<0.001	2.06	0.082 (2)	<0.001
Mean velocity (MIDI units)	0.31	0.014	0.998	0.30	0.012 (5)	0.992
Total variance explained by significant predictors in the forward-stepwise model: 0.170						

Marginal tests correspond to tests conducted separately on each expressive parameter taken in isolation. Sequential tests represent the relative contribution of each parameter after taking into account other parameters already included in the model, based on a forward selection procedure using the proportion of variance explained by each parameter as the criterion for selection (the order of entry is given in parentheses). Parameters marked with an asterisk were not included in the sequential tests due to multicollinearity.

Table 4 | Correlation matrices on the mean values for the expressive parameters.

	Mean tempo	Coefficient of variation of the tempo (%)	Mean overlap (% of note duration)	Mean root-mean-square asynchrony (ms)
Bergeries				
Mean tempo	1			
Coefficient of variation of the tempo (%)	0.37	1		
Mean overlap (% of note duration)	−0.53	−0.54	1	
Mean root-mean-square asynchrony (ms)	0.00	0.41	0.48	1
Mean velocity (MIDI units)	−0.29	−0.34	−0.38	−0.86
Partita				
Mean tempo	1			
Coefficient of variation of the tempo (%)	0.41	1		
Mean overlap (% of note duration)	−0.02	−0.17	1	
Mean root-mean-square asynchrony (ms)	0.64	0.67	0.02	1
Mean velocity (MIDI units)	−0.15	0.77	−0.34	0.39

Correlations computed on the mean values for each expressive parameter for each excerpt.

was excluded. No other performance parameters were highly correlated.

In the case of the *Bergeries*, tempo, tempo variation, and overlap were significant predictors of the perceptual distance between excerpts according to the marginal tests. Tempo, tempo variation, and asynchrony were significant predictors in the forward selection model (sequential tests). Separate analyses for musicians and non-musicians were also conducted to examine potential differences between groups (not shown in **Table 3**). The results for musicians were similar to the results on the entire group of participants, whereas asynchrony was also significant in the marginal tests for non-musicians, and the forward selection model included only tempo and asynchrony as significant predictors.

In the case of the *Partita*, all parameters were significant predictors of the perceptual distance between excerpts according to the marginal tests. Asynchrony, overlap, and tempo were significant predictors in the forward selection model. The results for musicians on the marginal tests were similar to the results on the entire group of participants, but the forward selection model included overlap, asynchrony, and velocity as significant predictors. All parameters except overlap were significant in the marginal tests for non-musicians, whereas tempo, asynchrony, and overlap were significant in the forward selection model.

To evaluate the relationship between the perceptual distance between both pieces and the performance parameters, standardized values (z-scores) were used for the parameters. Moreover, the distances for all within-piece comparisons were set to chance performance, thus leaving the sum of the distance matrix elements unchanged but confining the variance to the between-pieces quadrants (note that all distance-based regression models are based on square, symmetrical distance matrices and thus require some type of algebraic manipulation in order to enable the type of between-pieces comparison conducted here; similar problems arise with related methods such as redundancy analysis or non-metric multidimensional scaling). Tempo and asynchrony were significantly correlated with perceived distance according to marginal tests, and both parameters were significant predictors in the forward selection model. Similar results were obtained for musicians and non-musicians. The proportion of variance explained by the forward-stepwise model on all participants was considerably smaller (0.170) than that explained by the models considering only one piece (respectively 0.495 for the *Bergeries* and 0.476 for the *Partita*), in line with the observation that participants performed on average barely above chance in this situation (especially in the case of non-musicians).

The analyses conducted in this section have, until now, focused solely on the mean values for the performance parameters, computed over an entire excerpt. However, it is also plausible that listeners would pay attention to note-by-note (or event-by-event) expressive profiles, and that two excerpts with similar profiles would be judged as more likely to have been played by the same performer. The magnitude of the correlations between the expressive profiles corresponding to different performers can be used to evaluate the degree of similarity between these profiles. Hence, following the method outlined in Gingras et al. (2013), we computed Kendall's tau correlations between all pairs of performers for each piece and for each performance parameter (to avoid pseudoreplication, the values for the two recordings associated with each piece were averaged before computing the correlations). Four performance parameters were considered: tempo, overlap, asynchrony, and velocity. The correlation matrices thus obtained for each parameter were used as similarity matrices. Mantel tests were conducted to evaluate the degree of similarity between the similarity matrices corresponding to the expressive profiles associated with each performance parameter on the one hand, and the co-occurrence matrix corresponding to the perceptual distance between excerpts as judged by listeners on the other hand. The statistical significance of the Mantel tests was assessed using the Bonferroni correction procedure.

In the case of the *Bergeries*, the similarity matrices corresponding to the expressive profiles did not correlate significantly with the perceptual distance between excerpts, with the exception of asynchrony. However, although the uncorrected p-value for the asynchrony matrix was significant ($p = 0.014$ before the Bonferroni adjustment), the correlation was negative (Mantel $r = -0.52$), which means that this association was probably an artifact of another relationship as it is unlikely that listeners would group together excerpts whose asynchrony profiles differed markedly. Similar results were obtained for both musicians and non-musicians.

In the case of the *Partita*, no correlation between the similarity matrices corresponding to the expressive profiles and the co-occurrence matrix reached significance (all uncorrected p-values > 0.2). Similar results were observed for both musicians and non-musicians. These results suggest that, for either the *Bergeries* or the *Partita*, listeners did not rely on the degree of similarity between note-by-note expressive profiles when grouping excerpts together.

DISCUSSION

This study examined whether listeners are able to accurately group together short excerpts from two different harpsichord pieces (Couperin's *Bergeries* and Frescobaldi's *Partita*) played by the same performer, while taking into consideration both performer and listener expertise. Although most participants reported that they experienced the sorting task as being very difficult, an analysis of the categorization accuracy of individual participants revealed that 39 of 40 participants performed above chance, with 26 participants at a level significantly better than chance. As in earlier work by Gingras et al. (2011), musicians performed better than non-musicians on the sorting task, but in this case the difference between the two groups was significant, whereas it did not reach significance in the previous study. This result overlaps with Davidson's (1993) findings that music students performed significantly better than non-musicians when asked to distinguish between the performance manners of different pianists and violinists, suggesting that musical training has an important role in recognizing personal characteristics in a short musical excerpt. More generally, our findings are in line with Ashby and Maddox's (1992) observations that novices are generally less accurate than experienced categorizers.

The influence of musical training may have been stronger here than in Gingras et al. (2011) due to the fact that the participants had to compare two different pieces here, instead of two different interpretations of the same piece as in the earlier study. Indeed, whereas both musicians and non-musicians performed significantly better than chance on the *Bergeries–Bergeries* pairings, only musicians performed significantly better than chance when considering the *Partita–Partita* or *Bergeries–Partita* pairings. These findings suggest that the effect of musical training may be stronger with some pieces or musical styles, and that successfully matching excerpts from two different pieces to the same performer may require extensive musical training. The musicians' familiarity with specific musical cues may have had a positive impact on their performance on the task. Moreover, musicians may also have a better ability to retain the characteristics of the excerpts in memory, although this remains to be evaluated. Commitment to the task (as shown by the increased amount of time spent listening to the excerpts) was also shown to be a good predictor of the participants' performance, replicating the results reported in Gingras et al. (2011). However, gender was not related to the ability to perceive artistic individuality. Indeed, male and female participants performed very similarly on average.

We hypothesized that there would be an effect of performers' level of musical expertise on the listeners' grouping accuracy, that is, excerpts played by prize-winning performers would be easier to group accurately than excerpts played by non-prize-winners. This assumption was based on earlier results in a similar task (Gingras et al., 2011). However, our results did not show any significant effect of performers' expertise. It is possible that the consistency and distinctiveness of the performers, which were associated with the level of expertise in Gingras et al. (2011), were not as relevant here, especially since two different pieces were compared.

Because the two pieces selected for this experiment differed in tempo, melody, texture, duration, and meter, we considered the possibility that excerpts from one piece might be easier or more difficult to sort accurately than excerpts from the other piece. Indeed, the results suggested that there was a significant difference between the two pieces in terms of grouping accuracy, the *Bergeries* excerpts proving to be sorted more accurately. The question of interest here is explaining what made one piece more easily recognizable than the other. Although the results presented here do not provide a direct answer to that question, they do provide some plausible interpretations. As shown in **Table 2**, significant differences between performers could be found for all five expressive parameters analyzed here in the case of the *Bergeries*, whereas only four parameters yielded significant differences between performers for the *Partita*. Moreover, at least two performers were significantly different from the mean for each parameter in the case of the *Bergeries*, suggesting the possibility to differentiate perceptually between these performers based on the parameter in question, in contrast to the *Partita* where in most cases only one performer was significantly different from the mean. Another explanation for the better grouping accuracy observed for the *Bergeries* may be simply that the *Bergeries* excerpts contained more notes (75 versus 37 for the *Partita*) and were a few seconds longer on average than the *Partita* excerpts (11–15 s for the *Bergeries* versus 8–11 s for the *Partita*), thus giving participants more time to recognize the distinctive features of an excerpt. Although the fact that the excerpts did not have exactly the same length for both pieces may be considered a potential impediment when comparing the performance on the two pieces, we deemed it important to select stimuli that represented complete musical units with a sense of closure, and the excerpts chosen likely constituted the most appropriate selection in that regard. A further explanation for the difference in sorting accuracy between both pieces could be the greater distinctiveness of the *Bergeries* fragment selected for the experiment, with its lighter texture, regular rhythm, and clear melody, making it *a priori* easier to process than the corresponding *Partita* fragment, with its more complex polyphonic texture. In that regard, it is noteworthy that the difference in sorting accuracy observed between musicians and non-musicians was much more manifest in the case of the *Partita* than for the *Bergeries*, suggesting that non-musicians were especially affected by the difference in texture between the two pieces. However, it should be noted that both non-musicians and musicians performed above chance in a comparable task using an excerpt of organ music whose length, number of notes, style, complexity, and polyphonic texture were very similar to that of the *Partita* (compare **Figure 2** with **Figure 1** in Gingras et al., 2011).

We were also interested in examining whether listeners fared better in matching the excerpts from the same pieces (either the *Bergeries* or the *Partita*) or the excerpts from different pieces but played by the same performer. A significant difference in the sorting accuracy was found between grouping the excerpts from the same piece and grouping excerpts from different pieces (in addition to the difference observed between the *Bergeries* and the *Partita* described above) after correcting for chance performance, as indicated by *post hoc* tests. Moreover, only musicians performed significantly above chance when considering only pairings of excerpts from different pieces. These results indicate that participants were not very successful in matching excerpts from different pieces, especially in the case of non-musicians. Nevertheless, the fact that musicians could perform above chance in this situation suggests that some distinctive features associated with a performer's specific playing style can be recognized across different pieces, even in the case of unfamiliar performers (thus extending the work of Benadon, 2003 on famous performers) and on an instrument limiting the use of expressive strategies associated with timbre and dynamics. These results are in line with the findings reported in Gingras et al. (2013), and with the earlier work by Stamatatos and Widmer (2005) using an artificial intelligence approach. Although a direct comparison with the results reported by Stamatatos and Widmer is not possible due to the different experimental design, our results suggest that these authors were apparently correct to note that human listeners were not likely to match the accuracy of a learning ensemble when attempting to sort performances of two different pieces by their performer.

Additionally, we investigated the relationship between performance parameters and the perceptual distance between performers as established from the results of the sorting task using distance-based multivariate regression analysis. Although the resulting regression models differed between pieces, as well as between musicians and non-musicians in some cases, some overarching conclusions could nevertheless be gleaned from these analyses. First, we note that marginal tests for tempo were significant in all analyses, and that tempo entered practically all forward-stepwise models (the only exception being the *Partita* in the case of musicians). The fact that tempo was used prominently by listeners in such a task is in line with earlier results (Repp and Knoblich, 2004; Gingras et al., 2011). Second, non-musicians appeared to rely on note onset asynchrony to a greater extent than musicians: asynchrony entered all models for non-musicians, but was only a significant predictor in the case of the *Partita* and the between-pieces comparisons for musicians. Third, tempo and asynchrony were the only significant predictors of perceptual distance when comparing between pieces, for both musicians and non-musicians. This suggests that listeners found it difficult to rely on other expressive features such as overlap, velocity, or tempo variation when comparing across different pieces. In the case of tempo variation, the fact that the two pieces were written in a different meter, apart from the considerable textural differences, may explain why listeners did not rely on this parameter. On the other hand, in the case of velocity, some performers were consistent across both pieces (performer E, for instance, used significantly higher velocities than other performers in both pieces). However, it is possible that these differences in velocity, which lead to contrasts in sound intensity amounting to a few dB at most on the harpsichord (Penttinen, 2006), could not be easily perceived by listeners. Although overlap entered a few regression models, it appeared to be a generally secondary parameter, especially in comparison to organ performance where overlap was, along with tempo, a major predictor of perceptual distance between performers (Gingras et al., 2011). This may be partially explained by the fact that the harpsichord sound decays relatively quickly, at least in comparison to other keyboard instruments such as the piano or organ, thus making the contrast between *legato* and *staccato* articulation less striking on this instrument. Finally, correlational analyses between note-by-note expressive patterns and perceptual distances (as obtained from the sorting task) suggested that listeners did not appear to rely on note-by-note patterns in the sorting task. This leads us to surmise that they relied mostly on a holistic impression of the excerpts, which is captured in a rough manner by the distance-based multivariate regression models based on the mean values computed for each of the expressive parameters analyzed here.

Because performance errors could also conceivably contribute to the identification of individual performers, we listed the total number of performance errors for each recording in **Table 1**. As can be seen, the error totals were very low. Most of these errors (11 of 14) consisted in omissions, meaning that a note present in the score was not played. Such "silent" errors are likely to be inconspicuous and, as shown by Repp (1996), most performance errors are typically difficult to detect, even for trained musicians. Additionally, none of these errors occurred in the highest voice (or part), causing them to be less noticeable (Palmer and Holleran, 1994). Indeed, author Bruno Gingras, a trained musicologist, could not detect most of these performance errors even when listening to the recordings while following with the score (note that participants in the sorting task did not have access to the score of the pieces). For these reasons, it is very unlikely that performance errors could have been used reliably by listeners to discriminate between performers. One performer (harpsichordist C) committed a somewhat larger number of errors in the *Bergeries* excerpts, but all of these errors consisted in omissions in the left-hand part (lower voice) and were thus not conspicuous.

In conclusion, very few studies have so far investigated the ability of humans to process identity cues in music performance, especially with unfamiliar performers. To our knowledge, the present study is the first empirical study that investigated the participants' ability to accurately discriminate between unfamiliar performers playing excerpts from two different pieces. The study by Gingras et al. (2011), on which the current work was modeled, served as a good benchmark for comparison. Both studies showed that most participants, both musicians and non-musicians, are able to recognize and process identity cues in short musical excerpts (of approximately 10–15 s in both studies) and to correctly group excerpts that are played by the same performer at a level better than chance. Both studies also showed that sorting accuracy was significantly correlated with the time spent doing the sorting task. Although musicians performed better than non-musicians in both studies, the effect only reached significance in the present case. However, whereas Gingras et al. (2011) reported an effect of performer expertise, no such effect was observed here. Moreover, the present study underscored that the choice of musical excerpts may exert an important influence on the sorting accuracy. Nevertheless, the fact that these studies yielded generally similar results, even though the experiments were conducted using a different instrument and stylistic repertoire, in addition to being carried out in different countries, suggest that the findings are indeed valid and reliable. Overall, our results indicate that the performers' expertise may not be as essential in predicting individual recognition as is the musical background of the participants and the characteristics of the excerpts when more than one piece is involved (however, the influence of the performers' expertise should ideally be evaluated using a larger group of performers, as well as other measures of expertise besides performance prizes). Moreover, our findings suggest that specific features associated with a piece may play a crucial role in enabling listeners to pick up on its characteristics and recognize the identity of the performer, and that extensive musical training may be a prerequisite for perceiving identity cues across different pieces, at least in the case of short excerpts played by unfamiliar performers.

ACKNOWLEDGMENTS

This research was supported by a SSHRC postdoctoral fellowship to Bruno Gingras. We wish to thank Alain Poirier, former director of the Conservatoire National Supérieur de Musique et de Danse de Paris, for the loan of the MIDI harpsichord, Stephen McAdams, for the use of the CIRMMT facilities at McGill University (Montreal), and Lauren Stewart, for the use of the audiometer and sound-attenuated room at Goldsmiths College (London). We also thank Pierre-Yves Asselin, Bennett Smith, Meghan Goodchild, Julien Boissinot, and Harold Kilianski for their technical assistance, McGill Sound Recording for permission to use their equipment, and Martha de Francisco for her assistance with sound recording.

REFERENCES

Anderson, M. J. (2003). *DISTLM Forward: A FORTRAN Computer Program to Calculate a Distance-based Multivariate Analysis for a Linear Model using Forward Selection.* Department of Statistics, University of Auckland, New Zealand.

Andrews, T. J., and Ewbank, M. P. (2004). Distinct representations for facial identity and changeable aspects of faces in the human temporal lobe. *Neuroimage* 23, 905–913. doi: 10.1016/j.neuroimage.2004.07.060

Ashby, F. G., and Lee, W. W. (1991). Predicting similarity and categorization from identification. *J. Exp. Psychol. Gen.* 120, 150–172. doi: 10.1037/0096-3445.120.2.150

Ashby, F. G., and Maddox, W. T. (1992). Complex decision rules in categorization: contrasting novice and experienced performance. *J. Exp. Psychol. Hum. Percept. Perform.* 18, 50–71. doi: 10.1037/0096-1523.18.1.50

Ashby, F. G., and Maddox, W. T. (2005). Human category learning. *Annu. Rev. Psychol.* 56, 149–178. doi: 10.1146/annurev.psych.56.091103.070217

Ashby, F. G., and Perrin, N. A. (1988). Toward a unified theory of similarity and recognition. *Psychol. Rev.* 95, 124–150. doi: 10.1037/0033-295X.95.1.124

Belin, P., Fecteau, S., and Bédard, C. (2004). Thinking the voice: neural correlates of voice perception. *Trends Cogn. Sci.* 8, 129–135. doi: 10.1016/j.tics.2004.01.008

Benadon, F. (2003). "Spectrographic and calligraphic cues in the identification of jazz saxophonists," in *Proceedings of the Fifth ESCOM Conference*, eds R. Kopiez, A. C. Lehmann, I. Wolther, and C. Wolf (Osnabrück: epOs-Music).

Bigand, E., and Poulin-Charronnat, B. (2006). Are we experienced listeners? A review of the musical capacities that do not depend on formal musical training. *Cognition* 100, 100–130. doi: 10.1016/j.cognition.2005.11.007

Blake, R., and Shiffrar, M. (2007). Perception of human motion. *Annu. Rev. Psychol.* 58, 47–73. doi: 10.1146/annurev.psych.57.102904.190152

Blatchford, H., and Foulkes, P. (2006). Identification of voices in shouting. *Int. J. Speech Lang. Law* 13, 241–254. doi: 10.1558/ijsll.2006.13.2.241

Bresin, R., and Battel, G. U. (2000). Articulation strategies in expressive piano performance: analysis of legato, staccato, and repeated notes in performances of the Andante movement of Mozart's Sonata in G major (K 545). *J. New Music Res.* 29, 211–224. doi: 10.1076/jnmr.29.3.211.3092

Carey, S. (1992). Becoming a face expert. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 335, 95–103. doi: 10.1098/rstb.1992.0012

Chartrand, J. P., Peretz, I., and Belin, P. (2008). Auditory recognition expertise and domain specificity. *Brain Res.* 1220, 191–198. doi: 10.1016/j.brainres.2008.01.014

Cutting, E. J., and Kozlowski, T. L. (1977). Recognizing friends by their walk: gait perception without familiarity cues. *Bull. Psych. Soc.* 9, 353–356. doi: 10.3758/BF03337021

Davidson, J. W. (1993). Visual perception of performance manner in the movements of solo musicians. *Psychol. Music* 21, 103–113. doi: 10.1177/030573569302100201

Daws, J. T. (1996). The analysis of free-sorting data: beyond pairwise cooccurrences. *J. Classif.* 13, 57–80. doi: 10.1007/BF01202582

Efron, B., and Tibshirani, R. (1993). *An Introduction to the Bootstrap.* New York: Chapman & Hall. doi: 10.1007/978-1-4899-4541-9

Ell, S. W., and Ashby, F. G. (2012). The impact of category separation on unsupervised categorization. *Atten. Percept. Psychophys.* 74, 466–475. doi: 10.3758/s13414-011-0238-z

Flach, R., Knoblich, G., and Prinz, W. (2004). Recognising one's own clapping: the role of temporal cues. *Psychol. Res.* 69, 147–156. doi: 10.1007/s00426-003-0165-2

Gabrielsson, A. (2003). Music performance research at the millenium. *Psychol. Music* 31, 221–272. doi: 10.1177/03057356030313002

Gingras, B. (2008). *Expressive Strategies and Performer-Listener Communication in Organ Performance.* Unpublished Ph.D. dissertation, McGill University, Montreal.

Gingras, B., Asselin, P.-Y., and McAdams, S. (2013). Individuality in harpsichord performance: disentangling performer- and piece-specific influences on interpretive choices. *Front. Psychol.* 4:895. doi: 10.3389/fpsyg.2013.00895

Gingras, B., Lagrandeur-Ponce, T., Giordano, B. L., and McAdams, S. (2011). Perceiving musical individuality: performer identification is dependent on performer expertise and expressiveness, but not on listener expertise. *Perception* 40, 1206–1220. doi: 10.1068/p6891

Gingras, B., and McAdams, S. (2011). Improved score-performance matching using both structural and temporal information from MIDI recordings. *J. New Music Res.* 40, 43–57. doi: 10.1080/09298215.2010.545422

Gingras, B., McAdams, S., Schubert, P., and Utz, C. (2010). "The performer as analyst: a case study of JS Bach's "Dorian" Fugue (BWV 538)," in *Music Theory and Interdisciplinarity – Eighth Congress of the Gesellschaft für Musiktheorie Graz 2008* (Germany: Pfau-Verlag Saarbrücken), 305–318.

Haxby, J. V., Hoffman, E. A., and Gobbini, M. I. (2000). The distributed human neural system for face perception. *Trends Cogn. Sci.* 4, 223–233. doi: 10.1016/S1364-6613(00)01482-0

Hirsh, I. J. (1959). Auditory perception of temporal order. *J. Acoust. Soc. Am.* 31, 759–767. doi: 10.1121/1.1907782

Hubert, L., and Arabie, P. (1985). Comparing partitions. *J. Classif.* 2, 193–218. doi: 10.1007/BF01908075

Johansson, G. (1973). Visual perception of biological motion and a model for its analysis. *Percept. Psychophys.* 14, 201–211. doi: 10.3758/BF03212378

Juslin, P. N., and Sloboda, J. A. (2001). *Music and Emotion.* New York: Oxford University Press.

Kendall, R. A., and Carterette, E. C. (1990). The communication of musical expression. *Music Percept.* 8, 129–164. doi: 10.2307/40285493

Lehmann, A. C., Sloboda, J. A., and Woody, R. H. (2007). *Psychology for Musicians.* New York: Oxford University Press.

Link, H., Chaillou, G., Forest, A., Piepenburg, D., and Archambault, A. (2013). Multivariate benthic ecosystem functioning in the Arctic – benthic fluxes explained by environmental parameters in the southeastern Beaufort sea. *Biogeosciences* 10, 5911–5929. doi: 10.5194/bg-10-5911-2013

Lipsitz, S. R., Parzen, M., and Fitzmaurice, G. M. (2003). A two-stage logistic regression model for analyzing inter-rater agreement. *Psychometrika* 68, 289–298. doi: 10.1007/BF02294802

Loula, F., Prasad, S., Harber, K., and Shiffrar, M. (2005). Recognising people from their movement. *J. Exp. Psychol. Hum. Percept. Perform.* 31, 210–220. doi: 10.1037/0096-1523.31.1.210

McArdle, B. H., and Anderson, M. J. (2001). Fitting multivariate models to community data: a comment on distance-based redundancy analysis. *Ecology* 82, 290–297. doi: 10.1890/0012-9658(2001)082[0290:FMMTCD]2.0.CO;2

Miller, G. A. (1969). A psychological method to investigate verbal concepts. *J. Math. Psychol.* 6, 169–191. doi: 10.1016/0022-2496(69)90001-7

Palmer, C. (1989). Mapping musical thought to musical performance. *J. Exp. Psychol. Hum. Percept. Perform.* 15, 331–346. doi: 10.1037/0096-1523.15.2.331

Palmer, C., and Holleran, S. (1994). Harmonic, melodic, and frequency height influences in the perception of multivoiced music. *Percept. Psychophys.* 56, 301–312. doi: 10.3758/BF03209764

Palmer, C., Jungers, M. K., and Jusczyk, P. W. (2001). Episodic memory for musical prosody. *J. Mem. Lang.* 45, 526–545. doi: 10.1006/jmla.2000.2780

Penttinen, H. (2006). "On the dynamics of the harpsichord and its synthesis," in *Proceedings of Ninth International Conference on Digital Audio Effects (DAFx-06)*, ed. V. Verfaille (Montreal), 115–120. Available at: http://www.dafx.ca/proceedings/dafx06_cite.pdf

Rasch, R. A. (1979). Synchronization in performed ensemble music. *Acustica* 43, 121–131.

Repp, B. H. (1987). The sound of two hands clapping: an exploratory study. *J. Acoust. Soc. Am.* 81, 1100–1109. doi: 10.1121/1.394630

Repp, B. H. (1992). Diversity and commonality in music performance – an analysis of timing microstructure in Schumann's "Träumerei." *J. Acoust. Soc. Am.* 92, 2546–2568. doi: 10.1121/1.404425

Repp, B. H. (1995). Acoustics, perception, and production of legato articulation on a digital piano. *J. Acoust. Soc. Am.* 97, 3862–3874. doi: 10.1121/1.413065

Repp, B. H. (1996). The art of inaccuracy: why pianists' errors are difficult to hear. *Music Percept.* 14, 161–183. doi: 10.2307/40285716

Repp, B. H. (1997). The aesthetic quality of a quantitatively average music performance: two preliminary experiments. *Music Percept.* 14, 419–444.

Repp, B. H., and Keller, P. E. (2010). Self versus other in piano performance: detectability of timing perturbations depends on personal playing style. *Exp. Brain Res.* 202, 101–110. doi: 10.1007/s00221-009-2115-8

Repp, B. H., and Knoblich, G. (2004). Perceiving action identity: how pianists recognise their own performances. *Psychol. Sci.* 15, 604–609. doi: 10.1111/j.0956-7976.2004.00727.x

Riesenhuber, M., and Poggio, T. (2000). Models of object recognition. *Nat. Neurosci.* 3 (suppl.), 1199–1204. doi: 10.1038/81479

Sheffert, S. M., Pisoni, D. B., Fellowes, J. M., and Remez, R. E. (2002). Learning to recognise talkers from natural, sinewave, and reversed speech samples. *J. Exp. Psychol. Hum. Percept. Perform.* 28, 1447–1469. doi: 10.1037/0096-1523.28.6.1447

Sloboda, J. A. (2000). Individual differences in music performance. *Trends Cogn. Sci.* 4, 397–403. doi: 10.1016/S1364-6613(00)01531-X

Stamatatos, E., and Widmer, G. (2005). Automatic identification of musical performers with learning ensembles. *Artif. Int.* 165, 37–56. doi: 10.1016/j.artint.2005.01.007

Tang-Martinez, Z. (2001). The mechanisms of kin discrimination and the evolution of kin recognition in vertebrates: a critical re-evaluation. *Behav. Process.* 53, 21–40. doi: 10.1016/S0376-6357(00)00148-0

Thompson, J. C., and Hardee, J. E. (2008). The first time I ever saw your face. *Trends Cogn. Sci.* 12, 283–284. doi: 10.1016/j.tics.2008.05.002

Van Lancker, D., Kreiman, J., and Emmorey, K. (1985). Familiar voice recognition: patterns and parameters. Part I: recognition of backward voices. *J. Phon.* 13, 19–38.

Zar, J. H. (1999). *Biostatistical Analysis*, 4th Edn. Upper Saddle River, NJ: Prentice Hall.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 01 September 2013; accepted: 04 February 2014; published online: 24 February 2014.

Citation: Koren R and Gingras B (2014) Perceiving individuality in harpsichord performance. Front. Psychol. ***5****:141. doi: 10.3389/fpsyg.2014.00141*

This article was submitted to Cognitive Science, a section of the journal Frontiers in Psychology.

HYPOTHESIS AND THEORY ARTICLE
published: 05 November 2013
doi: 10.3389/fpsyg.2013.00825

The Linked Dual Representation model of vocal perception and production

Sean Hutchins*[1]* and Sylvain Moreno[1,2]

[1] *Rotman Research Institute at Baycrest Hospital, Toronto, ON, Canada*
[2] *Department of Psychology, University of Toronto, Toronto, ON, Canada*

Edited by:
Bruno Gingras, University of Vienna, Austria

Reviewed by:
Lena Rachel Quinto, Macquarie University, Australia
Sandra E. Trehub, University of Toronto, Canada

****Correspondence:***
Sean Hutchins, Rotman Research Institute, 3560 Bathurst St., Toronto, ON M6A 2E1, Canada
e-mail: shutchins@research.baycrest.org

The voice is one of the most important media for communication, yet there is a wide range of abilities in both the perception and production of the voice. In this article, we review this range of abilities, focusing on pitch accuracy as a particularly informative case, and look at the factors underlying these abilities. Several classes of models have been posited describing the relationship between vocal perception and production, and we review the evidence for and against each class of model. We look at how the voice is different from other musical instruments and review evidence about both the association and the dissociation between vocal perception and production abilities. Finally, we introduce the Linked Dual Representation (LDR) model, a new approach which can account for the broad patterns in prior findings, including trends in the data which might seem to be countervailing. We discuss how this model interacts with higher-order cognition and examine its predictions about several aspects of vocal perception and production.

Keywords: perception, production, voice, music, pitch, singing, models

INTRODUCTION

One of the most important abilities of humans is the capacity to communicate complex ideas quickly and efficiently. Although there are many ways of communicating with each other, including methods as diverse as body language, signing, and smoke signals, by far the most important medium is the voice. Singing and speech are cultural universals which rely on the voice being physically produced and perceived; these two processes are necessary for communication to occur. Understanding the relationship between vocal perception and production, then, is critical to understanding communication, the nature of the mental processes underlying it, and the most fundamental abilities of humanity.

Singing, even more than speech, has been one of the most profitable places to look for insights into vocal perception and production. On the production side, it involves a similar degree and type of vocal control as speech, and both create a similar type of signal to be perceived by a listener. Furthermore, because of the stylistic communication goals of music, small variations in the produced signal are generally more important than in speech and have thus been the focus of comparatively more research. Since speech and singing both use similar aspects of the vocal signal, the research on perception and production of the voice in a musical context can be informative of how people use their voices in the context of speech. Indeed, many who study this field consider music to have a special relationship with speech processing, due in large part to their overlap and the greater demands of precision of processing in music (see Moreno et al., 2009 or Patel, 2011). This makes singing a particularly interesting and fruitful place to understand the connection (or lack thereof) between perception and production. Furthermore, these findings may shed some insight on how other domains divide processing for these functions.

Three basic model architectures have been proposed to explain the relationship between vocal perception and production (**Figure 1**). The simplest such theory posits that perception necessarily precedes vocal production (**Figure 1**, left). Thus, when we imitate speech or music, we first construct a symbolic representation of the vocal stimulus. This symbolic representation is then used to construct the vocal-motor representation. These vocal-motor representations are used to issue the appropriate commands to the vocal tract to create the intended sounds. That is, we imitate our symbolic representation of the sound. This model has the benefit of being intuitive and straightforward. It predicts a causal connection between perception and production abilities such that a deficit in our conscious pitch perception abilities would impair our pitch production abilities, while pitch production impairments would not negatively affect our pitch perception abilities.

However, there are alternate models. A motor model of vocal perception (**Figure 1**, center) would predict the opposite processing stream, where vocal stimuli are first processed for their motor-relevant features, and only afterwards are relayed into our conscious perception for symbolic representation. Such a model preserves the correlation between perception and production, but makes the reverse predictions of the naïve model: vocal production impairments should negatively affect vocal perception abilities, but not vice-versa. Finally, dual-route models (**Figure 1**, right) predict that vocal stimuli are processed for motor-relevant features and conscious, symbolic representations along two different, independent pathways. This model predicts that vocal perception and production abilities should be uncorrelated, and each can be improved or impaired without affecting the other. These models all have analogues in the speech domain. To take just a few examples, the general auditory account (Diehl et al., 2004), the motor theory of speech processing (Liberman and

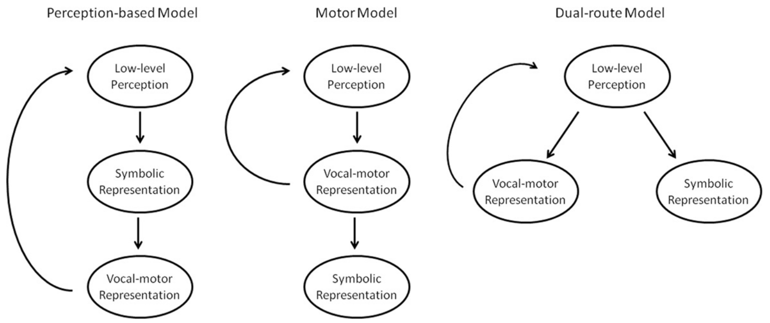


FIGURE 1 | Three proposed models of perception and production.

Mattingly, 1985), and the dual-stream model of speech (Hickok and Poeppel, 2007) mirror the general architectures of the models in **Figure 1** from left to right, respectively.

In this review, we will be examining the many factors that affect perception and production abilities, with an eye toward how perception and production might relate to each other and the neural mechanisms underlying each type of ability. We will look at the evidence for each basic type of model and show how different types of evidence point toward structurally different models. Based on this evidence, we introduce the Linked Dual Representation (LDR) model, a synthesis of the relevant features of these prior models that has the potential to explain why vocal perception and production can appear to be both correlated and dissociable abilities. Finally, we will look at the implications and predictions specific to the LDR model and lay out some possible lines of research.

PRODUCTION OF THE SINGING VOICE

Anybody who has ever been serenaded by "Happy Birthday" could tell you that there can be quite large individual differences in singing ability. Even among people who have never received any formal music training, we can find both potential future stars and those who cannot seem to find the key. One of the major reasons for individual differences in singing is the fact that singers have such a large number of variables to control simultaneously. To be a good singer, one needs to control the pitch, timbre, timing, and loudness of the voice, with many of these factors changing both between and within individual tones. Of course, part of what makes singing good or bad is culturally-dependent. For example, a Western operatic voice is inappropriate for a Hindustani raga, and vice-versa. Within cultures, too, there are stylistic factors that will affect the judgment of performances-a very skilled country-western singer may sound quite out of place in an R&B recording. Taking stylistic concerns into account, we can identify certain factors that contribute to a good singing performance within particular styles. For example, one of the more well-known and studied of these is the singer's formant. This feature, which is really a compression of the 4th and 5th formants (those regions of the frequency spectrum at which the voice is most resonant; these help define the timbre of the voice) into one large amplitude formant, is a marker of good singing in the Western operatic style (Sundberg, 1987) and is typically achieved by lowering the larynx. Producing a singer's formant can help a solo singer to be heard over an orchestra by concentrating amplitude at frequencies which are not as loud in an orchestra (Sundberg, 1987). Studies of the particular characteristics that make a good vocal style for musical theatre (i.e., belting; Sundberg et al., 1993; Cleveland et al., 2003), country music (i.e., "twang"; Sundberg and Thalén, 2010), and others (Borch and Sundberg, 2011) have also revealed unique techniques for those styles. On the other side of the spectrum, studies of poor singers have found a number of acoustical markers that differentiate them from good singers. These include jitter (which captures irregularity in the microstructure of pitch), shimmer (which captures irregularity in the microstructure of amplitude), and harmonic-to-noise ratio (which captures the strength of harmonic vs. inharmonic frequencies), among others (Titze, 2000; Sataloff, 2005).

However, across all singing styles, one of the most important factors in determining the quality of singing is pitch accuracy. For example, in a study assessing the views of music educators on the singing abilities of non-musicians, intonation (pitch accuracy) was rated as the single most important factor in whether or not a non-musician was perceived as having talent (Watts et al., 2003a). Because of its importance, pitch accuracy is also one of the most widely studied factors in the literature on singing ability (e.g., Dalla Bella et al., 2007; Pfordresher and Brown, 2007; Hutchins and Peretz, 2012a). For example, in a study of untrained singers asked to sing a well-known song in either a city park or a lab setting, Dalla Bella et al. (2007) found a range of singing abilities. These singers showed a great amount of variance in the number of pitch interval errors. All of the participants in the park setting had at least one pitch interval error of greater than a semitone, and a few sang incorrectly on over half of the intervals of the song (there were a total of 31 intervals in the song). Singers performing the same song in a laboratory setting had fewer errors, but nevertheless showed a great deal of variability in performance. Interestingly, the number of errors in the time dimension was much lower across all participants in both groups, indicating that timing accuracy does not seem to be as indicative of singing ability as pitch accuracy.

In another study of note, Pfordresher and Brown (2007) studied singers performing single pitches, single intervals, and short melodies. This study also found a range of abilities on each task, with most being able to sing with an average pitch within one semitone of a target pitch, but some being very inaccurate, as high as 250 cents in error (1 semitone = 100 cents). Their results also indicated that poor pitch singers tend to be inaccurate both in single tones and in intervals and melodies. Poor singers tended to compress intervals. A further investigation (Pfordresher et al., 2010) demonstrated the variability of both single tone and interval tuning, even within individual singers. Here, over 50% of participants showed a standard deviation of greater than 100 cents in their singing, indicating wide-spread imprecision and considerable variability both within and between singers. Numerous other studies have looked at pitch-related singing abilities in the population; these have found consistent variation within non-musicians and consistently better pitch abilities in musicians than non-musicians (e.g., Amir et al., 2003; Watts et al., 2003b; Demorest and Clements, 2007; Nikjeh et al., 2009; Hutchins and Peretz, 2012a). Pitch matching ability also tends to increase in children during their elementary and middle school years (Green, 1990; Yarbrough et al., 1991). Thus, it seems that there is a wide range of abilities in the general population to produce vocal pitches accurately. This wide range of abilities, in combination with the importance of pitch matching in singing, makes it one of the best ways to study vocal-motor control, providing an insight into the accuracy of individuals' vocal-motor representations.

FACTORS AFFECTING SINGING ABILITY

One of the most common assumptions about singing is that poor perception ability drives poor production ability. If people cannot hear pitches accurately, then it stands to reason that they will be inaccurate at imitating those pitches. This is the prediction of the perception-based model (**Figure 1**, left). Several studies have investigated this hypothesis, and the evidence is mixed. Using a variety of different singing and pitch perception tasks, some studies have found evidence of a correlation between the two abilities (e.g., Amir et al., 2003; Watts et al., 2005; Moore et al., 2007; Estis et al., 2009, 2011). However, many others, using similar designs, have failed to find a significant correlation (e.g., Bradshaw and McHenry, 2005; Dalla Bella et al., 2007; Pfordresher and Brown, 2007; Moore et al., 2008), which argues more for a dual-route model of perception and production (**Figure 1**, right), making the overall evidence mixed at best.

Two studies addressing this issue are worth pointing out in particular. First, in one of the few studies to use an experimental design, Zarate et al. (2010a) trained participants to better perceive small variations in pitch in the context of micromelodies. However, although they improved at perception, they did not improve in their abilities to produce these same small pitch changes. They concluded that perceptual training does not aid singing ability, thus contradicting the perceptual-based model. Second, in their 2007 study, Pfordresher and Brown found no correlation between pitch perception abilities and their imitation tasks, nor any problems with vocal pitch range in their sample. Thus, they posited that sensori-motor mismappings were the best remaining explanation for poor singing ability in most cases, such that perceived tones were incorrectly mapped onto motor outputs.

In order to sort out the causes of poor singing ability, Hutchins and Peretz (2012a) used a novel methodology involving a new instrument called a slider. This slider produced a synthesized vocal tone that was subject to many of the same limitations as the human voice, including a very fine scale of pitch control. Instead of using their vocal apparatus, though, the participant played the slider by pressing a finger onto a touch-sensitive strip. Thus, it provided a measurement of pitch matching ability independent of the ability to control one's vocal musculature. Pitch-matching ability on the slider was compared to the ability to vocally match a synthesized vocal tone and a prior recording of one's own voice. Participants who could match the pitch with the slider but not with their voice were thus likely to have a vocal-motor control impairment as their primary cause of singing inaccuracies. Those who could match the pitch with the slider and match the recording of their own voice (which had the same timbre as their attempts to match it), but not the synthesized vocal tone, were likely to have a sensori-motor impairment as their primary cause of singing inaccuracies. These singers had a specific difficulty in translating between the timbre of the synthesized voice and the timbre of their own voice. Because their primary deficit was neither in perceiving the relationships among tones, nor in controlling their vocal muscles, but in connecting their perception to an appropriate production, this is considered to be a type of sensori-motor impairment. Finally, those singers who failed at matching pitch both with the slider and the voice are likely to have a perceptual deficit.

The results showed about 20% of singers had a vocal-motor control impairment, 35% had a sensori-motor (timbre) deficit, and only 5% had a perceptual deficit. Participants were universally better at matching pitch with the slider than with their voice, and the results showed a wide range of singing abilities among non-musicians. Singing ability was not aided by multiple attempts, nor was it improved by a visualization of their produced pitch. Although these results show that perception is not a limiting factor in most people's pitch imitation ability, there was nevertheless a modest correlation among non-musicians ($r = 0.4$) between accuracy on the slider and with their voice. These results point to a strong effect of motor and sensori-motor factors on singing ability, with a moderate influence of perceptual ability. This pattern of results suggests aspects of both the perceptual-based model and the dual-route model of vocal perception and production.

Other studies have also shown effects of the target's timbre on pitch-matching ability. Singers are better able to match the pitch of vocal targets with a similar voice than the pitch of instruments (Watts and Hall, 2008) and better able to match the pitch of their own voice than the pitch of other targets (Moore et al., 2008). Poor singers are especially aided by using a human, rather than synthetic, target pitch (Léveque et al., 2012). Educators also report that children tend to be able to match pitch better when modeling a similar voice (reviewed in Goetze et al., 1990).

A number of functional imaging studies have investigated the brain areas that support singing production. These studies have localized the "singing network," which includes the auditory cortex, insula, supplementary motor area and anterior cingulated, as well as parts of the motor cortex specific to the mouth/lips and larynx. (Perry et al., 1999; Brown et al., 2004; Özdemir et al., 2006; Kleber et al., 2007). This network is involved in motor production, motor planning of sequences, motor initiation, and articulation.

Singing ability is also reflected in neural activation patterns. For example, as might be expected, highly trained singers show more recruitment of laryngeal and mouth areas of the somatosensory cortex than less-trained singers, an effect related to the amount of singing practice (Kleber et al., 2010). They also show more activation in non-cortical regions, such as the basal ganglia, the thalamus, and the cerebellum (Kleber et al., 2010). Other studies using a pitch-shift paradigm, in which the singer's auditory feedback is manipulated while producing the tones, have shown that experienced singers recruit more areas of the singing network than untrained singers (Zarate and Zatorre, 2008). This methodology has shown a particularly strong role of the dorsal premotor cortex in regulating and controlling responses to auditory feedback; this area is thus thought to be highly involved in the interface between perception and production (Zarate and Zatorre, 2008; Zarate et al., 2010b).

PERCEPTION OF THE SUNG VOICE

GENERAL PITCH PERCEPTION ABILITIES

While there has been a good amount of research on singing ability and the factors underlying singing ability, there has been quite a bit less research done of vocal perception. However, we know a great deal about auditory perception in general. In the case of pitch, we can measure just-noticeable differences (or difference limens); in some cases these can be as low as five cents (Zwicker and Fastl, 1999). Individual differences in pitch difference limens, which can be considerable, could contribute to differences in vocal pitch perception abilities. The timbre of tones can also affect pitch perception abilities. Changes in timbre interfere with pitch judgments (Melara and Marks, 1990a,b,c; Krumhansl and Iverson, 1992), and timbre and pitch have been shown not to be perceptually independent (Melara and Marks, 1990a,b,c; Krumhansl and Iverson, 1992; Pitt, 1994; Warrier and Zatorre, 2002). Musicians seem to be less susceptible to timbral interference of pitch processing, however, (Beal, 1985; Pitt and Crowder, 1992; Pitt, 1994).

There is also considerable variability in preferences and judgments of musical intervals. Listeners will show differences between what they consider to be an acceptably-tuned musical interval or note (Rakowski, 1990; Vurma and Ross, 2006; Hutchins et al., 2012), as well as differences in their identification judgments of intervals (Siegel and Siegel, 1977; Halpern and Zatorre, 1979). There are also individual differences related to musical training in preferences in listening to certain types of consonant vs. dissonant intervals (McDermott et al., 2010).

Experience and training can play a large role in pitch perception ability, as evidenced by the differences between musicians and non-musicians (e.g., Pitt, 1994; Moreno and Besson, 2006; Moreno et al., 2009; McDermott et al., 2010; Hutchins et al., 2012). Even among non-musicians, pitch discrimination abilities can be improved with extra training (Zarate et al., 2010a). Tone-language speakers, too, show better pitch perception abilities, presumably due to their greater experience in pitch processing (Pfordresher and Brown, 2009; Bidelman et al., 2013a). Among bilinguals, there is also evidence of causality running in the opposite direction, such that musical ability is predictive of the ability to discriminate and produce non-native speech sounds, both for linguistic tones (Gottfried et al., 2004; Alexander et al., 2005) and for non-tone phonemes (Slevc and Miyake, 2006). Musically trained participants are also better at detecting pitch changes in speech in a foreign language (Marques et al., 2007).

One of the most important neurological correlates of pitch processing ability is the auditory brainstem response (ABR). This response mimics the pitch and some timbral characteristics of a presented tone (Krishnan, 2007; Skoe and Kraus, 2010) and occurs very early in processing, being recorded typically with less than a 10 ms lag following the stimulus. One characteristic of the ABR that is of particular interest is the fact that trained musicians show a higher-fidelity ABR with a shorter lag than non-musicians; this higher fidelity ABR correlates with better ability to make behavioral pitch judgments (Kraus et al., 2009; Bidelman et al., 2011). This benefit is not limited to musicians but generalizes to other groups with high expertise in pitch, such as tonal language speakers (Krishnan et al., 2008; Bidelman et al., 2013b). Other studies have shown that the ABR preserves timbral characteristics more accurately in people with musical backgrounds (Kraus et al., 2009; Bidelman and Krishnan, 2010; Strait et al., 2012). This early benefit in pitch and timbre perception seems to precede cortical representations of pitch and timbre and may be transformed to a more conceptual-level representation of the response as it is transmitted upwards (Bidelman et al., 2013a). This response most likely occurs before any task-relevant effects have time to affect the neural representation. Thus, the fidelity of the brainstem response is a good candidate to affect the accuracy of both pitch perception and production, and may be an indicator of the earliest level of perceptual processing.

CONGENITAL AMUSIA

One way of learning about the causes and effects of pitch perception, as well as its relationship to production and to the domain of language, is by looking at cases where pitch perception is compromised. Congenital amusia, which is a neurogenetic disorder (Peretz et al., 2007) characterized by impaired music perception ability in the absence of brain damage or hearing or cognitive impairments (Peretz, 2008), provides this kind of test case. This condition is formally diagnosed by the Montreal Battery of Evaluation of Amusia (MBEA; Peretz et al., 2003). The majority of congenital amusics seem to suffer from a selective pitch perception deficit. Amusics are impaired at detecting pitch changes of less than a semitone (Peretz et al., 2002; Hyde and Peretz, 2004) and distinguishing between rising and falling pitches (Foxton et al., 2004; Liu et al., 2010). Amusics also seem to be somewhat impaired in timbre perception (Tillmann et al., 2009; Marin et al., 2012) and memory for pitch (e.g., Gosselin et al., 2009; Tillmann et al., 2009; Williamson et al., 2010). Their condition often leads to amusics not enjoying or seeking out music. Subjectively, they report that music seems like noise; thus it is reasonable to suspect a vicious circle here, where amusics tend to listen to music less often, thus gaining less experience with processing it, making listening even less rewarding than it otherwise might have been.

As would be expected from this type of condition, amusics are impaired in their singing abilities as well. Congenital amusics are judged as poor singers (Ayotte et al., 2002) and make considerably more pitch errors in singing a well-known song than do matched controls (Dalla Bella et al., 2009; Tremblay-Champoux et al., 2010). They are also well-below controls at matching single pitches (Hutchins et al., 2010). However, there are some signs that amusics are not uniformly poor at singing. Certain amusics seem to sing considerably better than would be predicted by their poor perceptual abilities (Dalla Bella et al., 2009; Hutchins et al., 2010; Tremblay-Champoux et al., 2010), and amusics as a whole are aided when directly imitating a model, rather than singing from memory (Tremblay-Champoux et al., 2010). For example, one amusic, ML, is able to sing an array of songs just as well as or better than unimpaired individuals despite her inability to hear errors in songs. These types of findings suggest that conscious perceptual ability may not be a hard limit on amusics' singing abilities. Further evidence for this and its implications will be reviewed later in this paper.

Anatomic and functional MRI studies have shown several differences between congenital amusics and unimpaired individuals. Congenital amusics typically show reduced white matter in the right inferior frontal gyrus, as well as thicker cortices in both that area and the right auditory cortex (Hyde et al., 2007). There is some evidence that there may be differences between amusics and controls in the left analogues of those regions as well (Mandell et al., 2007). In the right hemisphere, these two regions also show reduced functional connectivity (Hyde et al., 2011), and diffusion tensor imaging has shown reduced anatomical connectivity in the right arcuate fasciculus connecting these two regions (Loui et al., 2009). There is some evidence that different regions of the arcuate fasciculus may correlate with pitch perception ability and the discrepancy between perception and production ability (Loui et al., 2009), but this has yet to be corroborated.

Electrophysiological evidence also supports the relationship between pitch perception abilities and frontal-auditory connectivity. Amusics show a normal mismatch negativity (MMN) response (a pre-conscious response to deviations in sound generated in the auditory cortex, Näätänen et al., 2007) to small deviations in pitch which they are unable to consciously detect (Moreau et al., 2009; Peretz et al., 2009). These same deviations, however, generate no P3b response, normally indicative of attentive processing (Moreau et al., 2013). These components, then, seem to be markers of conscious and unconscious pitch perception ability. Taken together, the evidence indicates that frontal regions, auditory regions, and the connection between them regulate normal pitch perception ability, and that there may be anatomically and functionally distinct regions responsible for conscious and unconscious pitch processing. While the regions and processes investigated in these studies are not voice-specific, this type of pitch processing is likely a precursor to voice specific perception and production abilities, which may also be anatomically and functionally distinct.

IS VOCAL PITCH PERCEPTION SPECIAL?

One possible explanation of amusics' better-than-expected singing abilities is that our ability to perceive vocal pitch (and by extension, the processes underlying this ability) may be different from our ability to perceive the pitch of non-vocal tones, such as instruments or synthesized tones. While it is obvious that we can *distinguish* between the voice and other instruments, not many studies have examined the uniqueness of vocal musical perception. One clue that there may be fundamental differences between vocal and non-vocal pitch perception comes from the tuning perception literature. It has been noticed that pitch errors seem to be less noticeable when produced by a voice than by other instruments (Seashore, 1938; Sundberg, 1979). For example, Lindgren and Sundberg (as cited in Sundberg, 1979, 1982) showed that musically experienced listeners would accept as in-tune up to 50–70 cents of tuning errors in a recording of a highly trained singer. Another study looked at recordings of 10 professional singers performing the same song, and found that listeners were highly variable in their assessments of the tuning, with out-of-tune notes being accepted as in-tune and well-tuned notes sometimes being judged as out-of-tune (Sundberg et al., 1996). In contrast, studies of acceptable tuning in synthesized tones show a much smaller range of acceptable tuning, with listeners accepting only 10–15 cents of error (Fyk- in van Besouw et al., 2008). This seems to indicate that listeners use different criteria when judging the pitch of the voice vs. other instruments.

To investigate this effect in a well-controlled manner, Hutchins and Peretz (2012a) directly compared tuning judgments of real and synthesized voices. Musicians and non-musicians listened to pairs of tones and judged them as the same or different. Listeners were less likely to notice the differences in tuning when the tone pairs were real voices than when they were synthesized voices; this pattern held across musicians and non-musicians. Non-musicians needed the two tones to be 50 cents apart to reliably notice the difference between two real vocal tones, compared with only 30 cents for synthesized vocal tones. This pattern held in musicians as well. Hutchins et al. (2012) found very similar results for tuning judgments of a trained voice vs. a violin and extended these findings to a melodic context. This difference in acceptable and noticeable tuning between voices and other timbres was termed the Vocal Generosity Effect and may be evidence of special processing of voices in a musical context as it is consistent across different voices and instruments.

Different types of tuning errors between vocal and non-vocal stimuli are also found in production. Trained singers tend to show more tuning errors than trained instrumentalists. Trained singers have a propensity to begin a note flat (Seashore, 1938), and analyses of recordings of professional singers show deviations of more than 40 cents, both sharp and flat (Prame, 1997). In contrast, studies of violin and wind instruments show average deviations less than 20 cents. This difference in production ability comes despite the fact that people have considerable amounts of experience using their voice. In experts, though, there is a tendency for instrumentalists to practice much more than vocalists (as the voice tends to tire out after a couple of hours of practice). In addition, singers typically use considerably more vibrato than do performers on other instruments, such as the violin (Prame, 1997; Mellody and Wakefield, 2000). Vibrato is sometimes thought to be a way of hiding tuning errors (Yoo et al., 1998), although listeners are nevertheless capable of making quite accurate tuning judgments even for tones with very high-amplitude vibrato (Shonle and Horan, 1980). However, unlike the case of perception, many of these differences between voice and instruments can be explained by the unique motoric requirements of vocal production, which are substantially different from those required by any other instrument.

If the voice *is* processed differently from other instruments, then we should see special neural processes and regions devoted to vocal perception and production. And indeed, there is evidence for just such effects. Belin et al. (2000) showed evidence for subregions of the auditory cortex particularly sensitive to voice perception, called temporal voice areas. These are located bilaterally along the mid superior temporal sulcus, and respond to the voice independent of its linguistic content. Temporal voice areas become less active as the vocal signal is degraded by filtering, indicating a sensitivity to the quality of the input that was reflected in both fMRI and behavioral voice discrimination judgments. Electrophysiological studies also indicate special processing of the voice, with vocal sounds eliciting a fronto-temporal positivity/occipital negativity when compared to environmental sounds or birdsong, peaking around 200 ms post-stimulus (Charest et al., 2009). Another study found a similar frontal positivity of sung tones compared to instrumental sounds, but a bit later, likely due to the more similar acoustic characteristics of these stimuli (Levy et al., 2001), although an MEG study failed to show any differences between similar types of stimuli (Gunji et al., 2003). To the best of our knowledge, no one has yet run an fMRI study comparing activation from perceiving humming to that of perceiving instruments to look for vocal-specific regions involved in music processing. Given the specificity of the motor demands of singing, we would expect to find some such regions; such an experiment would provide an important contribution to the field.

THE RELATIONSHIP BETWEEN PERCEPTION AND PRODUCTION

To truly understand the nature of perception and production abilities, it is helpful to examine their relationship to each other, specifically the link between conscious vocal perception acuity and vocal production accuracy. The evidence reviewed so far shows a moderate, but not overwhelming correlation between perception and production abilities, which suggests a connection, rather than dissociation, between the two. This points more toward a perceptual-based or motor model of perception and production, rather than a dual route model (see **Figure 1**). However, other lines of evidence tend to argue against the simple and motor models, and dual-route models have been suggested to explain this pattern of findings (Griffiths, 2008).

PERCEPTION-PRODUCTION DISSOCIATIONS IN CONGENITAL AMUSIA

Some of the best evidence arguing for a dual-route model of perception and production comes from congenital amusics. Although most congenital amusics, who have severely impaired pitch perception abilities, are impaired in their singing ability, there is evidence that some amusics nevertheless retain the ability to sing accurately. Dalla Bella et al. (2009) identified three amusics (out of eleven tested) who were unimpaired at singing the correct intervals in a well-known song, including one who was unimpaired even without the aid of the lyrics—a condition in which most amusics fail to complete more than a few notes of the song. Hutchins et al. (2010) tested congenital amusics in a single-pitch matching task and found that despite amusics' overall inaccurate performances, they showed a consistent, linear relationship between the imitations and the target tones.

These studies hint that amusics may demonstrate better overall singing ability than would be predicted from their abilities on perceptual tasks. Recently, a number of studies have attempted to directly compare perception and production abilities in amusia, to serve as direct tests of vocal perception and production models. Loui et al. (2008) presented three amusics with two note sequences and asked amusics to imitate the interval, then to describe whether the second note had been higher or lower than the first. The amusics were impaired at describing the direction of the second note, but they performed similarly to controls at singing an interval that went in the correct direction, although they were still inaccurate at producing an interval of the correct distance.

Some of our recent work also demonstrates a similar discrepancy between pitch perception and production ability in amusics. In one ongoing study (Hutchins and Peretz, 2010), we tested amusics' pitch matching abilities with the slider and a vocal imitation condition (the same as used in Hutchins and Peretz, 2012a, Experiment 1; see above). As expected, amusics as a group performed worse than matched controls at both slider and vocal pitch matching. However, we found two participants who performed at levels comparable to normal participants on the vocal imitation task and, notably, better than their performance on the slider. This is a pattern of results not found among normal participants, who almost invariably show excellent pitch matching performance on the slider, even among non-musicians. This demonstrates that for these two amusics, their vocal pitch matching ability was not constrained by their pitch perception ability, arguing against the perceptual-based model of pitch perception and production.

Another of our studies looked at the pitch shift effect. This effect is an automatic compensatory response to a sudden shift in pitch of the feedback of a sung or spoken utterance. When most participants hear such a shift in their own voice, there is a quick reaction to change the pitch of their voice in the opposite direction. We tested amusics and controls in a pitch shift paradigm, where a pitch shift would occur in the middle of an imitative response. Our results showed that a subset of amusics showed a preserved pitch shift effect, showing normal pitch shift responses to both large (2 semitone) and small (25 cent) shifts. This is strong evidence that amusics do process even small pitch shifts when they are relevant to vocal-motor control. In addition, this study also found evidence of a correlation between the pitch shift effect and pitch matching accuracy (absent of any shift), strengthening the idea that this retained pitch shift response is related to generally preserved vocal-motor control. Together, this presents a strong contrast with amusics' previously documented disabilities in consciously perceiving small pitch changes.

We also see evidence for dissociation of vocal perception and production abilities in amusics' use of pitch in speech. Unlike in tone languages, pitch is non-lexical in most European languages. However, it plays a strong role in prosody and can determine the meaning of certain types of statement/question pairs. Liu et al. (2010) showed that amusics were somewhat poorer than controls at discriminating between statements and questions differing only in pitch contour. However, just as with intervals (Loui et al., 2008), they were better at imitating the pitch contour of these same sentences (although still below the level of matched controls). Hutchins and Peretz (2012b) tested amusics with speech examples containing pitch changes that did not systematically alter the meaning of the sentence. In this experiment, amusics showed an impaired ability to perceive pitch changes between sentences, but no impairment at imitating those same pitch differences, compared to controls. Similarly, in the pitch shift study (Hutchins and Peretz, 2013), we found no difference between pitch shift responses to spoken vs. sung utterances. The fact that pitch perception-production dissociation occurs across music and speech indicates that it is a function of vocal pitch perception and control, rather than a function of music.

Neural evidence also supports the dissociation between pitch perception and production in amusics. Loui et al. (2009) found that pitch perception abilities were correlated with tract density along the superior route of the arcuate fasciculus, whereas the lower route was correlated with the difference between their perception and production abilities. While a somewhat complicated story (all the more so because the association runs in the reverse direction to some other theories of dual-route processing, e.g., Goodale and Milner, 1992; Hickok and Poeppel, 2004), this is the first evidence of direct correlations between these dissociations in amusics and specific neuroanatomical structures.

EVIDENCE FOR PERCEPTION-PRODUCTION DISSOCIATIONS IN NORMAL SUBJECTS

A few studies have shown similar evidence for dissociations between perception and production abilities in an unimpaired population. In one study, Hafke (2008) used a vocal pitch shift paradigm to test trained singers. She found that they showed a normal pitch shift effect, even when the shifts were so small that the participants were unaware that they had occurred at all. This is similar to the pattern of results found among congenital amusics (Hutchins and Peretz, 2013). Vurma (2010) showed a related effect, demonstrating that trained singers' musical interval production abilities are more finely honed than their abilities to perceive the same intervals. Results such as these indicate that the independence of vocal-motor pitch control from conscious pitch perception is not limited to cases such as amusia, which again argues against a perceptual-based model.

The reverse pattern, better conscious perception than production ability, is even more common in normal participants. Hutchins and Peretz (2012a) showed that almost every participant was more capable of matching pitch with an instrument than with their voice in many cases over an order of magnitude better. This pattern held true for musicians and non-musicians alike and demonstrated that poor vocal pitch accuracy does not lead to poor pitch perception ability, as would be predicted by a motor theory. However, there was a moderate correlation between instrumental and vocal pitch matching abilities, arguing against a dual-route theory. A few other studies have found evidence of such perception-production connections (e.g., Amir et al., 2003; Watts et al., 2005; Moore et al., 2007; Estis et al., 2009, 2011), though others have failed to do so (Bradshaw and McHenry, 2005; Dalla Bella et al., 2007; Pfordresher and Brown, 2007; Moore et al., 2008). The preponderance of evidence shows a weak connection between pitch perception and singing ability, but also indicates that poor pitch perception ability is not necessarily the main cause of poor singing ability.

Similar evidence of this dissociation comes from second language learners. Many late second language learners will gain the ability to comprehend a second language, but will nevertheless be unable to speak it with any degree of fluency. Other second language learners, however, will show an opposite pattern, where their production ability will outstrip their comprehension ability. This latter pattern is typically shown by people who need to perform or deliver information in a second language, such as the singer who performs a Mozart opera without speaking a word of German, whereas the former is more characteristic of an immigrant immersed in a second language who does not have the opportunity or inclination to speak it often. Again, like with pitch in singing, perception and production ability in a second language will broadly correlate, but are nevertheless dissociable abilities.

THE LINKED DUAL REPRESENTATION MODEL

Across these studies, we see two main patterns emerging. First, there is a trend for people who are poor at pitch perception to be worse singers, holding across amusics and unimpaired people. This correlation is not perfect, however, and perception does not determine pitch matching abilities. Second, in many cases, people's production abilities can outstrip their perceptual limitations (or vice versa); this pattern can arise in both perceptually impaired and unimpaired people. To account for these two main patterns we propose a new model of adult human vocal perception and production: The LDR model (**Figure 2**). Like a dual-route model, the LDR model predicts that vocal information can be processed in two distinct ways. First, it can be encoded as a symbolic representation, such that we gain conscious knowledge of the identifiable features of the vocal stimulus. This process, which is what we normally equate with conscious perception, allows us to determine whether a tone is higher or lower than another, the same or different from another, and allows us to make identification and categorization judgments. Second, vocal information can be encoded as a motoric representation, such that it enables reproduction, imitation, or generative production.

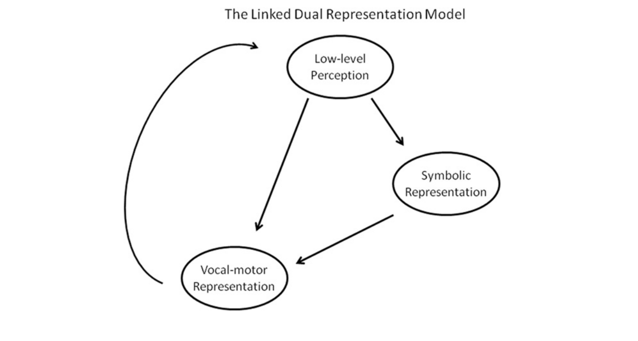

The Linked Dual Representation Model
Low-level Perception
Symbolic Representation
Vocal-motor Representation

FIGURE 2 | The Linked Dual Representation model.

The LDR model predicts that vocal information can be directly encoded as a motoric representation, without mediation through a symbolic representation. Just as a point in space can be represented with Cartesian or polar coordinates, each of which is better suited to particular calculations, these symbolic and motor representations support different kinds of behaviors.

However, unlike other dual-route models, the LDR model also predicts that the vocal-motor representation can be mediated by the symbolic representation (see **Figure 2**). Whereas most dual-route model fail to predict the broad correlation seen between vocal perception and production abilities (e.g., Goodale and Milner, 1992; Griffiths, 2008; Hutchins et al., 2010), this aspect of the model is designed to incorporate this effect. The LDR model predicts that a vocal-motor representation is influenced directly by the low-level perceptual information, but also indirectly by our conscious perception, identification, and category judgments of the information. This is a unidirectional link between the symbolic and vocal-motor representations; the latter cannot directly affect the former. Finally, there is a process of feedback from production back to low-level perception; this process is taken to reflect both auditory feedback from actual productions as well as efferent feedback from actualized motor plans.

All of these processes are variable in strength and are influenced by top-down mechanisms, similar to the way in which executive function can moderate transfer effects between speech and music (Moreno and Bidelman, 2013). The relative influence of the symbolic and direct motoric encoding of a tone on its production can be mediated by the task requirements and context. Even the degree to which a tone is initially encoded symbolically or motorically is influenced by the intention of the listener. A listener who is tasked with comparing a note to a template or identifying an interval will preferentially encode it symbolically, whereas the same input would lead to a stronger vocal-motor encoding in the context of an imitation task. These effects can be visualized as a change in the relative sizes of the arrows.

This model, although motivated by pitch, is intended to apply to other aspects of vocal processing, including timbre, loudness, and phonemic processing. There is nothing about symbolic representation or motoric encoding which does not apply equally to other aspects of vocal tones. This generalization is motivated by several factors, including amusics' impairment in speech perception but not production (Hutchins and Peretz, 2012b), and variability in speech perception and production abilities among normal participants in contexts such as second language learning. However, the applicability of this model to speech warrants further study. The model assumes that initial perception of these attributes can vary across individuals; this variance is passed along to subsequent steps and can influence the accuracy of both types of encoding. It also assumes that individuals can vary in skill in transforming between these different representations accurately, independently of their initial perceptual abilities. Together, these variances in different abilities can explain the patterns of individual difference in perception, discrimination, and imitation abilities.

Taken together, this model provides a more complete explanation of the data than previously proposed models by combining some of the features of previous models. For example, similar to other dual-route models that have been proposed, the LDR model is able to predict dissociations between perception and production among congenital amusics. This model posits that congenital amusics are impaired at encoding pitch symbolically and are thus poor at tasks such as categorization or identification of pitch. Because symbolic representations are responsible for our awareness of pitch, congenital amusics also have diminished awareness of pitch, leading to their lower enjoyment of music. However, they retain their ability to encode pitch as a vocal-motor code. Thus, in some cases, they retain their ability to imitate pitches and respond to pitch changes, often just as well as normal participants. However, they are still, on average, below the abilities of normal participants, which is due to the lack of contribution from a symbolic representation of pitch. A similar argument using naturally occurring variances in abilities can also explain why normal individuals will occasionally show a similar dissociation between conscious perception and production abilities.

However, straightforward dual-route models are unable to explain cases where there seems to be a relationship between perception and production. In contrast, the influence of the symbolic representation on the vocal-motor encoding in the LDR model allows it to explain the moderate correlation between pitch perception ability and imitation ability. Furthermore, this route of influence also allows us to explain the broad correspondence between what we produce and what we hear- most people's imitative responses broadly line up with their perceptual judgments (although not a one to one correspondence). This processing flow, and the independent variance in these abilities, can explain why individual differences in perception and production abilities co-vary but are not perfectly predictive.

FUTURE DIRECTIONS

The LDR model makes several predictions, which would be profitable to explore in future research. First, because this model is assumed to apply to all vocal abilities, rather than specifically to the domain of music or speech, this model predicts that vocal perception and production abilities should be domain-independent. We would expect to find that, in general, people who are better at singing should be better at using their voice for speaking and vice-versa. It has already been shown that congenital amusics are unimpaired at speech imitation (Hutchins and Peretz, 2012b), and they typically report no general speech production problems. The LDR model predicts that this general phenomenon should carry over to an unimpaired population as well. For example, trained singers should be better at speech imitation, and people skilled at manipulating their voices (such as voice actors) should be better than average at singing. This leads to the interesting prediction that training in singing should also help public speaking ability (above and beyond the benefit of simply becoming more comfortable performing in front of others). Similar relationships should also be found between experts in speech and music perception (such as speech therapists or piano tuners). However, the model also predicts that these abilities are task-dependent—better singers are not necessarily better at perceiving speech sounds. Showing such a pattern would help confirm the domain-generality of this model.

A particularly interesting aspect of this prediction arises when considering the case of dyslexia, which is fundamentally an impairment in reading and writing skills. Many instances of dyslexia are assumed to arise from an impairment of phonetic abilities (Bradley and Bryant, 1978, 1983; Bruck, 1992), which can be considered to be difficulty forming an adequate motor representation of speech sounds (Heilman et al., 1996; Hickok and Poeppel, 2004; D'Ausilio et al., 2009). The LDR model bears a few similarities to dual-route models of sentence reading, which assume that phonological and whole-word routes are mediated by separate neural pathways (e.g., Coltheart et al.'s Dual Route Cascade model, 1993). Both models explain dyslexics' particular difficulties with reading non-words. However, the LDR model puts the phonological difficulties of dyslexics in the context of a general impairment of vocal-motor encoding. Because of this, we would predict that dyslexics should be worse than non-dyslexics at tasks requiring speech imitation and that they would be particularly influenced by the mediating influence of the symbolic representation of phonemic sounds. Thus, dyslexics should be particularly sensitive to the categorical representations of sounds and less able than non-dyslexics at imitating within-category variations in speech sounds.

Another unique prediction of the LDR model comes from taking the dynamics of the system into account. Although a production response can be constructed directly from the input or mediated by the symbolic encoding of the input, the latter route to motor responses involves more steps and would thus take more time to perform. This explains several interesting facts about the timing of vocal responses. In the pitch shift task, for example, responses occur very rapidly and automatically, typically around 100–200 ms after the pitch shift. However, when asked to consciously control the pitch shift response (by inhibiting it, for example), participants are unable to do so as quickly and take another 200–300 ms to make a conscious adjustment to their automatic shift response (Burnett et al., 1998). Our model posits that the controlled response must come through conscious awareness via a symbolic representation of vocal pitch, whereas the automatic response comes directly from a motor-representation of the feedback, creating the different time courses of the two responses.

A similar effect can be found in speech shadowing. Listeners have the ability to shadow a stream of speech (e.g., Chistovich, 1960; Chistovich et al., 1960; Marslen-Wilson, 1973) with a delay as short as 150 ms. While both close and distant shadowing can be quite accurate, and are subject to the same global effects of context (Marslen-Wilson, 1973, 1985), those who shadow speech quickly typically report that they were repeating the material "before they understood [it]" (Chistovich et al., 1960, see also Marslen-Wilson, 1985), whereas the distant shadowers reported knowing what the words were before repeating them. Marslen-Wilson (1985) described evidence that, in certain cases, distant shadowers were more affected by the meaning of words than close shadowers, a fact that makes sense if close shadowers were using a direct encoding from vocal input to vocal motor code and distant shadowers made use of the slower route through symbolic representation of words in their shadowing. Interestingly, when close shadowers were forced to consider the meaning of the words they were shadowing, their performance became slower, more like that of close shadowers (Marslen-Wilson, 1985), a process which can also be explained by the latency of the two analysis paths. Our model would also make the counterintuitive prediction that variation in the speech sounds, such as in different regional accents, would be more likely to be preserved in close shadowers than distant shadowers, due to the normalization process inherent in creating symbolic representations of the stream of speech.

These dynamical properties of the model could be tested directly using absolute pitch possessors. We would predict that in a vocal matching task, requiring a speeded response would make more use of the direct route to a vocal-motor encoding, bypassing the symbolic representation of pitch. However, forcing a delayed response (past the length of the sensory buffer) would lead to greater mediation of the symbolic representation. Because absolute pitch listeners are able to categorize pitches into distinct pitch classes (Takeuchi and Hulse, 1993; Levitin and Rogers, 2005), we would expect that these listeners would be more influenced by their categorizations when making delayed responses, whereas non-absolute pitch listeners should merely show a general decrease in accuracy over longer timescales (as in Estis et al., 2009).

One final avenue worth considering is the connection between the LDR model and the mirror neuron system. This system, which is hypothesized to underlie our abilities to recognize the connections between our actions and those of others (Rizzolatti et al., 2001; Kohler et al., 2002; Rizzolatti and Craighero, 2004), may be of great importance in the ability to imitate others' actions (Brass and Heyes, 2005; Heyes, 2011) and may play a role in speech processing as well (Rizzolatti and Arbib, 1998; although the importance of mirror neurons is not universally agreed upon, see Hickok, 2009, for example). The LDR model's ability to represent an input as a motor code and a symbolic code may be related to the mirror neuron system's purported ability to mediate between these two codes, and it may well be that dissociations between perceptual and production abilities are more likely to be found in people with poorer mirror neuron systems. As both of these models intend to describe the relationship between perception and imitation tasks, further research into their connection (or lack thereof) could be very revealing.

CONCLUSION

There is a great deal of variability in vocal perception and performance abilities and only a modest correlation between the two. Vocal perception and production are highly related to speech and musical processing, and we see evidence of a relationship in abilities between the two domains. However, despite the link between vocal perception and production abilities, there is growing evidence supporting a dissociation between them, both in impaired and unimpaired individuals. The LDR model can explain both these broad trends in the data and makes several new predictions about speech imitation, singing, and response timing. We believe this model will help to interpret a wide variety of experiments and can create a common framework for understanding vocal perception and production.

ACKNOWLEDGMENTS

This work was supported by a grant from the Federal Economic Development Agency for Southern Ontario. The authors would like to thank Aline Moussard and Stefanie Hutka for their comments on previous versions of this manuscript, and Tristan Watson and Madeline Harris for editing assistance. The authors would also like to thank the contributions of our two reviewers for their comments on earlier drafts of this paper.

REFERENCES

Alexander, J., Wong, P. C. M., and Bradlow, A. (2005). "Lexical tone perception in musicians and nonmusicians, in *Proceedings of Interspeech 2005 - Eurospeech - 9th European Conference on Speech Communication and Technology* (Lisbon), 397–400.

Amir, O., Amir, N., and Kishon-Rabin, L. (2003). The effect of superior auditory skills on vocal accuracy. *J. Acoust. Soc. Am.* 113, 1102–1108. doi: 10.1121/1.1536632

Ayotte, J., Peretz, I., and Hyde, K. (2002). Congenital amusia: a group study of adults afflicted with a music-specific disorder. *Brain* 125, 238–251. doi: 10.1093/brain/awf028

Beal, A. L. (1985). The skill of recognizing musical structures. *Mem. Cogn.* 13, 405–412. doi: 10.3758/BF03198453

Belin, P., Zatorre, R. J., Lafaille, P., Ahad, P., and Pike, B. (2000). Voice-selective areas in human auditory cortex. *Nature* 403, 309–312. doi: 10.1038/35002078

Bidelman, G. M., Hutka, S., and Moreno, S. (2013a). Tone language speakers and musicians share enhanced perceptual and cognitive abilities for musical pitch: evidence for bidirectionality between the domains of language and music. *PLoS ONE* 8:e60676. doi: 10.1371/journal.pone.0060676

Bidelman, G. M., Moreno, S., and Alain, C. (2013b). Tracing the emergence of categorical speech perception in the human auditory system. *Neuroimage* 79, 201–212. doi: 10.1016/j.neuroimage.2013.04.093

Bidelman, G. M., Gandour, J. T., and Krishnan, A. (2011). Cross-domain effects of music and language experience on the representation of pitch in the human auditory brainstem. *J. Cogn. Neurosci.* 23, 425–434. doi: 10.1162/jocn.2009.21362

Bidelman, G. M., and Krishnan, A. (2010). Effects of reverberation on brainstem representation of speech in musicians and non-musicians. *Brain Res.* 1355, 112–125. doi: 10.1016/j.brainres.2010.07.100

Borch, D. Z., and Sundberg, J. (2011). Some phonatory and resonatory characteristics of the rock, pop, soul, and swedish dance band styles of singing. *J. Voice* 25, 532–537. doi: 10.1016/j.jvoice.2010.07.014

Bradley, L., and Bryant, P. (1978). Difficulties in auditory organization as a possible cause of reading backwardness. *Nature* 271, 746–747. doi: 10.1038/271746a0

Bradley, L., and Bryant, P. (1983). Categorizing sounds and learning to read—A causal connection. *Nature* 310, 419–421. doi: 10.1038/301419a0

Bradshaw, E., and McHenry, M. A. (2005). Pitch discrimination and pitch matching abilities of adults who sing inaccurately. *J. Voice* 19, 431–439. doi: 10.1016/j.jvoice.2004.07.010

Brass, M., and Heyes, C. (2005). Imitation: is cognitive neuroscience solving the correspondence problem. *Trends Cogn. Sci.* 9, 489–495. doi: 10.1016/j.tics.2005.08.007

Brown, S., Martinez, M. J., Hodges, D. A., Fox, P. T., and Parsons, L. M. (2004). The song system of the human brain. *Cogn. Brain Res.* 20, 363–375. doi: 10.1016/j.cogbrainres.2004.03.016

Bruck, M. (1992). Persistence of dyslexics' phonological awareness deficits. *Dev. Psychol.* 28, 874–886. doi: 10.1037/0012-1649.28.5.874

Burnett, T. A., Freedland, M. B., Larson, C. R., and Hain, T. C. (1998). Voice F0 responses to manipulations in pitch feedback. *J. Acoust. Soc. Am.* 103, 3153–3161. doi: 10.1121/1.423073

Charest, I., Pernet, C. R., Rousselet, G. A., Quiñones, I., Latinus, M., Fillion-Bilodeau, S., et al. (2009). Electrophysiological evidence for an early processing of human voices. *BMC Neurosci.* 10:127. doi: 10.1186/1471-2202-10-127

Chistovich, L. A. (1960). Classification of rapidly repeated speech sounds. *Akusticheskii Zh.* 6, 392–398.

Chistovich, L. A., Aliakrinskii, V. V., and Abulian, V. A. (1960). Time delays in speech repetition. *Vopr. Psikhol.* 1, 114–119.

Cleveland, T. F., Sundberg, P. J., and Prokop, J. (2003). Aerodynamic and acoustical measures of speech, operatic, and Broadway vocal styles in a professional female singer. *J. Voice* 17, 283–297. doi: 10.1067/S0892-1997(03)00074-2

Coltheart, M., Curtis, B., Atkins, P., and Haller, M. (1993). Models of reading aloud: dual-route and parallel-distributed-processing approaches. *Psychol. Rev.* 100, 589. doi: 10.1037/0033-295X.100.4.589

Dalla Bella, S., Giguère, J. F., and Peretz, I. (2007). Singing proficiency in the general population. *J. Acoust. Soc. Am.* 121, 1182–1189. doi: 10.1121/1.2427111

Dalla Bella, S., Giguère, J. F., and Peretz, I. (2009). Singing in congenital amusia: an acoustical approach. *J. Acoust. Soc. Am.* 126, 414–424. doi: 10.1121/1.3132504

D'Ausilio, A., Pulvermüller, F., Salmas, P., Bufalari, I., Begliomini, C., and Fadiga, L. (2009). The motor somatotopy of speech perception. *Curr. Biol.* 19, 381–385. doi: 10.1016/j.cub.2009.01.017

Demorest, S. M., and Clements, A. (2007). Factors influencing the pitch-matching of junior high boys. *J. Res. Music Educ.* 55, 190–203. doi: 10.1177/002242940705500302

Diehl, R. L., Lotto, A. J., and Holt, L. L. (2004). Speech Perception. *Annu. Rev. Psychol.* 55, 149–179. doi: 10.1146/annurev.psych.55.090902.142028

Estis, J. M., Coblentz, J. K., and Moore, R. E. (2009). Effects of increasing time delays on pitch-matching accuracy in trained singers and untrained individuals. *J. Voice* 23, 439–445. doi: 10.1016/j.jvoice.2007.10.001

Estis, J. M., Dean-Claytor, A., Moore, R. E., and Rowell, T. L. (2011). Pitch-matching accuracy in trained singers and untrained individuals: the impact of musical interference and noise. *J. Voice* 25, 173–180. doi: 10.1016/j.jvoice.2009.10.010

Foxton, J. M., Dean, J. L., Gee, R., Peretz, I., and Griffiths, T. (2004). Characterization of deficits in pitch perception underlying tone-deafness. *Brain* 127, 801–810. doi: 10.1093/brain/awh105

Fyk, J. (1982). Perception of mistuned intervals in melodic context. *Psychol. Music Spec. Ed.* 36–41. [cited in Van Besouw et al., 2008]. Available online at: http://psycnet.apa.org/psycinfo/1984-14060-001

Goetze, M., Cooper, N., and Brown, C. J. (1990). Recent research on singing in the general music classroom. *Bull. Counc. Res. Music Educ. Counc.* 104, 16–37.

Goodale, M. A., and Milner, A. D. (1992). Separate visual pathways for perception and action. *Trends Neurosci.* 15, 20–25. doi: 10.1016/0166-2236(92)90344-8

Gosselin, N., Jolicœur, P., and Peretz, I. (2009). Impaired memory for pitch in congenital amusia. *Ann. N.Y. Acad. Sci.* 1169, 270–272. doi: 10.1111/j.1749-6632.2009.04762.x

Gottfried, T. L., Staby, A. M., and Ziemer, C. J. (2004). Musical experience and Mandarin tone discrimination and imitation. *J. Acoust. Soc. Am.* 115, 2545. doi: 10.1121/1.4783674

Green, G. A. (1990). The effect of vocal modeling on pitch-matching accuracy of elementary schoolchildren. *J. Res. Music Educ.* 38, 225–231. doi: 10.2307/3345186

Griffiths, T. D. (2008). Sensory systems: auditory action streams? *Curr. Biol.* 18, R387–R388. doi: 10.1016/j.cub.2008.03.007

Gunji, A., Koyama, S., Ishii, R., Levy, D., Okamoto, H., Kakigi, R., et al. (2003). Magnetoencephalographic study of the cortical activity elicited by human voice. *Neurosci. Lett.* 348, 13–16. doi: 10.1016/S0304-3940(03)00640-2

Hafke, H. Z. (2008). Nonconscious control of fundamental voice frequency. *J. Acoust. Soc. Am.* 123, 273–278. doi: 10.1121/1.2817357

Halpern, A. R., and Zatorre, R. J. (1979). Identification, discrimination, and selective adaptation of simultaneous musical intervals. *Percept. Psychophys.* 26, 384–395. doi: 10.3758/BF03204164

Heilman, K. M., Voeller, K., and Alexander, A. W. (1996). Developmental dyslexia: a motor-articulatory feedback hypothesis. *Ann. Neurol.* 39, 407–412. doi: 10.1002/ana.410390323

Heyes, C. (2011). Automatic imitation. *Psychol. Bull.* 137, 463–483. doi: 10.1037/a0022288

Hickok, G. (2009). Eight problems for the mirror neuron theory of action understanding in monkeys and humans. *J. Cogn. Neurosci.* 21, 1229–1243. doi: 10.1162/jocn.2009.21189

Hickok, G., and Poeppel, D. (2004). Dorsal and ventral streams: a framework for understanding aspects of the functional anatomy of language. *Cognition* 92, 67–99. doi: 10.1016/j.cognition.2003.10.011

Hickok, G., and Poeppel, D. (2007). The cortical organization of speech processing. *Nat. Rev. Neurosci.* 8, 393–402. doi: 10.1038/nrn2113

Hutchins, S., and Peretz, I. (2010). "Double dissociation of pitch production and perception," in *Presented at the Seventeenth Annual Meeting of* *the Cognitive Neuroscience Society*, (Montreal, QC), 127. Available online at: http://www.cogneurosociety.org/wordpress/wp-content/themes/CNStheme/downloads/CNS2010_Program.pdf

Hutchins, S., and Peretz, I. (2012a). A frog in your throat or in your ear. Studying the causes of poor singing. *J. Exp. Psychol. Gene.* 141, 76–97. doi: 10.1037/a0025064

Hutchins, S., and Peretz, I. (2012b). Amusics can imitate what they cannot discriminate. *Brain Lang*, 123, 234–239 doi: 10.1016/j.bandl.2012.09.011

Hutchins, S., and Peretz, I. (2013). Vocal pitch shift in congenital amusia (pitch deafness). *Brain Lang.* 125, 106–117. doi: 10.1016/j.bandl.2013.01.011

Hutchins, S., Roquet, C., and Peretz, I. (2012). The vocal generosity effect: how bad can your singing be. *Music Percept.* 30, 147–159. doi: 10.1525/mp.2012.30.2.147

Hutchins, S., Zarate, J. M., Zatorre, R. J., and Peretz, I. (2010). An acoustical study of vocal pitch matching in congenital amusia. *J. Acoust. Soc. Am.* 127, 504–512. doi: 10.1121/1.3270391

Hyde, K. L., Lerch, J. P., Zatorre, R. J., Griffiths, T. D., Evans, A. C., and Peretz, I. (2007). Cortical thickness in congenital amusia: when less is better than more. *J. Neurosci.* 27, 13028–13032. doi: 10.1523/JNEUROSCI.3039-07.2007

Hyde, K. L., and Peretz, I. (2004). Brains that are out of tune but in time. *Psycholog. Sci.* 15, 356–360. doi: 10.1111/j.0956-7976.2004.00683.x

Hyde, K. L., Zatorre, R. J., and Peretz, I. (2011). Functional MRI evidence of an abnormal neural network for pitch processing in congenital amusia. *Cereb. Cortex* 21, 292–299. doi: 10.1093/cercor/bhq094

Kleber, B., Birbaumer, N., Veit, R., Trevorrow, T., and Lotze, M. (2007). Overt and imagined singing of an Italian aria. *Neuroimage* 36, 889–900. doi: 10.1016/j.neuroimage.2007.02.053

Kleber, B., Veit, R., Birbaumer, N., Gruzelier, J., and Lotze, M. (2010). The brain of opera singers: experience-dependent changes in functional activation. *Cereb. Cortex* 20, 1144–1152. doi: 10.1093/cercor/bhp177

Kohler, E., Keysers, C., Umiltà, M. A., Fogassi, L., Gallese, V., and Rizzolatti, G. (2002). Hearing sounds, understanding actions: action representation in mirror neurons. *Science* 297, 846–848. doi: 10.1126/science.1070311

Kraus, N., Skoe, E., Parbery-Clark, A., and Ashley, R. (2009). Experience-induced malleability in neural encoding of pitch, timbre, and timing. *Ann. N.Y. Acad. Sci.* 1169, 543–557. doi: 10.1111/j.1749-6632.2009.04549.x

Krishnan, A. (2007). "Human frequency following response," in *Auditory Evoked Potentials: Basic Principles and Clinical Application*, eds R. F. Burkard, M. Don, and J. J. Eggermont (Baltimore, MD: Lippincott Williams and Wilkins), 313–335.

Krishnan, A., Swaminathan, J., and Gandour, J. T. (2008). Experience-dependent enhancement of linguistic pitch representation in the brainstem is not specific to a speech context. *J. Cogn. Neurosci.* 21, 1092–1105. doi: 10.1162/jocn.2009.21077

Krumhansl, C. L., and Iverson, P. (1992). Perceptual interactions between musical pitch and timbre. *J. Exp. Psychol. Hum. Percept. Perform*. 18, 739–751. doi: 10.1037/0096-1523.18.3.739

Léveque, Y., Giovanni, A., and Schön, D. (2012). Pitch-matching in poor singers: human model advantage. *J. Voice* 26, 293–298. doi: 10.1016/j.jvoice.2011.04.001

Levitin, D. J., and Rogers, S. E. (2005). Absolute pitch: perception, coding, and controversies. *Trends Cogn. Sci.* 9, 26–33. doi: 10.1016/j.tics.2004.11.007

Levy, D. A., Granot, R., and Bentin, S. (2001). Processing specificity for human voice stimuli: electrophysiological evidence. *Neuroreport* 12, 2653–2657. doi: 10.1097/00001756-200108280-00013

Liberman, A. M., and Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition* 21, 1–36. doi: 10.1016/0010-0277(85)90021-6

Liu, F., Patel, A. D., Fourcin, A., and Stewart, L. (2010). Intonation processing in congenital amusia: discrimination, identification and imitation. *Brain* 133, 1682–1693. doi: 10.1093/brain/awq089

Loui, P., Alsop, D., and Schlaug, G. (2009). Tone deafness: a new disconnection syndrome. *J. Neurosci.* 29, 10215–10220. doi: 10.1523/JNEUROSCI.1701-09.2009

Loui, P., Guenther, F. H., Mathys, C., and Schlaug, G. (2008). Action–perception mismatch in tone-deafness. *Curr. Biol.* 18, R331–R332. doi: 10.1016/j.cub.2008.02.045

Mandell, J., Schulze, K., and Schlaug, G. (2007). Congenital amusia: an auditory-motor feedback disorder. *Restor. Neurol. Neurosci.* 25, 323–334.

Marin, M. M., Gingras, B., and Stewart, L. (2012). Perception of musical timbre in congenital amusia: categorization, discrimination and short-term memory. *Neuropsychologia* 50, 367–378. doi: 10.1016/j.neuropsychologia.2011.12.006

Marques, C., Moreno, S., and Besson, M. (2007). Musicians detect pitch violation in a foreign language better than nonmusicians: behavioral and electrophysiological evidence. *J. Cogn. Neurosci.* 19, 1453–1463. doi: 10.1162/jocn.2007.19.9.1453

Marslen-Wilson, W. (1973). Linguistic structure and speech shadowing at very short latencies. *Nature.* 244, 522–523 doi: 10.1038/244522a0

Marslen-Wilson, W. (1985). Speech shadowing and speech comprehension. *Speech Commun.* 4, 55–73. doi: 10.1016/0167-6393(85)90036-6

McDermott, J. H., Lehr, A. J., and Oxenham, A. J. (2010). Individual differences reveal the basis of consonance. *Curr. Biol.* 20, 1035–1041. doi: 10.1016/j.cub.2010.04.019

Melara, R. D., and Marks, L. E. (1990a). HARD and SOFT interacting dimensions: differential effects of dual context on classification. *Percept. Psychophys.* 47, 307–325. doi: 10.3758/BF03210870

Melara, R. D., and Marks, L. E. (1990b). Interaction among auditory dimensions: timbre, pitch, and loudness. *Percept. Psychophys.* 48, 169–178. doi: 10.3758/BF03207084

Melara, R. D., and Marks, L. E. (1990c). Perceptual primacy of dimensions: support for a model of dimensional interaction. *J. Exp. Psychol. Hum. Percept. Perform.* 16, 398–414. doi: 10.1037/0096-1523.16.2.398

Mellody, M., and Wakefield, G. H. (2000). The time-frequency characteristics of violin vibrato: modal distribution analysis and synthesis. *J. Acoust. Soc. Am.* 107, 598. doi: 10.1121/1.428326

Moore, R. E., Estis, J., Gordon-Hickey, S., and Watts, C. (2008). Pitch discrimination and pitch matching abilities with vocal and nonvocal stimuli. *J. Voice* 22, 399–407. doi: 10.1016/j.jvoice.2006.10.013

Moore, R. E., Keaton, C., and Watts, C. (2007). The role of pitch memory in pitch discrimination and pitch matching. *J. Voice* 21, 560–567. doi: 10.1016/j.jvoice.2006.04.004

Moreau, P., Jolicœur, P., and Peretz, I. (2009). Automatic brain responses to pitch changes in congenital amusia. *Ann. N.Y. Acad. Sci.* 1169, 191–194. doi: 10.1111/j.1749-6632.2009.04775.x

Moreau, P., Jolicœur, P., and Peretz, I. (2013). Pitch discrimination without awareness in congenital amusia: evidence from event-related potentials. *Brain Cogn.* 81, 337–344. doi: 10.1016/j.bandc.2013.01.004

Moreno, S., and Besson, M. (2006). Musical training and language-related brain electrical activity in children. *Psychophysiology* 43, 287–291. doi: 10.1111/j.1469-8986.2006.00401.x

Moreno, S., and Bidelman, G. M. (2013). Examining neural plasticity and cognitive benefit through the unique lens of musical training. *Hear. Res.* doi: 10.1016/j.heares.2013.09.012. [Epub ahead of print].

Moreno, S., Marques, C., Santos, A., Santos, M., and Besson, M. (2009). Musical training influences linguistic abilities in 8-year-old children: more evidence for brain plasticity. *Cereb. Cortex* 19, 712–723. doi: 10.1093/cercor/bhn120

Näätänen, R., Paavilainen, P., Rinne, T., and Alho, K. (2007). The mismatch negativity (MMN) in basic research of central auditory processing: a review. *Clin. Neurophysiol.* 118, 2544–2590. doi: 10.1016/j.clinph.2007.04.026

Nikjeh, D. A., Lister, J. J., and Frisch, S. A. (2009). Preattentive cortical-evoked responses to pure tones, harmonic tones, and speech: influence of music training. *Ear Hear.* 30, 432–446. doi: 10.1097/AUD.0b013e3181a61bf2

Özdemir, E., Norton, A., and Schlaug, G. (2006). Shared and distinct neural correlates of singing and speaking. *Neuroimage* 33, 628–635. doi: 10.1016/j.neuroimage.2006.07.013

Patel, A. (2011). Why would musical training benefit the neural encoding of speech. The OPERA hypothesis. *Front. Psychol.* 2:142. doi: 10.3389/fpsyg.2011.00142. doi: 10.3389/fpsyg.2011.00142

Peretz, I. (2008). Musical disorders from behavior to genes. *Curr. Dir. Psycholog. Sci* 17, 329–333. doi: 10.1111/j.1467-8721.2008.00600.x

Peretz, I., Ayotte, J., Zatorre, R. J., Mehler, J., Ahad, P., Penhune, V. B., et al. (2002). Congenital amusia: a disorder of fine-grained pitch discrimination. *Neuron* 33, 185–191. doi: 10.1016/S0896-6273(01)00580-3

Peretz, I., Brattico, E., Järvenpää, M., and Tervaniemi, M. (2009). The amusic brain: in tune but unaware. *Brain* 132, 1277–1286. doi: 10.1093/brain/awp055

Peretz, I., Champod, A. S., and Hyde, K. (2003). Varieties of musical disorders. *Ann. N.Y. Acad. Sci.* 999, 58–75. doi: 10.1196/annals.1284.006

Peretz, I., Cummings, S., and Dubé, M. P. (2007). The genetics of congenital amusia (tone deafness): a family-aggregation study. *Am. J. Hum. Genet.* 81, 582–588. doi: 10.1086/521337

Perry, D. W., Zatorre, R. J., Petrides, M., Alivisatos, B., Meyer, E., and Evans, A. C. (1999). Localization of cerebral activity during simple singing. *Neuroreport* 10, 3979–3984. doi: 10.1097/00001756-199912160-00046

Pitt, M. A. (1994). Perception of pitch and timbre by musically trained and untrained listeners. *J. Exp. Psychol. Hum. Percept. Perform*. 20, 976. doi: 10.1037/0096-1523.20.5.976

Pitt, M. A., and Crowder, R. G. (1992). The role of spectral and dynamic cues in imagery for musical timbre. *J. Exp. Psychol. Hum. Percept. Perform*. 18, 728. doi: 10.1037/0096-1523.18.3.728

Pfordresher, P. Q., and Brown, S. (2007). Poor-pitch singing in the absence of "tone deafness". *Music Percept.* 25, 95–115. doi: 10.1525/mp.2007.25.2.95

Pfordresher, P. Q., and Brown, S. (2009). Enhanced production and perception of musical pitch in tone language speakers. *Attent. Percept. Psychophys.* 71, 1385–1398. doi: 10.3758/APP.71.6.1385

Pfordresher, P. Q., Brown, S., Meier, K. M., Belyk, M., and Liotti, M. (2010). Imprecise singing is widespread. *J. Acoust. Soc. Am.* 128, 2182. doi: 10.1121/1.3478782

Prame, E. (1997). Vibrato extent and intonation in professional Western lyric singing. *J. Acoust. Soc. Am.* 102, 616. doi: 10.1121/1.419735

Rakowski, A. (1990). Intonation variants of musical intervals in isolation and in musical contexts. *Psychol. Music* 18, 60–72. doi: 10.1177/0305735690181005

Rizzolatti, G., and Arbib, M. A. (1998). Language within our grasp. *Trends Neurosci*. 21, 188–194. doi: 10.1016/S0166-2236(98)01260-0

Rizzolatti, G., and Craighero, L. (2004). The mirror-neuron system. *Annu. Rev. Neurosci*. 27, 169–192. doi: 10.1146/annurev.neuro.27.070203.144230

Rizzolatti, G., Fogassi, L., and Gallese, V. (2001). Neurophysiological mechanisms underlying the understanding and imitation of action. *Nat. Rev. Neurosci.* 2, 661–670. doi: 10.1038/35090060

Sataloff, R. T. (2005). *Professional Voice: the Science and Art of Clinical Care*. San Diego, CA: Plural Publishing.

Seashore, C. E. (1938). *Psychology of Music*. New York, NY: McGraw-Hill

Shonle, J. I., and Horan, K. E. (1980). The pitch of vibrato tones. *J. Acoust. Soc. Am.* 67, 246. doi: 10.1121/1.383733

Siegel, J. A., and Siegel, W. (1977). Categorical perception of tonal intervals: musicians can't tell sharp from flat. *Percept. Psychophys.* 21, 399–407. doi: 10.3758/BF03199493

Skoe, E., and Kraus, N. (2010). Auditory brain stem response to complex sounds: a tutorial. *Ear Hear.* 31, 302–324. doi: 10.1097/AUD.0b013e3181cdb272

Slevc, L. R., and Miyake, A. (2006). Individual differences in second-language proficiency does musical ability matter. *Psycholog. Sci.* 17, 675–681. doi: 10.1111/j.1467-9280.2006.01765.x

Strait, D. L., Chan, K., Ashley, R., and Kraus, N. (2012). Specialization among the specialized: auditory brainstem function is tuned in to timbre. *Cortex* 48, 360–362. doi: 10.1016/j.cortex.2011.03.015

Sundberg, J. (1979). Perception of singing. *Speech Transm. Lab. Q. Prog. Status Rep.* 20, 1–48.

Sundberg, J. (1982). In tune or not. A study of fundamental frequency in music practise. *Speech Transm. Lab. Q. Prog. Status Rep.* 23, 49–78.

Sundberg, J. (1987). *The Science of the Singing Voice*. Dekalb, IL: Northern Illinois University Press.

Sundberg, J., Gramming, P., and Lovetri, J. (1993). Comparisons of pharynx, source, formant, and pressure characteristics in operatic and musical theatre singing. *J. Voice* 7, 301–310. doi: 10.1016/S0892-1997(05)80118-3

Sundberg, J., Prame, E., and Iwarsson, J. (1996). "Replicability and accuracy of pitch patterns in professional singers," in *Controlling Complexity and Chaos: 9th Vocal Fold Physiology Symposium*, (San Diego, CA: Singular Press), 291–306.

Sundberg, J., and Thalén, M. (2010). What is "Twang". *J. Voice* 24, 654–660. doi: 10.1016/j.jvoice.2009.03.003

Takeuchi, A. H., and Hulse, S. H. (1993). Absolute pitch. *Psychol. Bull.* 113, 345. doi: 10.1037/0033-2909.113.2.345

Tillmann, B., Schulze, K., and Foxton, J. M. (2009). Congenital amusia: a short-term memory deficit for non-verbal, but not verbal sounds. *Brain Cogn.* 71, 259–264. doi: 10.1016/j.bandc.2009.08.003

Titze, I. R. (2000). *Principles of Voice Production. (Second Printing).* Iowa, IA: National Center for Voice and Speech.

Tremblay-Champoux, A., Dalla Bella, S., Phillips-Silver, J., Lebrun, M. A., and Peretz, I. (2010). Singing proficiency in congenital amusia: imitation helps. *Cogn. Neuropsychol*. 27, 463–476. doi: 10.1080/02643294.2011.567258

van Besouw, R. M., Brereton, J. S., and Howard, D. M. (2008). Range of tuning for tones with and without vibrato. *Music Percept.* 26, 145–155. doi: 10.1525/mp.2008.26.2.145

Vurma, A. (2010). Mistuning in two-part singing. *Logoped. Phoniatr. Vocol.* 35, 24–33. doi: 10.3109/14015430903581591

Vurma, A., and Ross, J. (2006). Production and perception of musical intervals. *Music Percept.* 23, 331–344. doi: 10.1525/mp.2006.23.4.331

Warrier, C. M., and Zatorre, R. J. (2002). Influence of tonal context and timbral variation on perception of pitch. *Percept. Psychophys.* 64, 198–207. doi: 10.3758/BF03195786

Watts, C. R., and Hall, M. D. (2008). Timbral influences on vocal pitch-matching accuracy. *Logoped. Phoniatr. Vocol*. 33, 74–82. doi: 10.1080/14015430802028434

Watts, C., Barnes-Burroughs, K., Adrianopoulos, M., and Carr, M. (2003a). Potential factors related to untrained singing talent: a survey of singing pedagogues. *J. Voice* 17, 298–307. doi: 10.1067/S0892-1997(03)00068-7

Watts, C., Murphy, J., and Barnes-Burroughs, K. (2003b). Pitch matching accuracy of trained singers, untrained subjects with talented singing voices, and untrained subjects with nontalented singing voices in conditions of varying feedback. *J. Voice* 17, 185–194. doi: 10.1016/S0892-1997(03)00023-7

Watts, C., Moore, R., and McCaghren, K. (2005). The relationship between vocal pitch-matching skills and pitch discrimination skills in untrained accurate and inaccurate singers. *J. Voice* 19, 534–543. doi: 10.1016/j.jvoice.2004.09.001

Williamson, V. J., McDonald, C., Deutsch, D., Griffiths, T. D., and Stewart, L. (2010). Faster decline of pitch memory over time in congenital amusia. *Adv. Cogn. Psychol*. 6, 15–22. doi: 10.2478/v10053-008-0073-5

Yarbrough, C., Green, G. A., Benson, W., and Bowers, J. (1991). Inaccurate singers: an exploratory study of variables affecting pitch-matching. *Bull. Counc. Res. Music Educ.* 107, 23–34.

Yoo, L., Sullivan, D. S. Jr., Moore, S., and Fujinaga, I. (1998). "The effect of vibrato on response time in determining the pitch relationship of violin tones," in *Proceedings of the 5th International Conference on Music Perception and Cognition.* (Seoul), 209–211.

Zarate, J. M., Delhommeau, K., Wood, S., and Zatorre, R. J. (2010a). Vocal accuracy and neural plasticity following micromelody-discrimination training. *PLoS ONE* 5:e11181. doi: 10.1371/journal.pone.0011181

Zarate, J. M., Wood, S., and Zatorre, R. J. (2010b). Neural networks involved in voluntary and involuntary vocal pitch regulation in experienced singers. *Neuropsychologia* 48, 607–618. doi: 10.1016/j.neuropsychologia.2009.10.025

Zarate, J. M., and Zatorre, R. J. (2008). Experience-dependent neural substrates involved in vocal pitch regulation during singing. *Neuroimage* 40, 1871–1887. doi: 10.1016/j.neuroimage.2008.01.026

Zwicker, E., and Fastl, H. (1999). *Psychoacoustics: Facts and models,* Vol. 2. Berlin: Springer. doi: 10.1007/978-3-662-09562-1

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 30 July 2013; accepted: 17 October 2013; published online: 05 November 2013.

Citation: Hutchins S and Moreno S (2013) The Linked Dual Representation model of vocal perception and production. Front. Psychol. ***4****:825. doi: 10.3389/fpsyg.2013.00825*

This article was submitted to Cognitive Science, a section of the journal Frontiers in Psychology.

frontiers in
PSYCHOLOGY

ORIGINAL RESEARCH ARTICLE
published: 16 January 2014
doi: 10.3389/fpsyg.2013.01024

Perceptual pitch deficits coexist with pitch production difficulties in music but not Mandarin speech

Wu-xia Yang[1,2]†, Jie Feng[1,2]†, Wan-ting Huang[1,2]†, Cheng-xiang Zhang[1,2]† and Yun Nan[1,2,3]*

[1] State Key Laboratory of Cognitive Neuroscience and Learning, Beijing Normal University, Beijing, China
[2] International Data Group/McGovern Institute for Brain Research, Beijing Normal University, Beijing, China
[3] Center for Collaboration and Innovation in Brain and Learning Sciences, Beijing Normal University, Beijing, China

Edited by:
Bruno Gingras, University of Vienna, Austria

Reviewed by:
Manuela Maria Marin, University of Vienna, Austria
Karen Wise, Guildhall School of Music & Drama, UK

****Correspondence:***
Yun Nan, State Key Laboratory of Cognitive Neuroscience and Learning, Beijing Normal University, 19 Xin-Wai Street, Hai-Dian District, Beijing 100875, China
e-mail: dr.yunnan@gmail.com

†Wu-xia Yang, Jie Feng, Wan-ting Huang, and Cheng-xiang Zhang have contributed equally to this work.

Congenital amusia is a musical disorder that mainly affects pitch perception. Among Mandarin speakers, some amusics also have difficulties in processing lexical tones (tone agnosics). To examine to what extent these perceptual deficits may be related to pitch production impairments in music and Mandarin speech, eight amusics, eight tone agnosics, and 12 age- and IQ-matched normal native Mandarin speakers were asked to imitate music note sequences and Mandarin words of comparable lengths. The results indicated that both the amusics and tone agnosics underperformed the controls on musical pitch production. However, tone agnosics performed no worse than the amusics, suggesting that lexical tone perception deficits may not aggravate musical pitch production difficulties. Moreover, these three groups were all able to imitate lexical tones with perfect intelligibility. Taken together, the current study shows that perceptual musical pitch and lexical tone deficits might coexist with musical pitch production difficulties. But at the same time these perceptual pitch deficits might not affect lexical tone production or the intelligibility of the speech words that were produced. The perception-production relationship for pitch among individuals with perceptual pitch deficits may be, therefore, domain-dependent.

Keywords: congenital amusia, tone agnosia, lexical tone, musical pitch, perception, production

INTRODUCTION

Successful communication relies on the seamless integration of auditory perception and vocal production. It is widely acknowledged that auditory perception strongly affects vocal production. To a certain extent, impaired auditory perception hinders vocal production (Peng et al., 2004; Han et al., 2007; Xu et al., 2011). This is true for both music and language. As the two most important communication vehicles, music and language share a vital element, namely pitch. Exploring the impact of impaired pitch perception upon pitch production across music and language domains is thus the key to understanding the influence of impaired auditory perception on vocal production.

In the last decade, a developmental perceptual pitch deficit known as congenital amusia (Peretz, 2001; Peretz et al., 2002) has increasingly attracted research attention. This characteristic acoustical pitch deficit (Hyde and Peretz, 2004; Pfeuty and Peretz, 2010) was initially related to deficient musical pitch perception (e.g., Foxton et al., 2004), which occurs independently of neurological trauma, mental retardation, autism, deafness, or lack of musical exposure. Subsequent research has suggested that its origin not only is related to the impaired fine-grained pitch processing (Hyde and Peretz, 2003; Foxton et al., 2004), but also involves compromised pitch working memory (Gosselin et al., 2009; Tillmann et al., 2009; Williamson and Stewart, 2010; Williamson et al., 2010; Albouy et al., 2013), timbre perception deficits (Marin et al., 2012), and emotional prosody perception difficulties (Thompson et al., 2012).

Recent studies have evidenced amusia's related pitch deficits in the speech domain, where pitch also plays an important role. Individuals with amusia (hereafter, "amusics") may demonstrate lexical tone deficits among speakers of tonal (Nan et al., 2010; Liu et al., 2012) and non-tonal languages (Tillmann et al., 2011). Likewise, amusics have been shown to suffer from parallel speech intonation problems, a finding that has held true among speakers of both tonal (Jiang et al., 2010) and non-tonal languages (Liu et al., 2010). One of our earlier studies showed that a minority subgroup of amusic Mandarin speakers also had difficulty with lexical tone discrimination and identification in Mandarin speech (hereafter "tone agnosics") (Nan et al., 2010).

This cross-domain perceptual pitch deficit offers an ideal opportunity to understand the influence of auditory perception on vocal production in music and speech. The current study sets out to investigate to what extent these perceptual pitch deficits may be related to pitch production impairments in music and Mandarin speech among Mandarin speakers. Mandarin Chinese is a tone language which relies on pitch variations to alter the meaning of words. Although amusic individuals who speak non-tonal languages also demonstrate similar problems with lexical tones (Tillmann et al., 2011), to study amusics and tone agnosics among tone language speakers will present unique perspectives on the impact of impaired pitch perception on pitch production. This is because in a tone language environment, the perception and production of lexical tones are basic communication needs of daily necessity.

In research on speakers of non-tonal languages, perceptual pitch deficits have generally been associated with poor pitch production in music (Dalla Bella et al., 2009; Hutchins et al., 2010). According to the vocal sensorimotor loop model (VSL model, Berkowska and Dalla Bella, 2009), perception is a necessary but insufficient element of vocal production. Motor components, such as motor planning and auditory-motor mapping, also play vital roles (Berkowska and Dalla Bella, 2009). In accordance with the assumptions of the VSL model, the existence of normal pitch perception may not necessarily preclude poor pitch singing, due to the possibility of independently deficient motor-related functions (Dalla Bella et al., 2007). The inclusion of both overt and covert perceptual components in the VSL model offers greater explanatory power. Within the amusic population, cases have been noted in which pitch perception impairments do not necessarily cause vocal production deficits (Loui et al., 2008; Dalla Bella et al., 2009). This could be explained by preserved covert but impaired overt pitch perceptual abilities; such an instance would corroborate the findings of previous mismatch negativity (MMN) studies on near-normal neural processing of fine-grained pitch differences without awareness in congenital amusia (Moreau et al., 2009; Peretz et al., 2009). More specifically, the auditory cortex may function relatively normally in these amusics, but its connectivity with the pars orbitalis of the right inferior frontal gyrus may function aberrantly (Hyde et al., 2011). For the other amusics, however, it is very likely that their abnormalities reside not only in the fronto-temporal pathway but also in the auditory cortices, as shown by a more recent study using magnetoencephalography and voxel-based morphometry (Albouy et al., 2013).

With regard to speech, however, linguistic tone deficits are not necessarily always correlated with production impairments, as suggested by a recent study showing that amusic speakers of non-tonal languages are unable to discriminate between speech intonations despite production being intact (Hutchins and Peretz, 2012). A more recent study, in contrast, reported impaired speech and song imitation among Mandarin speaking amusics (Liu et al., 2013). This sometimes asymmetrical perception-production relationship between music and speech domains could not be explained by the apparent acoustical difference between musical pitch and linguistic intonation – i.e., fine-grained for musical pitch but coarse-grained for linguistic intonation. As shown in one recent study (Dalla Bella et al., 2011), after controlling for acoustic pitch differences across domains, a young university student with intact musical pitch perception but impaired musical pitch imitation was shown to have intact linguistic tone production.

Among tonal language speakers, similar perceptual lexical tone deficits have been observed among individuals whose lexical tone production is intact (Nan et al., 2010). However, pitch production in music (singing abilities) among Mandarin speakers, especially among amusic and tone agnosic Mandarin speakers, has not yet been fully examined (but see Liu et al., 2013 for impaired pitch production in both music and speech for Mandarin speaking amusics only). It should be noted that, so far, the exact nature of tone agnosia is not yet clear. As shown in our early work (Nan et al., 2010), the tone agnosics had little problem identifying lexical tones carried by the same segments (e.g., word onsets and rhymes), but they had difficulties in tones embedded in different segments. This might be due to low executive or attentional control in these individuals. In the current study, we controlled these factors by matching the control, the amusic, and the tone agnosic groups on measures of executive functions and working memory. The tone agnostics were thus also amusics but with additional perceptual lexical tone disorders, as the tone agnosics and the amusics demonstrated similar levels of music perception deficits [as indicated by the similar melodic Montreal Battery of Evaluation of Amusia (MBEA) tests scores between these two groups].

In the current study, we tested musical pitch and lexical tone production among age- and IQ-matched amusics and tone agnosics (i.e., amusics who were, at the same time, tone agnosics) relative to normal controls, with three specific research aims. The first was to understand how musical pitch perception deficit is related to musical pitch production among Mandarin speakers. Second, we wanted to explore how the lexical tone impairments observed in tone agnosics would be related to musical pitch production relative to the other amusics who had intact lexical tone perception. Based on previous results of the dissociation between musical perceptual and productive abilities in amusia, we speculated that musical pitch production difficulties might be present in some but not all amusic participants. It is possible that the same holds true in tone agnosics, since they were also amusics. Alternatively, tone agnostics might show more severe musical pitch production deficits compared to the amusics, if the lexical tone deficit were detrimental to musical pitch production. The third research aim was to test lexical tone production using a novel set of objective analyses. It is possible that the subjective rating system used in our earlier study (Nan et al., 2010) has been insufficient for detecting subtle lexical tone production deficits, especially among tone agnosics.

MATERIALS AND METHODS

PARTICIPANTS

Sixteen amusics (six females) and twelve matched controls (five females) participated in the study. Among these 16 amusics, eight were also impaired in lexical tone perception and were thus identified as tone agnosics. A summary of all participants' characteristics is provided in **Table 1**. All participants were university students in Beijing and native Mandarin speakers without formal musical training. They reported no vocal, neurological, or audiological deficits. Their binaural audiometric thresholds were at or below 20 dB hearing level for octaves from 250 to 8000 Hz. Additionally, the controls reported no difficulty singing. Among all participants, 23 were right-handed and five were left-handed, as assessed by the Edinburgh Handedness Inventory (Oldfield, 1971). Informed written consent was obtained from all participants. This research was approved by the Institutional Review Board of Beijing Normal University.

All participants were assessed with the six tests of the MBEA (Peretz et al., 2003) and the lexical tone perception tests employed in our previous study (Nan et al., 2010). The MBEA includes three melodic pitch-based tests (scale, contour and interval), two time-based tests (rhythm and meter), and one memory test. All amusic participants scored below the cut-off score of 71.7%, which corresponds to two SDs below the mean of the normal controls that was obtained in our earlier study (Nan et al., 2010). The lexical tone perception test contains identification and discrimination tasks. Among the 16 amusics, eight tone agnosics were identified based on the same criteria (i.e., performance below the cut-off scores of

Table 1 | The characteristics of the controls, the amusics, and the tone agnosics with percentages of correct responses on the MBEA and lexical tone perception tests.

	Control ($n = 12$)	Amusia ($n = 8$)	Agnosia ($n = 8$)
Mean age (range)	22.5 (19–26)	21.8 (20–25)	24.5 (19–28)
Male/female	7/5	5/3	5/3
Right/left handedness	10/2	6/2	5/3
Performance IQ (SD)	115.8 (8.1)	111.3 (6.5)	110.3 (6.0)
Verbal IQ (SD)	128.8 (6.0)	126.9 (5.5)	124.9 (8.0)
Executive function (SD)	13.6 (0.9)	13.3 (0.6)	12.8 (1.1)
Working memory (SD)	15.1 (1.9)	14.1 (1.2)	13.6 (2.2)
MBEA mean (SD)			
Scale	91.7 (6.9)	60.4 (13.3)	57.4 (8.6)
Contour	90.0 (8.4)	63.3 (6.7)	55.8 (11.1)
Interval	85.8 (11.1)	60.8 (8.5)	60.4 (7.9)
Rhythm	92.8 (6.2)	65.0 (13.1)	64.3 (11.9)
Meter	81.4 (16.6)	60.4 (21.3)	65.5 (10.4)
Memory	93.4 (4.0)	74.6 (9.6)	67.6 (7.6)
Global	89.2 (6.2)	64.1 (3.0)	61.8 (5.4)
Lexical tone mean (SD)			
Mean	96.5 (4.1)	96.1 (2.8)	62.8 (12.6)
Discrimination (different segments)	94.3 (4.7)	92.6 (7.3)	65.0 (8.6)

79.2% for the lexical tone discrimination test with different segments – viz., word onsets and rhymes – and 80% for the average lexical tone perception tests, both of which correspond to three SDs below the means of the normal controls) as employed in our previous study (Nan et al., 2010). Except with respect to the lexical tone tests, the tone agnosics performed equivalently to the amusics on the MBEA tests (all $ps > 0.1$).

The amusic and tone agnosic groups were matched for age, handedness, performance IQ, and verbal IQ based on Wechsler Adult Intelligence Scale-Revised by China (WAIS-RC; Gong, 1992) with the control group (all $ps > 0.1$). Additionally, these two groups were also matched on measures of executive functions and working memory as derived from WAIS-RC with the control group (both $ps > 0.1$). Specifically, executive/attentional functions were indexed by the block design and similarities tests. The block design test taps attentional aspects of executive function (Chase et al., 1984; Kazui et al., 2011), whereas the similarities test may reflect abstraction and reasoning (Stern and Prohaska, 1996). The working memory index included arithmetic and digit span tests (Wechsler, 1981).

MATERIALS AND PROCEDURES

Music production test

All of the musical stimuli were computer-synthesized with a piano-like timbre. There were two imitation conditions: one-note and three-note. One-note condition included 13 trials, each being one of the 13 notes (G3, A3, B3, C4, D4, E4, F4, G4, A4, B4, C5, D5, and E5). Each note lasted 500 ms. Three-note condition consisted of eight trials, with two trials for each of the four different directions ("up", "down", "down up", and "up down") (**Figure 1**). Except for G3, A3, and D5, the remaining notes in one-note condition were also used in three-note condition. A trial in three-note condition lasted 2500 ms, including three 500 ms notes and two 500 ms gaps in between.

Lexical tone production test

A female voice actor who was a native Mandarin speaker produced a list of one-syllable words for the lexical tone production tests. Recordings were made in a sound-proof booth using a Sony 60EC digital recorder and an NT1 microphone with a Samson MDR8 mixer, at a sampling rate of 44.1 kHz. There were two conditions: monosyllabic word production and trisyllabic nonsense word production. The monosyllabic word production test contained eight one-syllable words, with two lexical tones from each category (four categories: 1 = level, 2 = mid-rising, 3 = dipping, and 4 = high-falling). The trisyllabic nonsense word production test had eight trials, each formed by three one-syllable words. As a result, the trisyllabic word condition contained 24 one-syllable words in total, with six lexical tones from each tone category. For both lexical tone production conditions, the durations of the syllables (on which the lexical tones were carried) were not significantly different (mean ± SD for the monosyllabic word condition: 537.5 ± 126.9 ms; for the trisyllabic nonsense word condition: 531.7 ± 131.6 ms; $p > 0.1$ between conditions). As a result, a trial in the trisyllabic nonsense word condition lasted 2595.0 ± 225.6 ms (including two 500 ms gaps), about five times the mean duration of the monosyllabic word condition. The acoustic characteristics of the lexical tones for production tests are listed in **Table 2**.

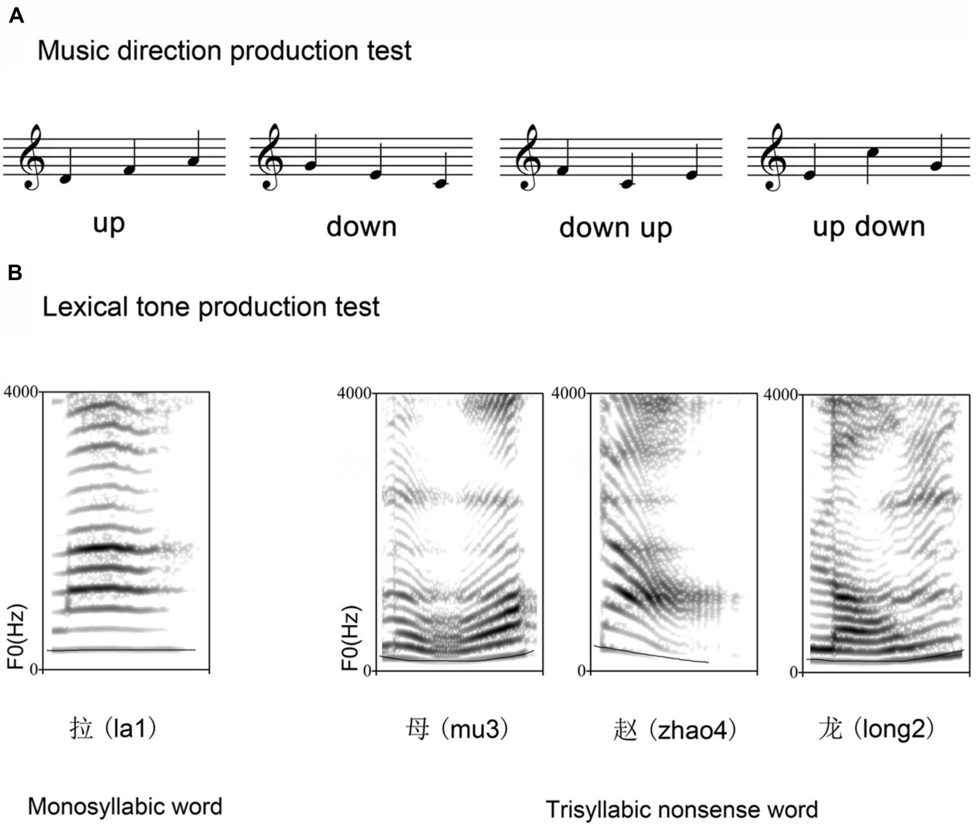


FIGURE 1 | The exemplar stimuli used in the music three-note imitation test (A) and lexical tone production test (B).

Procedure

All of the words and musical stimuli were equalized for sound intensity with 10 ms linear onset/offset ramps using Praat (Boersma, 2001). All of the production tests were completed in a sound-treated booth in a single session. The stimuli were presented to the participants binaurally through Sennheiser HD 201 headphones with individually adjustable volume. Participants were asked to vocally imitate the musical note sequences or the Mandarin words that they heard as closely as possible. They were to match not only pitch (for Mandarin words the accurate lexical tonal contours were emphasized) but also tempo of the target stimuli. The order of the music production test and the lexical tone production test was counterbalanced across participants. Before each test, detailed instructions and a warm-up phase were given to ensure that all participants understood the task. In the music production test, the participants were encouraged to use syllable /la/ when imitating the note sequences. All imitation samples were recorded onto a Marantz PMD-620 digital recorder (Marantz Professional, Itasca, IL, USA), with a sampling rate of 44.1 kHz. The whole test lasted approximately 30 min for each participant.

Table 2 | Acoustic characteristics of the lexical tones for production tests.

	Mean F0 (Hz)	F0 range (Hz)	Pitch glide size (Hz)	Duration (ms)
Tone 1	288.5 (8.1)	281.6∼306.2	24.6 (11.3)	541.3 (91.3)
Tone 2	213.0 (23.1)	177.9∼289.2	111.1 (20.3)	580.0 (78.0)
Tone 3	156.6 (9.0)	95.9∼200.9	105.0 (23.0)	647.5 (45.0)
Tone 4	234.9 (9.9)	112.5∼332.3	219.8 (31.2)	363.8 (79.5)

The numbers in parentheses are standard deviations.

DATA ANALYSIS

Music production test

The fundamental frequency (F0) and duration of each produced note were extracted using Praat (Boersma, 2001) based on the identified steady-state phase of each sung note. Accordingly, three F0-based measures and one duration-related measure were calculated: note deviation, interval deviation, direction accuracy, and duration deviation. F0 measurements in hertz were converted to cents (100 cents = 1 semitone).

Note deviation. Note deviation referred to the absolute difference between the produced F0 and the target F0 (e.g., Pfordresher et al., 2010). Octave errors were corrected.

Interval deviation. Interval deviation was calculated as the absolute difference between the produced interval and the target interval for the three-note condition.

Direction accuracy. Direction accuracy represented the rates of correctly produced directions in the three-note condition. A response was defined as correct if the successively produced three-note sequence shared the same direction as the target sequence.

Duration deviation. Duration deviation indicated the average absolute differences of duration between the produced note and the target note (Dalla Bella et al., 2009).

Note deviation and duration deviation were applicable for both one-note and three-note imitation conditions, whereas interval deviation and direction accuracy were entirely based on the three-note condition.

Lexical tone production test

Subjective assessment. Three independent raters (two female, native Mandarin speakers with a mean age of 24 years) classified each of the produced lexical tones from each participant as tone 1, 2, 3, or 4. When correct, the lexical tone that was produced was considered a hit. For each participant, the average scores for the monosyllabic word and the trisyllabic nonsense word production conditions were calculated separately.

Objective analysis. For each produced lexical tone, the mean F0, pitch glide size (i.e., the mean difference between minimum and maximum F0; Nan et al., 2010), and the duration were extracted using Praat. Octave errors were corrected for the mean F0 when necessary. These measures were then compared to those of the target lexical tones to calculate the mean F0 deviation, pitch glide size deviation, and duration deviation. All these deviation measures were calculated as the absolute difference between the produced lexical tone and the target one.

Statistical analysis

For all ANOVAs, the assumptions of normality and homogeneity of variance were met. If violated, then the non-parametric alternative to the planned ANOVA would be conducted instead. For all the repeated measures ANOVAs, however, an additional assumption of sphericity of the covariance matrix was also ensured. The Greenhouse–Geisser correction was applied when the sphericity assumption was violated. Bonferroni corrections were applied in multiple *post hoc* tests.

RESULTS

MUSIC PRODUCTION RESULTS

Note deviation

A mixed-model two-way repeated-measure ANOVA of note deviation with condition (2) as a within-subjects factor and group (3) as a between-subjects factor found a main effect of group [$F(2,25) = 12.093$, $p < 0.001$] and an interaction between condition and group [$F(2,25) = 5.607$, $p = 0.01$]. As shown in **Table 3**, the controls significantly outperformed the amusics and tone agnosics (both ps < 0.01), whereas the latter two groups performed similarly on note deviation ($p > 0.5$). Simple effect analysis of the observed interaction between condition and group suggested that the controls performed significantly better in three-note imitation condition than in one-note imitation condition ($p < 0.01$), whereas the amusics and tone agnosics both performed similarly in these two conditions (both ps > 0.05).

Table 3 | The music production performances among the three groups.

	Control	Amusia	Agnosia
Note deviation (semitone)	1.9 (0.7)	2.9 (0.5)	2.8 (0.4)
Interval deviation (semitone)	1.4 (0.7)	1. 5 (0.3)	1.6 (0.8)
Direction accuracy (%)	98.9 (3.6)	75.0 (24.1)	70.3 (24.0)
Duration deviation (ms)	323.2 (171.6)	253. 5 (96.0)	229.1 (120.3)

The numbers in parentheses are standard deviations.

Interval deviation

A one-way ANOVA of interval deviation revealed no main effect of group. All three groups performed indistinguishably on interval deviation [$F(2,25) = 1.102$, $p = 0.348$; **Table 3**].

Direction accuracy

Direction accuracy of the controls violated the normality assumption for ANOVA (one-sample Kolmogorov–Smirnov test: $Z = 1.837$, $p = 0.002$). The Kruskal–Wallis H test revealed a statistically significant difference between the three groups [$H(2) = 10.772$, $p = 0.005$], with a mean rank of 19.75 for the controls, 11.44 for the amusics, and 9.69 for the tone agnosics (**Table 3**). Pairwise comparison using Mann–Whitney U tests suggested that the controls significantly outperformed the amusics and tone agnosics (both ps < 0.01), whereas the latter two groups performed similarly on direction accuracy ($p > 0.5$).

Duration deviation

A mixed-model two-way ANOVA of duration deviation with condition (2) as a within-subjects factor and group (3) as a between-subjects factor did not reveal any significant main effects or interactions. The three groups performed equivalently on duration deviation (**Table 3**), suggesting that, although some of the amusics and tone agnosics showed impairments on the frequency dimension of musical pitch production, their performances on the time dimension seemed relatively unaffected.

LEXICAL TONE PRODUCTION RESULTS

Subjective assessment

Lexical tone production was highly accurate in all three groups, with 100% correct for both the controls and the amusics in both the monosyllabic and trisyllabic conditions. The tone agnosic group also yielded perfect lexical tone production scores for both the monosyllabic (99.0 ± 2.3%) and the trisyllabic (99.5 ± 1.4%) conditions. No main effects or interactions were observed in a mixed-model two-way ANOVA.

Objective analysis

Separate mixed-model three-way ANOVAs (2 conditions × 3 groups × 4 tone categories) of the results of the objective acoustic analysis, including the mean F0 deviation, pitch glide size deviation, and duration deviation, did not show any significant

Table 4 | The results of the objective analysis for the two lexical tone production tests across the three groups.

	Control	Amusia	Agnosia
Mean F0 deviation (semitone)			
Tone 1	1.9 (1.7)	2.9 (1.7)	1.9 (1.4)
Tone 2	2.5 (1.4)	3.3 (1.1)	2.2 (1.1)
Tone 3	3.4 (0.9)	3.7 (1.4)	4.0 (0.8)
Tone 4	2.6 (1.6)	2.7 (1.4)	2.1 (1.5)
Pitch glide size deviation (semitone)			
Tone 1	0.7 (0.3)	0.6 (0.3)	0.9 (0.5)
Tone 2	2.4 (0.9)	2.4 (1.3)	2.4 (0.8)
Tone 3	5.7 (1.4)	6.7 (1.6)	5.4 (2.0)
Tone 4	7.9 (1.7)	9.3 (2.1)	7.5 (2.3)
Duration deviation (ms)			
Tone 1	92.8 (45.6)	78.8 (42.1)	94.0 (32.7)
Tone 2	106.2 (42.4)	96.2 (30.5)	57.5 (39.5)
Tone 3	131.7 (66.4)	113.2 (84.6)	91.6 (50.0)
Tone 4	107.8 (68.6)	106.6 (57.7)	83.1 (35.5)

The numbers in parentheses are standard deviations. Tone 1 is the level tone, tone 2 mid-rising, tone 3 dipping, and tone 4 high-falling.

main effects or interactions involving group (see **Table 4**). This is consistent with the results obtained in the subjective assessments, suggesting that neither the tone agnosics nor the amusics were impaired in lexical tone production relative to the controls.

CORRELATION ANALYSIS

Spearman's Rank Correlation analyses were conducted to examine the relationship between musical pitch perception, lexical tone perception, musical pitch production, and objective lexical tone production measures across and within the three groups of participants. The subjective lexical tone production scores were not taken into account due to the ceiling effect. Pitch perception measures included the melodic MBEA scores (averaged across the scale, contour, and interval tests) and the average lexical tone perception scores. Pitch production measures included note deviation, interval deviation, direction accuracy, and duration deviation for music as well as the mean F0 deviation, pitch glide size deviation, and duration deviation for lexical tones.

The results showed that the melodic MBEA scores were positively correlated with the average lexical tone perception scores across the three groups [$r_s(28) = 0.575$, $p = 0.001$]. More importantly, the melodic MBEA scores were also significantly and negatively correlated with note deviation across the three groups (**Figure 2**), $r_s(28) = -0.695$, $p < 0.001$. However, neither of these two correlations held within each individual group (all *p*s > 0.1). There was no significant correlation between pitch perception measures (both melodic MBEA scores and the average lexical tone tests scores) and other music pitch production measures (including direction accuracy, interval deviation, and duration deviation) or lexical tone production measures (all *p*s > 0.1). There was no significant correlation between the equivalent music production and perception measures (i.e., direction accuracy and MBEA contour score or interval deviation and MBEA interval score) either (both *p*s > 0.1).

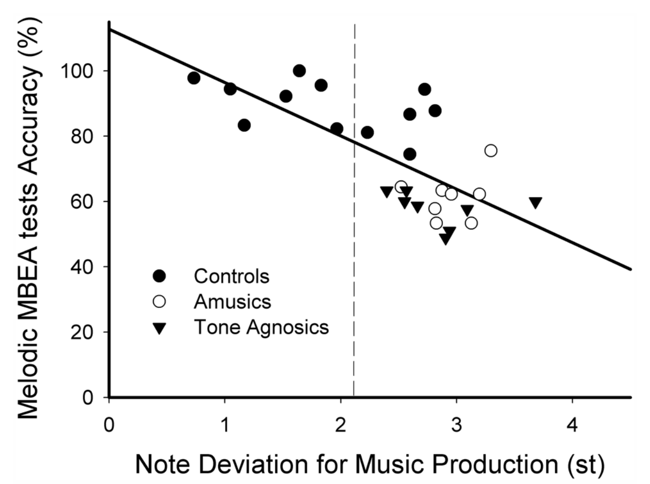


FIGURE 2 | A significant negative correlation between the melodic MBEA score and note deviation for music production tests among the three groups. A dashed line divides two distinct groups based on individual performances on note deviation for the controls, the amusic, and tone agnosic groups.

These results suggest that musical pitch perception is tightly linked to lexical tone perception, and musical pitch production and perception are significantly correlated (**Figure 2**).

INDIVIDUAL-LEVEL ANALYSIS

We conducted individual-level analyses based on note deviation and direction accuracy; we selected these two measures on account of their clear group differences.

For all 28 participants from the three groups, K-means cluster analysis using note deviation yielded two distinct groups (good pitch imitation vs. poor pitch imitation). As shown in **Figure 2**, a subgroup of the controls ($n = 7$) imitated musical pitch with significantly smaller note deviations (around 1.4 semitones) than the rest of the controls and all the amusics and tone agnosics (around 2.8 semitones).

On the other hand, K-means cluster analysis using direction accuracy yielded two different distinct groups. As shown in **Figure 3**, all controls ($n = 12$), five amusics, and four tone agnosics demonstrated relatively better music direction production performance (i.e., the "good direction imitation" group, with direction accuracy around 95%), whereas three amusics and four tone agnosics fell into the other group (the "poor direction imitation" group, with direction accuracy around 50%).

These results suggest that note deviation and direction accuracy are two different measures describing musical pitch production. Measured by note deviation, five controls and all the amusics and tone agnosics demonstrated poor musical production, whereas measured by direction accuracy, none of the controls and only subgroups (about half) of the amusics and tone agnosics showed poor musical production.

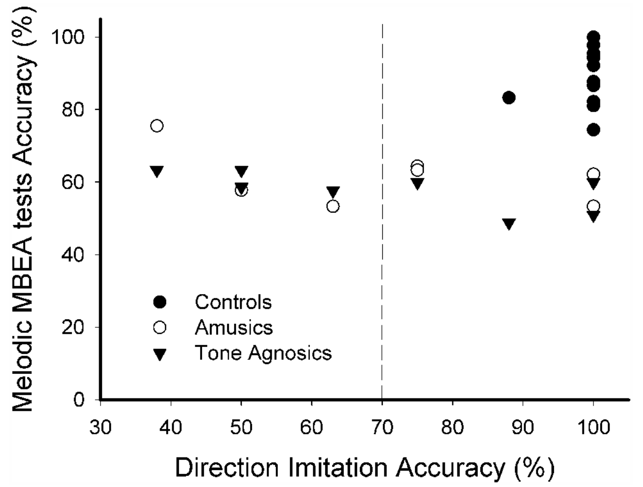


FIGURE 3 | Direction accuracy of three-note imitation across the three groups. A dashed line divides two distinct groups based on individual performances on direction accuracy of three-note imitation for the controls, the amusic, and tone agnosic groups.

DISCUSSION

Auditory perception and vocal production are closely related, although the exact nature of this link is still hotly debated (e.g., Lotto et al., 2009; Dalla Bella et al., 2011). The present study tried to determine to what extent perceptual pitch deficits across domains are related to pitch production difficulties in music and Mandarin speech. Our results showed that the amusics and tone agnosics both had musical pitch production difficulties. However, the perceptual pitch deficits across domains did not affect lexical tone production.

The current study found that the controls outperformed both the amusic and tone agnosic individuals on note deviation and music direction accuracy. Moreover, across the three groups, note deviation performance was significantly and negatively correlated with the melodic MBEA scores, suggesting an association between musical pitch perception and production. These findings corroborates previous results found among speakers of non-tonal languages (Dalla Bella et al., 2009) and further supports the notion that pitch perception and pitch production are closely coupled in the music domain (Berkowska and Dalla Bella, 2009; Dalla Bella et al., 2011) by using data from speakers of a tonal language. Furthermore, there was also a close link between lexical tone and musical pitch perception, corroborating the notion that lexical tone deficits among tone agnosics are associated with musical pitch disorders (Nan et al., 2010). However, it should be noted that according to the current results, tone agnosics were not more impaired than the amusics in musical pitch production, suggesting that lexical tone perception deficits are not necessarily detrimental to musical pitch production.

More interestingly, all participants with lexical tone perception impairments demonstrated intact lexical tone production, corroborating our previous results (Nan et al., 2010). This dissociation between perception and production of linguistic tones among amusics is also in line with previous results obtained using speakers of non-tonal languages (Hutchins and Peretz, 2012). A recent event-related potential paradigm (ERP) study reports a similar dissociation between production and perception of lexical tones for Cantonese (Law et al., 2013), suggesting that the observed independence between perception and production for lexical tones is relatively robust. This is partly in line with results from a recent study, which showed partial support for domain specific pitch processing, but at the same time a close association between song and speech imitation performance (Mantell and Pfordresher, 2013).

Nonetheless, neither the VSL model (Berkowska and Dalla Bella, 2009) nor the vocal-motor encoding theory (Hutchins and Peretz, 2012) can easily incorporate the distinct perception-production relationships for music and Mandarin speech that were observed in the current study. It is possible that pitch production in music and Mandarin speech involve independent but interactive systems, similar to the recently proposed dual routes for verbal repetition (Yoo et al., 2012). The production of lexical tones may mainly engage the acoustic–phonetic systems primarily relying on the high temporal resolution of the left hemisphere, whereas the production of musical pitches as well as cross-domain pitch perception (including both musical pitch and lexical tone perception) may mainly require the high frequency resolution of the right hemisphere (Zatorre and Belin, 2001; Luo et al., 2006), although the left inferior frontal gyrus is often implicated in lexical tone (Hsieh et al., 2001) or lexical tone and music pitch perception (Nan and Friederici, 2013) in Mandarin speakers as well. Thus, this classical view of hemispheric asymmetries in spectral and temporal processing (Zatorre and Belin, 2001) may account for how Mandarin lexical tone production can be preserved simultaneously with impaired production in musical pitch and impaired pitch perception across domains.

Moreover, the current results provide more insights on the nature of congenital amusia from the perspective of music production. Both the amusics and tone agnosics performed similarly to the controls on three-note imitation when measured by interval deviation, but these two groups were significantly impaired relative to the controls as measured by direction accuracy. This corroborates the accumulating results on music perception: compared to the controls, the amusics are not necessarily impaired in pitch discrimination thresholds (Foxton et al., 2004; Tillmann et al., 2009; Albouy et al., 2013), but they show relatively consistent difficulties in discriminating pitch direction (Foxton et al., 2004; Liu et al., 2010). Hence, the current study adds more evidence on the notion that the related deficits of congenital amusia may as well arise at a relatively higher stage of pitch processing, e.g., perceiving and producing pitch directions (for a review, see Stewart, 2011). Furthermore, the observed clear split of good and poor musical pitch production groups among amusics and tone agnosics as measured by music direction accuracy is in line with previous results on the possible existence of subgroups within the amusic population (Dalla Bella et al., 2009; Nan et al., 2010), converging on the notion that congenital amusia is indeed a complex disorder and often involves variously mixed presentations (Stewart, 2011).

It should be noted that, in the current study, all of the participants' performances in musical pitch production (except direction accuracy) were relatively lower when compared to previous similar studies (e.g., Amir et al., 2003; Watts et al., 2005; Dalla Bella et al., 2007; Pfordresher et al., 2007; Wise and Sloboda, 2008). This may be due to the effects of tasks and stimuli for the music production tests. First, despite using similar stimuli, our imitation task was more demanding compared to the tasks used in previous studies (Amir et al., 2003; Watts et al., 2005; Pfordresher et al., 2007). These previous studies usually presented each stimulus several times during the test, whereas in the present study, all stimuli were presented only once. Second and more importantly, for music production tasks, the present study used piano tones in a pitch range that was suitable for females but not for male participants. As we have more male participants ($n = 17$) than female participants ($n = 11$), this gender effect would have contributed to the overall lower musical production performance observed in the current study. However, it should be noted that, when gender was considered as an additional between-subjects factor, no significant main effect or interactions involving gender were found, probably due to the small sample size. Likewise, in lexical tone production task, female voices were used. This might also have caused male participants more difficulties in lexical tone imitation than females, although no gender effect was statistically significant.

Additionally, it is important to point out the fact that the lower production performances we observed occurred not only for the pitch dimension but also for the time dimension, in contrast to the findings of previous research (Dalla Bella et al., 2009). Based upon the same criterion, the current study found an average duration accuracy of approximately 30% across the three groups, whereas a previous study (Dalla Bella et al., 2009) with a familiar melody reported average duration accuracy more than 90% among the controls and the amusics. Nonetheless, the current data demonstrated that the amusics as well as tone agnosics were not impaired relative to the controls in the time dimension for music production, corroborating the previous study (Dalla Bella et al., 2009). Together with data from musical pitch perception (Hyde and Peretz, 2004; Nan et al., 2010), the current results further support the notion that pitch deficits in both perception and production related to congenital amusia mainly affect the frequency dimension but not the time dimension (Hyde and Peretz, 2004; Dalla Bella et al., 2009; Nan et al., 2010).

It should also be noted that a more recent study reported impaired speech and song imitation among Mandarin-speaking amusics (Liu et al., 2013), whereas our results showed that the amusics were only impaired in musical pitch production but not lexical tone production. This discrepancy of speech tone production in Mandarin-speaking amusics between our results and those of Liu et al. (2013) might be caused by different stimuli employed in these studies. The current study used speech stimuli which contained equal numbers of four lexical tones in Mandarin, whereas Liu et al. (2013) did not control for the number of lexical tones from each tone category. As shown in **Table 2** (Liu et al., 2013), among the set of selected speech stimuli used in the experiment, there were 28 level tones (tone 1) and five dipping tones (tone 3). With such a high rate of level tones (28 among 60 syllables, almost half), the speech stimuli did not well represent Mandarin which has four main lexical tones.

The observed intact lexical tone production among amusics and tone agnosics in our present study, however, might also be due to the fact that the current speech stimuli were mainly drawn from everyday materials. It is inevitable that the speech stimuli employed in the current study were more familiar to the participants than the music stimuli. Over years of experience with daily production needs, amusics and especially tone agnosics might have learned how to produce speech tones despite their perceptual impairments. But clearly the extent of their exposure and daily production pressure for music pitches is much lower. The resulted disparity of the learning processes between music and speech domains might thus also account for the intact lexical tone production but impaired music pitch production among amusics and tone agnosics relative to controls. Additionally, the timbre difference between the speech (human voices) and music stimuli (piano tones) might have played a role. As suggested by previous research (e.g., Leveque et al., 2012), piano timbre is more difficult to imitate than human voice.

Nonetheless, it should be noted that although the F0 deviations of the produced lexical tones in all the three groups were not negligible, the intelligibility of the speech words were not affected. This indicates different functional standards for pitch production in lexical tones than that in music. More interestingly, based on the current results, we might tentatively speculate that the pitch-related production skills necessary for intelligibility of speech (such as those measured by direction accuracy and interval deviation) are largely intact in the amusics and tone agnosics as tested in the current study, while the aspect that is more specific to music (for instance the one represented by note deviation) is the one that is most compromised.

CONCLUSION

Individuals with perceptual pitch deficits known as congenital amusics and tone agnosics represent unique opportunities to understand the intriguing relationships between pitch production and perception in music and language. The current study examined to what extent the perceptual pitch deficits involved in both the amusics and tone agnosics are related to pitch production difficulties in music and Mandarin speech among Mandarin speakers. For music, our results demonstrated that pitch production difficulties may be present in both the amusic and tone agnostic groups, resulting in significantly enlarged note deviation and decreased direction accuracy in these two groups relative to the controls. Moreover, tone agnosics were not more impaired than the amusics in musical pitch production, suggesting that lexical tone perception deficits are not necessarily detrimental to musical pitch production. For language, on the other hand, all three groups were able to imitate lexical tones with perfect intelligibility, suggesting that the perceptual pitch deficits across domains may coexist with intact lexical tone production. Taken together, the current results imply that the perception-production relationship for pitch among individuals with perceptual pitch deficits may be domain-dependent.

ACKNOWLEDGMENTS

This work was supported by the National Natural Science Foundation of China (30700225 and 31221003), by the Ministry of Science and Technology of China (2014CB846100), and by the 111 project (B07008). We owe our thanks to members of the music group at the State Key Laboratory of Cognitive Neuroscience and Learning for their input. We thank the two reviewers for their invaluable and insightful comments on the manuscript, and we thank the participants for their support.

REFERENCES

Albouy, P., Mattout, J., Bouet, R., Maby, E., Sanchez, G., Aguera, P. E., et al. (2013). Impaired pitch perception and memory in congenital amusia: the deficit starts in the auditory cortex. *Brain* 136, 1639–1661. doi: 10.1093/brain/awt082

Amir, O., Amir, N., and Kishon-Rabin, L. (2003). The effect of superior auditory skills on vocal accuracy. *J. Acoust. Soc. Am.* 113, 1102–1108. doi: 10.1121/1.1536632

Berkowska, M., and Dalla Bella, S. (2009). Acquired and congenital disorders of sung performance: a review. *Adv. Cogn. Psychol.* 5, 69–83. doi: 10.2478/v10053-008-0068-2

Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glot Int.* 5, 341–345.

Chase, T. N., Fedio, P., Foster, N. L., Brooks, R., Di, C. G., and Mansi, L. (1984). Wechsler Adult Intelligence Scale performance. Cortical localization by fluorodeoxyglucose F 18-positron emission tomography. *Arch. Neurol.* 41, 1244–1247. doi: 10.1001/archneur.1984.04050230026012

Dalla Bella, S., Berkowska, M., and Sowinski, J. (2011). Disorders of pitch production in tone deafness. *Front. Psychol.* 2:164. doi: 10.3389/fpsyg.2011.00164

Dalla Bella, S., Giguere, J. F., and Peretz, I. (2007). Singing proficiency in the general population. *J. Acoust. Soc. Am.* 121, 1182–1189. doi: 10.1121/1.2427111

Dalla Bella, S., Giguere, J. F., and Peretz, I. (2009). Singing in congenital amusia. *J. Acoust. Soc. Am.* 126, 414–424. doi: 10.1121/1.3132504

Foxton, J. M., Dean, J. L., Gee, R., Peretz, I., and Griffiths, T. D. (2004). Characterization of deficits in pitch perception underlying 'tone deafness'. *Brain* 127, 801–810. doi: 10.1093/brain/awh105

Gong, Y. X. (1992). *Wechsler Adult Intelligence Scale-Revised in China Version.* Changsha: Hunan Medical College.

Gosselin, N., Jolicoeur, P., and Peretz, I. (2009). Impaired memory for pitch in congenital amusia. *Ann. N. Y. Acad. Sci.* 1169, 270–272. doi: 10.1111/j.1749-6632.2009.04762.x

Han, D., Zhou, N., Li, Y., Chen, X., Zhao, X., and Xu, L. (2007). Tone production of Mandarin Chinese speaking children with cochlear implants. *Int. J. Pediatr. Otorhinolaryngol.* 71, 875–880. doi: 10.1016/j.ijporl.2007.02.008

Hsieh, L., Gandour, J., Wong, D., and Hutchins, G. D. (2001). Functional heterogeneity of inferior frontal gyrus is shaped by linguistic experience. *Brain Lang.* 76, 227–252. doi: 10.1006/brln.2000.2382

Hutchins, S., and Peretz, I. (2012). Amusics can imitate what they cannot discriminate. *Brain Lang.* 123, 234–239. doi: 10.1016/j.bandl.2012.09.011

Hutchins, S., Zarate, J. M., Zatorre, R. J., and Peretz, I. (2010). An acoustical study of vocal pitch matching in congenital amusia. *J. Acoust. Soc. Am.* 127, 504–512. doi: 10.1121/1.3270391

Hyde, K. L., and Peretz, I. (2003). "Out-of-pitch" but still "in-time". An auditory psychophysical study in congenital amusic adults. *Ann. N. Y. Acad. Sci.* 999, 173–176. doi: 10.1196/annals.1284.023

Hyde, K. L., and Peretz, I. (2004). Brains that are out of tune but in time. *Psychol. Sci.* 15, 356–360. doi: 10.1111/j.0956-7976.2004.00683.x

Hyde, K. L., Zatorre, R. J., and Peretz, I. (2011). Functional MRI evidence of an abnormal neural network for pitch processing in congenital amusia. *Cereb. Cortex* 21, 292–299. doi: 10.1093/cercor/bhq094

Jiang, C., Hamm, J. P., Lim, V. K., Kirk, I. J., and Yang, Y. (2010). Processing melodic contour and speech intonation in congenital amusics with Mandarin Chinese. *Neuropsychologia* 48, 2630–2639. doi: 10.1016/j.neuropsychologia.2010.05.009

Kazui, H., Yoshida, T., Takaya, M., Sugiyama, H., Yamamoto, D., Kito, Y., et al. (2011). Different characteristics of cognitive impairment in elderly schizophrenia and Alzheimer's disease in the mild cognitive impairment stage. *Dement. Geriatr. Cogn. Dis. Extra* 1, 20–30. doi: 10.1159/000323561

Law, S. P., Fung, R., and Kung, C. (2013). An ERP study of good production vis-à-vis poor perception of tones in Cantonese: implications for top-down speech processing. *PLoS ONE* 8:e54396. doi: 10.1371/journal.pone.0054396

Leveque, Y., Giovanni, A., and Schon, D. (2012). Pitch-matching in poor singers: human model advantage. *J. Voice* 26, 293–298. doi: 10.1016/j.jvoice.2011.04.001

Liu, F., Jiang, C., Pfordresher, P. Q., Mantell, J. T., Xu, Y., Yang, Y., et al. (2013). Individuals with congenital amusia imitate pitches more accurately in singing than in speaking: implications for music and language processing. *Atten. Percept. Psychophys.* 1783–1798. doi: 10.3758/s13414-013-0506-1

Liu, F., Jiang, C., Thompson, W. F., Xu, Y., Yang, Y., and Stewart, L. (2012). The mechanism of speech processing in congenital amusia: evidence from Mandarin speakers. *PLoS ONE* 7:e30374. doi: 10.1371/journal.pone.0030374

Liu, F., Patel, A. D., Fourcin, A., and Stewart, L. (2010). Intonation processing in congenital amusia: discrimination, identification and imitation. *Brain* 133, 1682–1693. doi: 10.1093/brain/awq089

Lotto, A. J., Hickok, G. S., and Holt, L. L. (2009). Reflections on mirror neurons and speech perception. *Trends Cogn. Sci.* 13, 110–114. doi: 10.1016/j.tics.2008.11.008

Loui, P., Guenther, F. H., Mathys, C., and Schlaug, G. (2008). Action-perception mismatch in tone-deafness. *Curr. Biol.* 18, R331–R332. doi: 10.1016/j.cub.2008.02.045

Luo, H., Ni, J. T., Li, Z. H., Li, X. O., Zhang, D. R., Zeng, F. G., et al. (2006). Opposite patterns of hemisphere dominance for early auditory processing of lexical tones and consonants. *Proc. Natl. Acad. Sci. U.S.A.* 103, 19558–19563. doi: 10.1073/pnas.0607065104

Mantell, J. T., and Pfordresher, P. Q. (2013). Vocal imitation of song and speech. *Cognition* 127, 177–202. doi: 10.1016/j.cognition.2012.12.008

Marin, M. M., Gingras, B., and Stewart, L. (2012). Perception of musical timbre in congenital amusia: categorization, discrimination and short-term memory. *Neuropsychologia* 50, 367–378. doi: 10.1016/j.neuropsychologia.2011.12.006

Moreau, P., Jolicoeur, P., and Peretz, I. (2009). Automatic brain responses to pitch changes in congenital amusia. *Ann. N. Y. Acad. Sci.* 1169, 191–194. doi: 10.1111/j.1749-6632.2009.04775.x

Nan, Y., and Friederici, A. D. (2013). Differential roles of right temporal cortex and broca's area in pitch processing: evidence from music and Mandarin. *Hum. Brain Mapp.* 34, 2045–2054. doi: 10.1002/hbm.22046

Nan, Y., Sun, Y., and Peretz, I. (2010). Congenital amusia in speakers of a tone language: association with lexical tone agnosia. *Brain* 133, 2635–2642. doi: 10.1093/brain/awq178

Oldfield, R. C. (1971). The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia* 9, 97–113. doi: 10.1016/0028-3932(71)90067-4

Peng, S. C., Tomblin, J. B., Cheung, H., Lin, Y. S., and Wang, L. S. (2004). Perception and production of Mandarin tones in prelingually deaf children with cochlear implants. *Ear Hear.* 25, 251–264. doi: 10.1097/01.AUD.0000130797.73809.40

Peretz, I. (2001). Brain specialization for music. New evidence from congenital amusia. *Ann. N. Y. Acad. Sci.* 930, 153–165. doi: 10.1111/j.1749-6632.2001.tb05731.x

Peretz, I., Ayotte, J., Zatorre, R. J., Mehler, J., Ahad, P., Penhune, V. B., et al. (2002). Congenital amusia: a disorder of fine-grained pitch discrimination. *Neuron* 33, 185–191. doi: 10.1016/S0896-6273(01)00580-3

Peretz, I., Brattico, E., Jarvenpaa, M., and Tervaniemi, M. (2009). The amusic brain: in tune, out of key, and unaware. *Brain* 132, 1277–1286. doi: 10.1093/brain/awp055

Peretz, I., Champod, A. S., and Hyde, K. (2003). Varieties of musical disorders. The Montreal Battery of Evaluation of Amusia. *Ann. N. Y. Acad. Sci.* 999, 58–75. doi: 10.1196/annals.1284.006

Pfeuty, M., and Peretz, I. (2010). Abnormal pitch – time interference in congenital amusia: evidence from an implicit test. *Atten. Percept. Psychophys.* 72, 763–774. doi: 10.3758/APP.72.3.763

Pfordresher, P. Q., Brown, S., Meier, K. M., Belyk, M., and Liotti, M. (2010). Imprecise singing is widespread. *J. Acoust. Soc. Am.* 128, 2182–2190. doi: 10.1121/1.3478782

Pfordresher, P. Q., Palmer, C., and Jungers, M. K. (2007). Speed, accuracy, and serial order in sequence production. *Cogn. Sci.* 31, 63–98. doi: 10.1080/03640210709336985

Stern, R. A., and Prohaska, M. L. (1996). "Neuropsychological evaluation of executive functioning," in *Academic Psychiatric Press Review of Psychiatry*, eds L. J. Dickstein, M. B. Riba, and J. M. Oldham (Washington: American Psychiatric Press), 243–266.

Stewart, L. (2011). Characterizing congenital amusia. *Q. J. Exp. Psychol. (Hove)* 64, 625–638. doi: 10.1080/17470218.2011.552730

Thompson, W. F., Marin, M. M., and Stewart, L. (2012). Reduced sensitivity to emotional prosody in congenital amusia rekindles the musical protolanguage hypothesis. *Proc. Natl. Acad. Sci. U.S.A.* 109, 19027–19032. doi: 10.1073/pnas.1210344109

Tillmann, B., Burnham, D., Nguyen, S., Grimault, N., Gosselin, N., and Peretz, I. (2011). Congenital amusia (or tone-deafness) interferes with pitch processing in tone languages. *Front. Psychol.* 2:120. doi: 10.3389/fpsyg.2011.00120

Tillmann, B., Schulze, K., and Foxton, J. M. (2009). Congenital amusia: a short-term memory deficit for non-verbal, but not verbal sounds. *Brain Cogn.* 71, 259–264. doi: 10.1016/j.bandc.2009.08.003

Watts, C., Moore, R., and McCaghren, K. (2005). The relationship between vocal pitch-matching skills and pitch discrimination skills in untrained accurate and inaccurate singers. *J. Voice* 19, 534–543. doi: 10.1016/j.jvoice.2004.09.001

Wechsler, D. (1981). *Wechsler Adult Intelligence Scale-Revised.* New York: Psychological Corp.

Williamson, V. J., McDonald, C., Deutsch, D., Griffiths, T. D., and Stewart, L. (2010). Faster decline of pitch memory over time in congenital amusia. *Adv. Cogn. Psychol.* 6, 15–22. doi: 10.2478/v10053-008-0073-5

Williamson, V. J., and Stewart, L. (2010). Memory for pitch in congenital amusia: beyond a fine-grained pitch discrimination problem. *Memory* 18, 657–669. doi: 10.1080/09658211.2010.501339

Wise, K., and Sloboda, J. A. (2008). Establishing an empirical profile of self-defined 'tone deafness': perception, singing performance, and self-assessment. *Music Sci.* 12, 3–23. doi: 10.1177/102986490801200102

Xu, L., Chen, X., Lu, H., Zhou, N., Wang, S., Liu, Q., et al. (2011). Tone perception and production in pediatric cochlear implants users. *Acta Otolaryngol.* 131, 395–398. doi: 10.3109/00016489.2010.536993

Yoo, S., Chung, J. Y., Jeon, H. A., Lee, K. M., Kim, Y. B., and Cho, Z. H. (2012). Dual routes for verbal repetition: articulation-based and acoustic-phonetic codes for pseudoword and word repetition, respectively. *Brain Lang.* 122, 1–10. doi: 10.1016/j.bandl.2012.04.011

Zatorre, R. J., and Belin, P. (2001). Spectral and temporal processing in human auditory cortex. *Cereb. Cortex* 11, 946–953. doi: 10.1093/cercor/11.10.946

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 27 May 2013; accepted: 31 December 2013; published online: 16 January 2014.

Citation: Yang W-x, Feng J, Huang W-t, Zhang C-x and Nan Y (2014) Perceptual pitch deficits coexist with pitch production difficulties in music but not Mandarin speech. Front. Psychol. ***4****:1024. doi: 10.3389/fpsyg.2013.01024*

This article was submitted to Cognitive Science, a section of the journal Frontiers in Psychology.

frontiers in
PSYCHOLOGY

ORIGINAL RESEARCH ARTICLE
published: 01 November 2013
doi: 10.3389/fpsyg.2013.00811

Cross-modal signatures in maternal speech and singing

Sandra E. Trehub*, Judy Plantinga, Jelena Brcic and Magda Nowicki

Music Development Laboratory, Department of Psychology, University of Toronto Mississauga, Mississauga, ON, Canada

Edited by:
Bruno Gingras, University of Vienna, Austria

Reviewed by:
Sean Hutchins, Rotman Research Institute at Baycrest Hospital, Canada
Claudia Kubicek, University of Giessen, Germany
Emily Zimmerman, Northeastern University, USA

****Correspondence:***
Sandra E. Trehub, Department of Psychology, University of Toronto Mississauga, 3359 Mississauga Road North, Mississauga, ON L5L 1C6, Canada
e-mail: sandra.trehub@utoronto.ca

We explored the possibility of a unique cross-modal signature in maternal speech and singing that enables adults and infants to link unfamiliar speaking or singing voices with subsequently viewed silent videos of the talkers or singers. In Experiment 1, adults listened to 30-s excerpts of speech followed by successively presented 7-s silent video clips, one from the previously heard speaker (different speech content) and the other from a different speaker. They successfully identified the previously heard speaker. In Experiment 2, adults heard comparable excerpts of singing followed by silent video clips from the previously heard singer (different song) and another singer. They failed to identify the previously heard singer. In Experiment 3, the videos of talkers and singers were blurred to obscure mouth movements. Adults successfully identified the talkers and they also identified the singers from videos of different portions of the song previously heard. In Experiment 4, 6— to 8-month-old infants listened to 30-s excerpts of the same maternal speech or singing followed by exposure to the silent videos on alternating trials. They looked longer at the silent videos of previously heard talkers and singers. The findings confirm the individuality of maternal speech and singing performance as well as adults' and infants' ability to discern the unique cross-modal signatures. The cues that enable cross-modal matching of talker and singer identity remain to be determined.

Keywords: speech, singing, infants, adults, cross-modal, identification

INTRODUCTION

Mothers around the world talk and sing to their pre-verbal infants (Trehub et al., 1993; Trehub and Trainor, 1998), presumably to gain their attention, modulate their arousal, share feelings, and strengthen dyadic ties (Fernald, 1992; Shenfield et al., 2003; Trehub et al., 2010). Maternal or infant-directed (ID) speech is generally regarded as a distinct speech register or genre (Fernald, 1992; Papoušek, 1994) although some consider it to be little more than highly expressive speech—happier, more loving, and more comforting than typical adult-directed (AD) speech (e.g., Kitamura and Burnham, 1998; Trainor et al., 2000; Singh et al., 2002). Indeed, the characteristically happy manner of North American ID speech shares some features with vocal elation or high-arousal happiness in AD speech (Banse and Scherer, 1996).

Research on ID speech has focused primarily on its acoustic features across languages and cultures (e.g., Ferguson, 1964; Grieser and Kuhl, 1988; Fernald et al., 1989) and secondarily on its consequences for infant attention, affect, and learning (e.g., Fernald, 1985; Werker and McLeod, 1989; Papoušek et al., 1990; Thiessen et al., 2005). The exaggerated pitch contours, rhythmicity, and repetitiveness of ID speech give it a notably musical flavor (Fernald, 1989; Trehub, 2009). In fact, the acoustic features of ID speech are more similar to those of ID song than to AD speech (Corbeil et al., 2013), leading some scholars to characterize ID speech as a form of music (Brandt et al., 2012). Differences in syntactic and semantic aspects of ID and AD speech, although substantial (e.g., Ferguson, 1964; Papoušek, 1994), presumably have less impact on pre-verbal listeners than do expressive aspects of such speech. In fact, there is evidence that the expressivity of ID speech to 12-month-olds is somewhat attenuated as compared with speech to younger infants (Stern et al., 1983; Kitamura and Burnham, 2003).

With attention focused largely on common features and cultural variations of ID speech, there has been relatively little interest in questions of individuality. Bergeson and Trehub (2007) found, however, that mothers used individually distinctive melodies, or *signature tunes*, in their speech to infants. In two recording sessions separated by a week or so, they found that mothers repeatedly used a small set of individually distinctive tunes (i.e., specific interval sequences that were unrelated to musical scales), varying the verbal content that accompanied those tunes. Such tunes—their pitch patterns and rhythms—could provide important cues to speaker identity. Just as communicative intentions are more transparent in ID than in AD speech (Fernald, 1989), even across disparate cultures (Bryant and Barrett, 2007), prosodic cues to identity may be more transparent in ID than in AD speech. It is unclear, however, whether phonetic or articulatory cues (i.e., talkers' idiolect) are individually distinctive in ID speech, as they are in AD speech (Fellowes et al., 1997; Sheffert et al., 2002).

In interactions with infants, mothers also use exaggerated facial (Chong et al., 2003) and body gestures (Brand et al., 2002; Brand and Shallcross, 2008) that feature greater repetitiveness and range of motion than AD gestures. To date, however, there has been no attempt to ascertain whether these visual aspects of ID speech are individually distinctive. Adults recognize familiar individuals from facial motion (Hill and Johnston, 2001), which provides visual correlates of prosody and articulation, and from point-light displays derived from the teeth, tongue, and face of talkers (Rosenblum et al., 2007), which provide visual cues to idiolect. Adults perform modestly but above chance levels in a delayed matching-to-sample task involving unfamiliar voices and silent videos from the same or different utterances (Kamachi et al., 2003; Lachs and Pisoni, 2004; Lander et al., 2007). In one condition, Kamachi et al. (2003) and Lander et al. (2007) presented adults with a scripted utterance followed by successively presented silent videos, one from the previously heard speaker articulating the same utterance (or a different scripted utterance in another condition) and the other from a different speaker. Performance was somewhat better for cross-modal matching of the same utterances than for different utterances. Performance was equivalent, however, for participants who experienced the stimuli in reverse order, for example, a silent video followed by two successively presented utterances. The results imply the presence of signature features in the audible and visible aspects of speech, perhaps based on rhythmic structures or expressiveness (Lander et al., 2007).

In previous research, the importance of temporal cues was indicated by adults' inability to match audible and visible aspects of speech when the stimuli were played backward rather than forward (Kamachi et al., 2003; Lachs and Pisoni, 2004). The manner or style of speech seems to make a critical contribution to performance. For example, changing the manner from statement to question form, from conversational style to clear (i.e., carefully articulated) speech, or from conversational to rushed casual speech significantly reduces identification accuracy (Lander et al., 2007). By contrast, electronic speeding or slowing of speech does not impair the accuracy of cross-modal matching (Lander et al., 2007), which implies that relational rather than absolute timing cues are implicated.

The goal of the present research was to ascertain whether auditory and visible aspects of maternal speech and song have a common signature that is perceptible to adults who are unfamiliar with the talkers and singers. The perceptibility of that signature would enable adults, perhaps even infants, to match auditory and visual components of maternal speech and song in the context of a delayed matching-to-sample task. As is the case for ID speech, research on ID song has focused largely on its acoustic features (e.g., Rock et al., 1999; Nakata and Trehub, 2011) and its consequences for infant attention (Trainor, 1996; Tsang and Conrad, 2010; Corbeil et al., 2013), arousal (Shenfield et al., 2003) and learning (Volkova et al., 2006; Lebedeva and Kuhl, 2010). Although mothers perform the same ID songs at nearly identical pitch level and tempo on different occasions (Bergeson and Trehub, 2002), it is unclear whether their performances of different songs exhibit comparable stability and uniqueness. In any case, pitch level and tempo are not considered reliable cues to the identity of speakers (Kunzel, 1989; Lander et al., 2007).

EXPERIMENT 1

In the present experiment, we sought to ascertain whether adults could link person-specific auditory and visual components of ID speech in a delayed matching-to-sample task. The procedure was modeled on that of Kamachi et al. (2003) who found that adults performed no differently when visual images were matched to previously heard voices or voices were matched to previously seen visual images. For our purposes, adults on each trial were exposed to a 30-s sample of natural ID speech from one of two unfamiliar women followed by two silent videos of speech presented sequentially, one from the previously heard woman, the second from the other woman. Their task was to identify which video corresponded to the previously heard speaker. The stimuli in previous face-voice matching studies featured the same scripted words or utterances for all speakers (e.g., Kamachi et al., 2003; Lachs and Pisoni, 2004; Lander et al., 2007) in contrast to the present experiment, which involved maternal speech extracted from natural interactions with infants. As a result, message content differed from one mother to another and for different parts of the discourse of the same mother. In principle, adults would be capable of lipreading some of the verbal content from silently articulating mothers, which necessitated the use of different speech passages from each mother at exposure and test phases. In other words, the verbal content differed from exposure to test and between the two test stimuli (familiar and unfamiliar women).

METHOD

The Office of Research Ethics at the University of Toronto approved all research reported here.

Participants

The participants were 44 young adults (24 women, 20 men) who were enrolled in an undergraduate course in introductory psychology. All were healthy and free of hearing loss, according to self-report.

Apparatus

Testing took place in a double-walled sound-attenuating booth (Industrial Acoustics) with two Audiological GSI loudspeakers located to the left and right of the seated participant at a 45-degree angle. Stimulus presentation and response recording were controlled by customized software (Real Basic) on a Windows workstation and amplifier (Harmon Kardon 3380) located outside the booth. Visual stimuli were presented on a monitor (Dell LCD, 33.5 × 26.5 cm) directly in front of the participants (at a distance of ∼1 m), who entered their responses on a hand-held keypad (Targus) connected to the computer.

Stimuli

Audio stimuli consisted of 30-s excerpts from previously recorded QuickTime videos (Sony 360X recorder) of mothers talking to their 11- to 12-month-old infants. Video stimuli, which filled the entire screen, were silent 7-s clips from different portions of the original videos (head and shoulders view of mother). Four pairs of mothers were selected from a larger set to minimize within-pair differences in physical appearance (e.g., race, stature, hair style, clothing).

Procedure

Participants were tested individually in a delayed matching-to-sample task. Before each of the four test trials, they were instructed to listen carefully to the speech excerpt and then to watch the two silent videos in succession. After the second video, static images of the two women from the videos appeared side by side on the monitor, and participants were required to judge which woman had been heard previously. A schematic view of the procedure is presented in **Figure 1**. Participants entered their

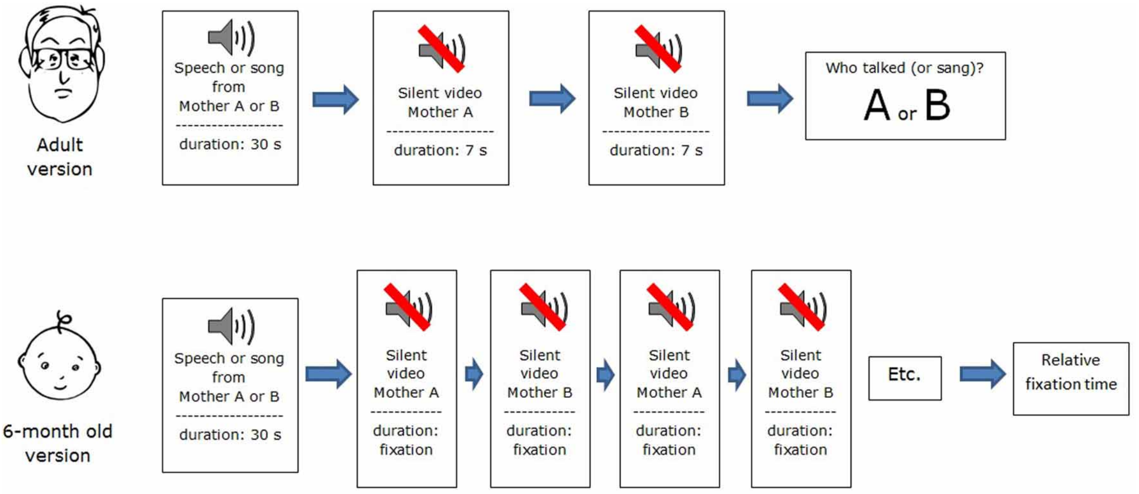


FIGURE 1 | Flow chart depicting adult and infant versions of the procedure.

choice on a hand-held keypad, which they also used to control the onset of trials. Half of the participants heard the audio excerpts of one woman from each pair and half heard the audio excerpts of the other woman. Matching and non-matching videos were presented in random order.

RESULTS AND DISCUSSION

As can be seen in **Figure 2** (solid bar), adults matched person-specific auditory and visual aspects of speech imperfectly ($M = 0.70$, $SD = 0.24$) but well above the proportion correct expected by chance (0.50), $t_{(43)} = 19.292$, $p < 0.001$. Moreover, women did not perform better than men, and performance did not differ across stimulus pairs. Adults' success at identifying previously heard maternal speakers on the basis of dynamic visual depictions of those speakers confirms the presence of individually distinctive cross-modal features in maternal speech. The nature of those features remains to be determined. Although two pairs of mothers exhibited differences in speaking rate ($M = 2.77$ vs. 2.03 and 2.90 vs. 1.57 syllables per sec), the other two pairs exhibited little difference ($M = 2.63$ vs. 2.67 and 2:63 vs. 2.60 syllables per sec). Nevertheless, participants performed no better on pairs that differed in speaking rate than those that did not, indicating that speech rate could not account for successful matching in this delayed matching-to-sample task.

The present findings add to the growing literature on adults' perception of face-voice relations in speech (Kamachi et al., 2003; Lachs and Pisoni, 2004; Munhall and Buchan, 2004; Rosenblum et al., 2006; Lander et al., 2007). They are consistent with the view that aspects of speech manner, independent of verbal content and modality, are person-specific. The unique contribution of the present experiment is its focus on ID speech and the use of speech from natural interactions rather than scripted portrayals. Despite the fact that ID speech to pre-verbal infants has many common features within and across cultures (Ferguson, 1964; Grieser and Kuhl, 1988; Fernald et al., 1989), it retains individually distinctive acoustic features that have perceptible visual correlates.

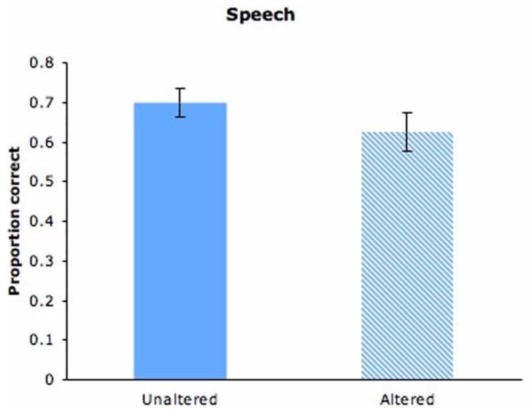


FIGURE 2 | Adults' proportion of correct responses for maternal speech with unaltered videos (solid bar) or altered videos (hatched bar). Error bars are standard errors.

EXPERIMENT 2

Our goal here was to ascertain whether adults could link person-specific auditory and visual components of ID singing in the delayed matching-to-sample task of Experiment 1. It is clear that visual features of sung performances carry music-related information. For example, singers provide cues to the magnitude of isolated intervals (i.e., two successive notes) by their facial and head movements (Thompson et al., 2010). Listeners' judgment of the affective connotation of such intervals is influenced by singers' facial expression (Thompson et al., 2008). To date, however, no study has investigated cross-modal identification of unfamiliar singers. On each trial of the present study, adults were exposed to a 30-s excerpt from an ID song performed by one of two unfamiliar women followed by two silent videos of a different song, presented one after the other. One silent video was from the previously heard singer, the other from the unheard singer. Their task was to identify which video corresponded to the previously heard singer.

METHOD

Participants

The participants were 20 young adults (14 women, 6 men), mostly undergraduates. All were healthy and free of hearing loss, according to self-report.

Apparatus

The apparatus was the same as that in Experiment 1.

Stimuli

Singing excerpts from four pairs of mothers were roughly 30 s in duration and were drawn from maternal interactions with infants. The pairs were selected to minimize gross differences in appearance. Silent video excerpts from each mother were from different songs to preclude the use of lipreading cues to song identity. Because mothers sang well-known nursery songs and different mothers sang different songs, song identity and therefore singer identity could have been obvious from visual features alone.

Procedure

The procedure was identical to that of Experiment 1 except for the use of maternal singing rather than speech (see **Figure 1**).

Results and discussion

Adults' selection of the matching videos ($M = 0.5$, $SD = 0.60$) was at chance levels (see **Figure 3**, solid bar), indicating that different maternal songs did not provide a common audiovisual signature, as was the case for maternal speech in Experiment 1. Previous research revealed that altering the manner of speech (e.g., statement to question; conversational speech to clear speech) between auditory familiarization and visual test impaired adults' performance on the delayed matching-to-sample task (Lander et al., 2007). When singing to infants, mothers may alter their performing style across songs to highlight the distinctiveness of each song or their own expressive intentions. It is possible, however, that cross-modal correspondences in maternal singing would be evident in the context of specific songs.

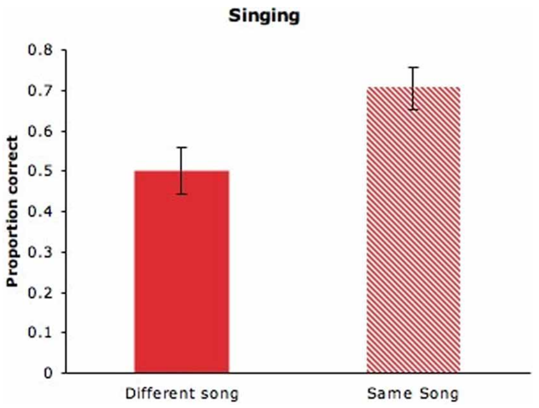


FIGURE 3 | Adults' proportion of correct responses for maternal singing with unaltered videos from different songs (solid bar) or altered videos from the same song (hatched bar). Error bars are standard errors.

EXPERIMENT 3

Adults successfully matched the speech of specific mothers to subsequent silent depictions of different utterances (Experiment 1). Interestingly, they failed to do so with audible and visible (silent) excerpts from different songs. Because the auditory and visual excerpts of speech and singing differed from exposure to test, correct person identification could not be achieved by relating the heard message to the lipread content. Prosody is known to contribute to person identification (Lander et al., 2007), as does the idiosyncratic manner of articulation or idiolect (Fellowes et al., 1997; Lachs and Pisoni, 2004) in auditory, visual, and audiovisual contexts. Prosodic and articulation features were available to participants in Experiments 1 and 2 and to the participants in previous studies of cross-modal identification (Kamachi et al., 2003; Lachs and Pisoni, 2004; Lander et al., 2007). In the present experiment, we asked whether adults could link person-specific auditory and visual components of ID speech and singing with mouth movements occluded. With lipreading cues eliminated, it was possible to examine adults' ability to link auditory and visual features from different portions of the same song rather than different songs (Experiment 2).

METHOD

Participants

The participants were 28 young adults (20 women, 8 men), mainly undergraduates, who were healthy and free from hearing loss, according to self-report.

Apparatus and stimuli

The apparatus was as described in Experiment 1. The audio excerpts of maternal speech were identical to those used in Experiment 1. The video excerpts were also the same except that Adobe Premiere Pro software was used to blur the mouth region of each speaker frame by frame. The audio excerpts of maternal singing were those used in Experiment 2. The video excerpts differed, however, in that they were selected from different portions of the same song. Adobe Premiere Pro software was used in a comparable manner to blur the mouth region of each singer.

Procedure

Participants were tested individually and in the same manner as in Experiments 1 and 2. Speech and singing trials were presented in blocks, and trials within blocks were randomized for each participant. On each trial, matching and non-matching video excerpts (i.e., same or different person) were presented in random order. The first trial block (speech or singing) and the first stimulus within blocks were counterbalanced across participants. Each participant completed eight test trials (i.e., audio excerpts from four different speakers and four different singers).

Results and discussion

As can be seen in **Figures 2** and **3** (hatched bars), adults succeeded in matching the altered video to audio samples of speech

($M = 0.63$, $SD = 0.25$), $t_{(27)} = 2.646$, $p = 0.013$, and singing ($M = 0.71$, $SD = 0.27$), $t_{(27)} = 3.986$, $p < 0.001$, and performance did not differ across speech and singing, $F_{(1,\ 26)} = 0.090$, n.s. In other words, adults successfully identified the previously heard speaker and singer on the basis of dynamic visual cues. The absence of cues from the mouth region did not significantly impair adults' ability to identify the speaker, as revealed by comparisons between the present speech condition and that of Experiment 1, $F_{(1,\ 69)} = 1.344$, n.s. It is likely, then, that prosodic cues and visual correlates of those cues were largely responsible for adults' success on this task. As can be seen in **Figure 4**, which displays the number of individuals who obtained scores of 0–4 on speaking and singing tasks, there was considerable variation in performance. One might expect individuals who perform well on speaker identification to perform well on singer identification, but performance on speaking and singing blocks was uncorrelated, $r_{(26)} = -0.017$, $p = 0.932$.

Recall that adults in Experiment 2 failed to identify the singers from video portions of different songs. Adults' performance in the present experiment on auditory and visual excerpts from the same song significantly exceeded their performance in Experiment 2 involving visual excerpts from different songs, $F_{(1,\ 46)} = 6.949$, $p < 0.01$. Unlike professional singers, mothers and other occasional singers may not have a uniform singing style, resulting in potential variations in style or manner across songs. For mothers, in particular, song performances may have different expressive intentions, for example, attention capture in some instances (e.g., *If You're Happy and You Know It*) and attention maintenance (e.g., *Twinkle, Twinkle*) or soothing (e.g., lullabies) in others. In any case, adults' ability to match audible to visible features from different portions of the same song confirms the presence of cross-modal cues to identity.

Lander et al. (2007) speculate that global aspects of expressiveness rather than single acoustic features underlie cross-modal matching in speech, but they did not attempt to quantify gradations in expressiveness. In a supplementary experiment, we had 15 undergraduates rate individual audio and silent video (unblurred) excerpts from each mother on a scale from 1 or neutral to 5 or very expressive/animated. Mean ratings of expressiveness for the four pairs of talking mothers and the four pairs of singing mothers (same song) are shown in **Table 1**. Although variations in rated expressiveness were evident across mothers, higher ratings of vocal expressiveness were not reliably associated with higher ratings of visual expressiveness. In other words, a mother who spoke or sang more expressively than her paired counterpart did not appear to be more visually expressive than the other mother.

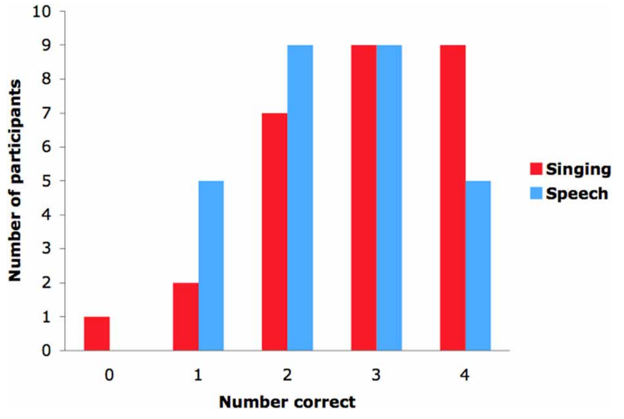


FIGURE 4 | Number of adults who obtained scores of 0–4 correct on the speech and singing tasks in Experiment 3.

EXPERIMENT 4

The findings of Experiments 1 and 3 confirmed the presence of unspecified cues to identity in auditory and visual aspects of maternal speech and singing. Recall that discernible cues to identity were found only within but not across songs. In the present experiment we investigated infants' ability to make use of cross-modal cues to identity.

In the early postnatal period, infants differentiate their mother's face from that of a stranger on the basis of static or dynamic images (Sai and Bushnell, 1988). They also differentiate the mother's voice from that of a stranger (DeCasper and Fifer, 1980). At 8 but not 4 months of age, they match auditory and visual cues to gender (Patterson and Werker, 2002), presumably on the basis of acquired knowledge of intermodal correspondences. They integrate emotional information from the face and voice, as indicated by ERP responses to simultaneously presented faces and voices (happy or angry) that are emotionally incongruent (Grossmann et al., 2006). The aforementioned unimodal and intermodal discriminations depend on learning. Nevertheless, infants perceive some cross-modal correspondences that may be independent of learning, arising from as yet unspecified amodal cues. For example, 4- to 5-month-old infants look longer at one of two simultaneously presented visual articulatory displays that matches a repeating vowel sound (/a/ or /i/) presented simultaneously and synchronously (Kuhl and Meltzoff,

Table 1 | Adults' mean expressiveness ratings (and standard deviations) of audio and video excerpts from each mother on a 5-point scale (1 = neutral, 5 = highly animated).

Talking pairs	Audio Mom A	Video Mom A	Audio Mom B	Video Mom B
1	4.8 (0.56)	3.33 (1.23)	2.47 (1.13)	1.93 (0.70)
2	2.68 (0.98)	1.07 (0.26)	1.87 (0.83)	3.23 (0.90)
3	3.93 (1.16)	1.43 (0.50)	2.27 (0.88)	1.77 (0.62)
4	3.40 (1.06)	1.20 (0.41)	4.00 (1.00)	3.50 (1.05)
Singing pairs	**Audio Mom A**	**Video Mom A**	**Audio Mom B**	**Video Mom B**
1	3.93 (0.80)	4.17 (0.79)	3.93 (0.96)	3.00 (1.25)
2	2.80 (0.86)	3.27 (0.96)	3.47 (0.64)	3.20 (0.78)
3	4.33 (0.90)	3.63 (0.81)	2.67 (0.72)	2.93 (0.70)
4	3.73 (0.96)	3.67 (0.98)	2.87 (0.83)	3.07 (0.80)

Ratings of speech are presented in the upper section and ratings of singing in the lower section. Columns indicate ratings for different pairs of mothers (1–4) and rows indicate ratings for each pair (Mom A, Mom B). Ratings of talking are for the unaltered excerpts, as in Experiments 1 and 4. Ratings of singing are for unaltered excerpts of the same song, as in Experiment 4.

1982; Patterson and Werker, 1999, 2002). Infants seem to perceive some connection between mouth shape and vowel category, perhaps because of redundant amodal cues (Bahrick et al., 2004). Remarkably, 6-month-old infants also perceive the links between syllables that they hear (/ba/ or /va/) and dynamic visual images presented before and after the auditory stimuli (Pons et al., 2009). By 10–12 months of age, they link the sounds of their native language to dynamic images of that language, indicating their perception of amodal cues to the identity of a familiar language (Lewkowicz and Pons, 2013).

The focus of the present experiment was on audiovisual cues to identity, as in Experiments 1–3. In contrast to previous cross-modal matching tasks with infants, which usually featured simultaneous visual displays (Kuhl and Meltzoff, 1982; Patterson and Werker, 1999, 2002; Pons et al., 2009; Lewkowicz and Pons, 2013), we used sequential presentation of visual stimuli. The procedure was in line with Experiments 1 and 2, with adjustments to accommodate the needs of 6- to 8-month-old participants. On the basis of individual identification across species (Ghazanfar et al., 2007; Pollard and Blumstein, 2011), one might expect some cues to identity—auditory, visual, and audiovisual—to be extracted automatically and effortlessly, even in early life.

Infants were tested with the familiarization-preference procedure (e.g., Hannon and Trehub, 2005; Plantinga and Trehub, 2013), which was modified to accommodate cross-modal matching. The procedure was similar, in some respects, to the intermodal matching procedure used by Pons et al. (2009), such as auditory stimuli presented separately from visual stimuli, but it differed in several respects including the sequential presentation of visual stimuli. First, infants were exposed to 30-s samples of ID speech or singing after which they received silent videos of the previously heard speaker or singer and another speaker or singer on alternating trials (see **Figure 1**). In other words, they saw the silent video of the previously heard speaker on every other trial and the silent video of the unheard speaker on intervening trials. If infants perceived amodal cues to identity in the auditory and visual excerpts, they should exhibit differential attention to the video excerpts. For example, they could look longer at videos of the familiar or previously heard speaker or singer or at the videos of the unheard speaker or singer. Infants' success, if evident, would stem from implicit memory for amodal cues, in contrast to adults, who might have explicit memory for person-specific features. Rhythmic factors could be implicated in both cases.

METHOD

Participants

The participants consisted of a total of 144 infants 6–8 months of age, 48 ($M = 30.08$ weeks, $SD = 3.16$; 25 girls, 25 boys) tested on audio and visual samples of speech, 48 ($M = 31.41$ weeks, $SD = 3.54$; 23 girls, 25 boys) on singing samples with videos from different songs, and 48 ($M = 32.73$, $SD = 1.80$; 21 girls, 27 boys) on the same singing samples with videos from different portions of the same song. All infants were healthy, born at term, and had no personal history of ear infections or family history of hearing loss, according to parental report.

Apparatus and stimuli

Infants were tested in a dimly lit sound-attenuating booth with the equipment described in Experiment 1 except for the presence of two additional monitors and a camcorder (Sony 360X) that transmitted images of the infant to the experimenter outside the booth. Infants were seated on their mother's lap facing the central monitor, with two other monitors 1 m away and at a 45-degree angle to their left and right. Parents wore headphones with masking music to prevent them from hearing the auditory stimuli presented to infants. Because of limited numbers of 6-month-old infants available at the time of testing, only three of the four pairs of stimuli from Experiments 1 and 2 (selected for best audio and video quality) were used. The video stimuli for the speech and singing segments were roughly 30 s in duration and were unaltered (i.e., no blurring of mouth area, as in Experiment 3). An experimenter outside the booth viewed the infant on a monitor and maintained a continuous record of infant looking to and away from the side monitors.

Procedure

Infants were first familiarized with the audio segments of speech or singing stimuli for 30 s during which time a silent video of a rotating globe was presented to help maintain infants' attention. Infants had 15 s of familiarization with the auditory stimulus paired with the silent video on one side followed by 15 s of the same auditory and visual stimuli on other side. Immediately after the familiarization phase, infants' attention was attracted to one of the side monitors by a flashing light on that monitor. When infants looked at that monitor, a silent video of the relevant condition (speech, different song, same song) was presented and continued to play until they looked away for 2 s. Infants' attention was then attracted to the monitor on the other side, and the contrasting silent speech or singing video from the same condition was presented until infants looked away for 2 s. The two silent video trials continued in alternation for a total of 10 trials. Half of the infants tested on each pair of speech or singing stimuli were familiarized with the audio sample of one mother and half with the audio sample of the other mother. In addition, the order of videos (target mother, other mother) and the side of first video trial (left or right) were counterbalanced.

Results and discussion

Because a number of infants in the speech condition failed to complete the full 10 trials, 6 trials (3 with each of the two video stimuli) were used for all infants in that condition. As can be seen in **Table 1**, the silent talking videos were rated lower in expressiveness than the silent singing videos. The full 10 trials were used in the singing conditions and are reported here. Proportions of infant looking time to the matching silent videos of speakers and singers in the three conditions are shown in **Figure 5**. Proportion of looking at the videos of previously heard speakers ($M = 0.573$, $SD = 0.128$) significantly exceeded chance levels (0.5), $t_{(47)} = 3.64$, $p = 0.001$, confirming infants' detection of cross-modal cues to speaker identity. By contrast, proportion of looking to the matching silent videos from different songs ($M = 0.497$) was at chance (see **Figure 2**). For videos featuring different portions of the same song, however, proportion

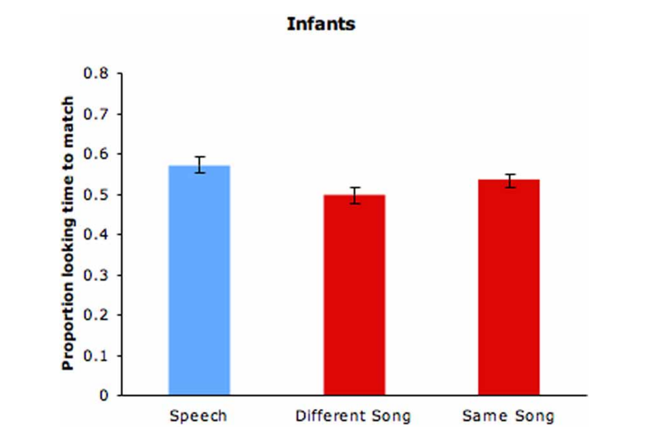


FIGURE 5 | Infants' proportion of looking time to the speech and singing videos of previously heard speakers and singers (same and different songs). Error bars are standard errors.

of looking at the matching videos of previously heard singers ($M = 0.534$, $SD = 0.116$) significantly exceeded chance levels, $t_{(47)} = 2.032$, $p = 0.048$. Differences in infant looking times are modest, but they are comparable to the levels reported in other familiarization-preference studies with 6-month-old infants that involve sequential presentation of stimuli (e.g., Hannon and Trehub, 2005). Overall, the findings from infants paralleled those from adults, with infants detecting cross-modal cues to identity for ID speech and for different portions of the same ID song but not for different ID songs.

GENERAL DISCUSSION

Adults and infants detected cross-modal cues to identity in maternal speech and singing. Adults' success in the present study confirms and extends the available evidence on cross-modal matching of talkers. It indicates that adults can identify maternal talkers from audio and video excerpts presented sequentially even when the excerpts are based on different verbal content (Kamachi et al., 2003; Lander et al., 2007). Previous research indicated that the manner of speech plays an important role such that changing manner across modalities (e.g., statement to question, conversational to clear speech) impairs cross-modal matching of speakers (Lander et al., 2007).

The manner of speech in the present study differed from that of earlier studies not only in its ID status but also in its derivation from natural interactions rather than portrayals. When "conversational" speech was used in previous studies of cross-modal matching (Lander et al., 2007), the adult "actors"' were instructed to memorize and produce a single scripted utterance ("I'm going to the library in the city.") and to "speak it in their usual natural manner (conversational statement)" (p. 906). By contrast, natural, conversational samples of ID speech in the present study were derived from playful maternal interactions with infants. As a result, the dynamic visual stimuli in each pair were based on speech samples that differed from each other as well as from the auditory stimuli. The range of possible variation across content, style, and modality was considerable. It would be of interest to ascertain whether adults would be capable of matching cross-modal cues to identity when auditory and visual cues are selected from contrasting registers such as conversational ID and AD speech, which vary considerably in expressiveness (Corbeil et al., 2013). Although female college students performed no better than their male counterparts on matching maternal voices to visual gestures, it is possible that mothers would perform better than non-mothers.

In the case of singing, adults perceived cross-modal cues to identity when the auditory and visual excerpts from each singer were from different portions of the same song with mouth movement obscured (Experiment 3) but not from different songs with intact movement (Experiment 2). Because all mothers sang different songs (i.e., songs that they typically sang to their infants), it is possible that adults in the present study simply identified the excerpts belonging to the same song rather than the same singer. Unfortunately, the design of the present study makes it impossible to rule out that interpretation. Identifying a well-known song from one of two silent videos, even with mouth movements obscured, may seem easy, but performance on the cross-modal singing task was modest and not significantly better than that on the speech task. Tempo appears to be an obvious cross-modal cue, but artificially speeding up or slowing down speech between familiarization and test stimuli does not interfere with adults' cross-modal matching (Lander et al., 2007). However, tempo is probably more salient in singing than in speech. In future research, artificial slowing or speeding of the tempo of maternal singing could indicate the relative contribution of absolute (i.e., tempo) and relative duration cues (i.e., rhythm).

Adults succeeded in identifying unfamiliar talkers and singers from cross-modal cues, but their performance in the present study and in earlier studies of talker identification was modest, roughly 70% correct or less. This kind of task is obviously difficult, even with 30-s passages of speech rather than the single words (Lachs and Pisoni, 2004) or single sentences (Kamachi et al., 2003; Lander et al., 2007) used in previous studies. Lachs and Pisoni (2004) argue that cross-modal matching is facilitated by the kinematics of articulation, but that may apply primarily to situations involving common lexical content across modalities. Removal of mouth cues in Experiment 3 did not significantly reduce performance accuracy, which suggests that global prosodic timing or rhythm was the primary amodal cue. Identifying the subtle visual rhythms that accompany speech and singing is an important challenge for the future.

Infants are presumed to use amodal cues when matching repeated vowels (/a/ or /i) to dynamic visual displays presented simultaneously and synchronously (Kuhl and Meltzoff, 1982; Patterson and Werker, 1999, 2002) and when matching repeating consonant-vowel syllables (/ba/ or /va/) to dynamic visual displays presented sequentially (Pons et al., 2009). Infants' use of amodal cues to identity in the present study, which involved sequential presentation of highly complex auditory and visual stimuli, is especially impressive. What did infants retain from the auditory familiarization phase, and what drove their longer looking times to videos of the previously heard speaker or singer? Perhaps adults formed intuitive impressions of the talkers and singers as they listened to the stimuli, even imagining what they might look like. Then they had an opportunity to watch both silent videos before deciding who was more likely to be the previously heard speaker or singer. Our supplementary rating experiment ruled out the most obvious factor in this regard, which was expressiveness or liveliness.

Adults typically have difficulty linking voices to static facial images (Kamachi et al., 2003; Lachs and Pisoni, 2004), but a recent study revealed poor but above-chance performance with static images presented sequentially (Mavica and Barenholtz, 2013). It is possible that adults generate expectations of a speaker's or singer's physical appearance or visual gestures while listening to that person, but infants are unlikely to do so. Nevertheless, the ID talking or singing in the present study primed infants for subsequent engagement with the talker's or singer's dynamic visual images. Something about each woman's ID speech or singing was engaging to infants as well as individually distinctive, memorable, and recognizable across modalities. As noted, global temporal features involving rhythmic prosody (Kamachi et al., 2003; Lander et al., 2007) are more likely candidates than local temporal features involving the fine-grained dynamics of articulation (Patterson and Werker, 1999; Lachs and Pisoni, 2004).

There was no indication that mouth movements contributed to adults' performance (Experiment 3), but they could have affected infants' performance. When exposed to audiovisual speech, 4-month-old infants fixate more on the eyes than on the mouth, 6-month-olds distribute their fixations equally across eye and mouth regions, and 8-month-olds focus more on the mouth than on the eyes (Lewkowicz and Hansen-Tift, 2012). Although there is no evidence that infants extract or retain person-specific cues to articulation, as older children do (Vongpaisal et al., 2010; van Heugten et al., in press), they may capitalize on other idiosyncratic features involving lip movements.

In sum, the present study revealed that mothers provide signature bimodal performances of speech and singing for their pre-verbal infants. Moreover, adults discern cross-modal cues to the identity of maternal speakers and singers and, remarkably, infants do so as well. An important task for future research is to specify the critical bimodal cues for infants and adults.

ACKNOWLEDGMENTS

Portions of this research were described summarily in an earlier publication (Trehub et al., 2009) and referenced as Trehub, Plantinga, and Brcic (manuscript in preparation). Funding was provided by the Social Sciences and Humanities Research Council of Canada and Advances in Interdisciplinary Research in Singing (AIRS).

REFERENCES

Bahrick, L. E., Lickliter, R., and Flom, R. (2004). Intersensory redundancy guides the development of selective attention, perception, and cognition in infancy. *Curr. Dir. Psychol. Sci.* 13, 99–102. doi: 10.1111/j.0963-7214.2004.00283.x

Banse, R., and Scherer, K. R. (1996). Acoustic profiles in vocal emotion expression. *J. Pers. Soc. Psychol.* 70, 614–636. doi: 10.1037/0022-3514.70.3.614

Bergeson, T. R., and Trehub, S. E. (2002). Absolute pitch and tempo in mothers' songs to infants. *Psychol. Sci.* 13, 72–75. doi: 10.1111/1467-9280.00413

Bergeson, T. R., and Trehub, S. E. (2007). Signature tunes in mothers' speech to infants. *Infant Behav. Dev*. 30, 648–654. doi: 10.1016/j.infbeh.2007.03.003

Brand, R. J., Baldwin, D. A., and Ashburn, L. A. (2002). Evidence for 'motionese': modifications in mothers' infant-directed action. *Dev. Sci.* 5, 72–83. doi: 10.1111/1467-7687.00211

Brand, R. J., and Shallcross, W. L. (2008). Infants prefer motionese to adult-directed action. *Dev. Sci.* 11, 853–861. doi: 10.1111/j.1467-7687.2008.00734.x

Brandt, A., Gebrian, M., and Slevc, L. R. (2012). Music and early language acquisition. *Front. Psychol.* 3:327. doi: 10.3389/fpsyg.2012.00327

Bryant, G. A., and Barrett, H. C. (2007). Recognizing intentions in infant-directed speech. *Psychol. Sci.* 18, 746–751. doi: 10.1111/j.1467-9280.2007.01970.x

Chong, S. C. F., Werker, J. F., Russell, J. A., and Carroll, J. M. (2003). Three facial expressions mothers direct to their infants. *Infant Child Dev*. 12, 211–232. doi: 10.1002/icd.286

Corbeil, M., Trehub, S. E., and Peretz, I. (2013). Speech vs. singing: infants choose happier sounds. *Front. Psychol.* 4:372. doi: 10.3389/fpsyg.2013.00372

DeCasper, A. J., and Fifer, W. P. (1980). Of human bonding: newborns prefer their mothers' voices. *Science* 208, 1174–1176. doi: 10.1126/science.7375928

Fellowes, J. M., Remez, R. E., and Rubin, P. E. (1997). Perceiving the sex and identity of a talker without natural vocal timbre. *Percept. Psychophys.* 59, 839–849. doi: 10.3758/BF03205502

Ferguson, C. (1964). Baby talk in six languages. *Am. Anthropol.* 66, 103–114. doi: 10.1525/aa.1964.66.suppl_3.02a00060

Fernald, A. (1985). Four-month-old infants prefer to listen to motherese. *Infant Behav. Dev*. 8, 181–195. doi: 10.1016/S0163-6383(85)80005-9

Fernald, A. (1989). Intonation and communicative intent in mothers' speech to infants: is the melody the message? *Child Dev.* 60, 1497–1510. doi: 10.2307/1130938

Fernald, A. (1992). "Meaningful melodies in mothers' speech to infants," in *Nonverbal Vocal Communication: Comparative and Developmental Approaches*, eds H. Papoušek and U. Jürgens (New York, NY: Cambridge University Press), 262–282.

Fernald, A., Taeschner, T., Dunn, J., Papoušek, M., de Boysson-Bardies, B., and Fukui, I. (1989). A cross-language study of prosodic modifications in mothers' and fathers' speech to preverbal infants. *J. Child Lang.* 16, 477–501. doi: 10.1017/S0305000900010679

Ghazanfar, A. A., Turesson, H. K., Maler, J. X., van Dinther, R., Patterson, R. D., and Logothetis, N. K. (2007). Vocal tract resonances as indexical cues in rhesus monkeys. *Curr. Biol.* 17, 425–430. doi: 10.1016/j.cub.2007.01.029

Grieser, D. A. L., and Kuhl, P. K. (1988). Maternal speech to infants in a tonal language: support for universal prosodic features in motherese. *Dev. Psychol.* 24, 14–20. doi: 10.1037/0012-1649.24.1.14

Grossmann, T., Striano, T., and Friederici, A. D. (2006). Crossmodal integration of emotional information from face and voice in the infant brain. *Dev. Sci.* 9, 309–315. doi: 10.1111/j.1467-7687.2006.00494.x

Hannon, E. E., and Trehub, S. E. (2005). Metrical categories in infancy and adulthood. *Psychol. Sci.* 16, 48–55. doi: 10.1111/j.0956-7976.2005.00779.x

Hill, H., and Johnston, A. (2001). Categorizing sex and identity from the biological motion of faces. *Curr. Biol.* 11, 880–885. doi: 10.1016/S0960-9822(01)00243-3

Kamachi, M., Hill, H., Lander, K., and Vatikiotis-Bateson, E. (2003). 'Putting the face to the voice': matching identity across modality. *Curr. Biol.* 13, 1709–1714. doi: 10.1016/j.cub.2003.09.005

Kitamura, C., and Burnham, D. (1998). The infant's response to maternal vocal affect. *Adv. Inf. Res.* 12, 221–236.

Kitamura, C., and Burnham, D. (2003). Pitch and communicative intent in mothers' speech: adjustments for age and sex in the first year. *Infancy* 4, 85–110. doi: 10.1207/S15327078IN0401_5

Kuhl, P. K., and Meltzoff, A. N. (1982). The bimodal development of speech in infancy. *Science* 218, 1139–1141. doi: 10.1126/science.7146899

Kunzel, H. J. (1989). How well does the average fundamental frequency correlate with speaker height and weight? *Phonetica* 46, 117–125. doi: 10.1159/000261832

Lachs, L., and Pisoni, D. B. (2004). Crossmodal source identification in speech perception. *Ecol. Psychol.* 16, 159–187. doi: 10.1207/s15326969eco1603_1

Lander, K., Hill, H., Kamachi, M., and Vatikiotis-Bateson, E. (2007). It's not what you say but the way you say it: matching faces and voices. *J. Exp. Psychol. Hum. Percept. Perform.* 33, 905–914. doi: 10.1037/0096-1523.33.4.905

Lebedeva, G. C., and Kuhl, P. K. (2010). Sing that tune: infants' perception of melody and lyrics and the facilitation of phonetic recognition in songs. *Inf. Behav. Dev*. 33, 419–430. doi: 10.1016/j.infbeh.2010.04.006

Lewkowicz, D. J., and Hansen-Tift, A. M. (2012). Infant deploy selective attention to the mouth of a talking face when learning speech. *Proc. Natl. Acad. Sci. U.S.A.* 109, 1431–1436. doi: 10.1073/pnas.1114783109

Lewkowicz, K. J., and Pons, F. (2013). Recognition of amodal language identity emerges in infancy. *Int. J. of Behav. Dev*. 37, 90–94. doi: 10.1177/0165025412467582

Mavica, L. W., and Barenholtz, E. (2013). Matching face and voice identity from static images. *J. Exp. Psychol. Hum. Percept. Perform.* 39, 307–312. doi: 10.1037/a0030945

Munhall, K. G., and Buchan, J. N. (2004). Something in the way she moves. *Trends Cog. Sci.* 8, 51–53. doi: 10.1016/j.tics.2003.12.009

Nakata, T., and Trehub, S. E. (2011). Expressive timing and dynamics in infant-directed and non-infant-directed singing. *Psychomusicol Music Mind Brain* 21, 45–53. doi: 10.1037/h0094003

Papoušek, M. (1994). Melodies in caregivers' speech: a species—specific guidance towards language. *Early Dev. Parenting* 3, 5–17. doi: 10.1002/edp.2430030103

Papoušek, M., Bornstein, M. H., Nuzzo, C., Papoušek, H., and Symmes, D. (1990). Infant responses to prototypical melodic contours in parental speech. *Infant Behav. Dev.* 13, 539–545. doi: 10.1016/0163-6383(90)90022-Z

Patterson, M., and Werker, J. F. (1999). Matching phonetic information in lips and voice is robust in 4.5-month-old infants. *Inf. Behav. Dev.* 22, 237–247. doi: 10.1016/S0163-6383(99)00003-X

Patterson, M., and Werker, J. F. (2002). Infants' ability to match dynamic phonetic and gender information in the face and voice. *J. Exp. Child Psychol*. 81, 93–115. doi: 10.1006/jecp.2001.2644

Plantinga, J., and Trehub, S. E. (2013). Revisiting the innate preference for consonance. *J. Exp. Psychol. Hum. Percept. Perform.* doi: 10.1037/a0033471. [Epub ahead of print].

Pollard, K. A., and Blumstein, D. T. (2011). Social group size predicts the evolution of individuality. *Curr. Biol.* 21, 413–417. doi: 10.1016/j.cub.2011.01.051

Pons, F., Lewkowicz, D. W., Soto-Faraco, S., and Sebastián-Gallés, N. (2009). Narrowing of intersensory speech perception in infancy. *Proc. Natl. Acad. Sci. U.S.A.* 106, 10598–10602. doi: 10.1073/pnas.0904134106

Rock, A. M. L., Trainor, L. J., and Addison, T. L. (1999). Distinctive messages in infant-directed lullabies and play songs. *Dev. Psychol.* 35, 527–534. doi: 10.1037/0012-1649.35.2.527

Rosenblum, L. D., Smith, N. M., and Niehus, R. P. (2007). Look who's talking: recognizing friends from visible articulation. *Percept*. 36, 157–159. doi: 10.1068/p5613

Rosenblum, L. D., Smith, N. M., Nichols, S. M., Hale, S., and Lee, J. (2006). Hearing a face: cross-modal speaker matching using isolated visible speech. *Percept. Psychophys*. 68, 84–93. doi: 10.3758/BF03193658

Sai, F., and Bushnell, I. W. R. (1988). The perception of faces in different poses by 1-month-olds. *Br. J. Dev. Psychol.* 6, 35–41. doi: 10.1111/j.2044-835X.1988.tb01078.x

Sheffert, S. M., Pisoni, D. B., Fellowes, J. M., and Remez, R. (2002). Learning to recognize talkers from natural, sinewave, and reversed speech samples. *J. Exp. Psychol. Hum. Percept. Perform.* 28, 1447–1469. doi: 10.1037/0096-1523.28.6.1447

Shenfield, T., Trehub, S. E., and Nakata, T. (2003). Maternal singing modulates infant arousal. *Psychol. Music* 31, 365–375. doi: 10.1177/03057356030314002

Singh, L., Morgan, J. L., and Best, C. T. (2002). Infants' listening preferences: baby talk or happy talk? *Infancy* 3, 365–394. doi: 10.1207/S15327078IN0303_5

Stern, D., Spieker, S., Barnett, R. J., and MacKain, K. (1983). The prosody of maternal speech: infant age and context related changes. *J. Child Lang*. 10, 1–15. doi: 10.1017/S0305000900005092

Thiessen, E. D., Hill, E. A., and Saffran, J. R. (2005). Infant-directed speech facilitates word segmentation. *Infancy* 7, 53–71. doi: 10.1207/s15327078in0701_5

Thompson, W. F., Russo, F. A., and Livingstone, S. R. (2010). Facial expressions of singers influence perceived pitch relations. *Psychon. B. Rev*. 17, 317–322. doi: 10.3758/PBR.17.3.317

Thompson, W. F., Russo, F. A., and Quinto, L. (2008). Audio-visual integration of emotional cues in song. *Cogn. Emot.* 22, 1457–1470. doi: 10.1080/02699930701813974

Trainor, L. J. (1996). Infant preferences for infant-directed versus non infant-directed playsongs and lullabies. *Infant Behav. Dev*. 19:1, 83–92. doi: 10.1016/S0163-6383(96)90046-6

Trainor, L. J., Austin, C. M., and Desjardins, R. N. (2000). Is infant-directed speech prosody a result of the vocal expression of emotion? *Psychol. Sci*. 11, 188–195. doi: 10.1111/1467-9280.00240

Trehub, S. E. (2009). "Music lessons from infants," in *The Oxford Handbook of Music Psychology,* eds S. Hallam, I. Cross, and M. Thaut (Oxford: Oxford University Press), 229–234.

Trehub, S. E., Hannon, E. E., and Schachner, A. (2010). "Perspectives on music and affect in the early years," in *Handbook of Music and Emotion: Theory, Research, Applications,* eds P. N. Juslin and J. A. Sloboda (Oxford: Oxford University Press), 645–668.

Trehub, S. E., Plantinga, J., and Brcic, J. (2009). Infants detect cross-modal cues to identity in speech and singing. *Ann. N.Y. Acad. Sci.* 1169, 508–511. doi: 10.1111/j.1749-6632.2009.04851.x

Trehub, S. E., and Trainor, L. J. (1998). Singing to infants: lullabies and play songs. *Adv. Inf. Res.* 12, 43–78.

Trehub, S. E., Trainor, L. J., and Unyk, A. M. (1993). Music and speech processing in the first year of life. *Adv. Child Dev. Behav*. 24, 1–35. doi: 10.1016/S0065-2407(08)60298-0

Tsang, C. D., and Conrad, N. J. (2010). Does the message matter? The effect of song type on infant preferences for lullabies and playsongs. *Inf. Behav. Dev*. 33, 96–100. doi: 10.1016/j.infbeh.2009.11.006

van Heugten, M., Volkova, A., Trehub, S. E., and Schellenberg, E. G. (in press). Children's recognition of spectrally degraded cartoon voices. *Ear Hear.*

Volkova, A., Trehub, S. E., and Schellenberg, E. G. (2006). Infants' memory for musical performances. *Dev. Sci.* 9, 583–589. doi: 10.1111/j.1467-7687.2006.00536.x

Vongpaisal, T., Trehub, S. E., Schellenberg, E. G., van Lieshout, P., and Papsin, B. C. (2010). Children with cochlear implants recognize their mother's voice. *Ear Hear*, 31, 555–566. doi: 10.1097/AUD.0b013e3181daae5a

Werker, J., and McLeod, P. (1989). Infant preference for both male and female infant-directed talk: a developmental study of attentional and affective responsiveness. *Can. J. Psychol.* 43, 230–246. doi: 10.1037/h0084224

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 02 August 2013; accepted: 14 October 2013; published online: 01 November 2013.

Citation: Trehub SE, Plantinga J, Brcic J and Nowicki M (2013) Cross-modal signatures in maternal speech and singing. Front. Psychol. ***4****:811. doi: 10.3389/fpsyg.2013.00811*

This article was submitted to Cognitive Science, a section of the journal Frontiers in Psychology.

frontiers in
PSYCHOLOGY

ORIGINAL RESEARCH ARTICLE
published: 29 April 2014
doi: 10.3389/fpsyg.2014.00262

Singing emotionally: a study of pre-production, production, and post-production facial expressions

Lena R. Quinto[1], William F. Thompson[1]*, Christian Kroos[1] and Caroline Palmer[2]

[1] Department of Psychology, Macquarie University, Sydney, NSW, Australia
[2] Department of Psychology, McGill University, Montreal, QC, Canada

Edited by:
Bruno Gingras, University of Vienna, Austria

Reviewed by:
Rosemary A. Cowell, University of Massachusetts Amherst, USA
Clemens Wöllner, University of Hamburg, Germany

***Correspondence:**
William F. Thompson, Department of Psychology, Macquarie University, Australian Hearing Hub Street, Balaclava Road, Sydney, NSW 2109, Australia
e-mail: bill.thompson@mq.edu.au

Singing involves vocal production accompanied by a dynamic and meaningful use of facial expressions, which may serve as ancillary gestures that complement, disambiguate, or reinforce the acoustic signal. In this investigation, we examined the use of facial movements to communicate emotion, focusing on movements arising in three epochs: before vocalization (pre-production), during vocalization (production), and immediately after vocalization (post-production). The stimuli were recordings of seven vocalists' facial movements as they sang short (14 syllable) melodic phrases with the intention of communicating happiness, sadness, irritation, or no emotion. Facial movements were presented as point-light displays to 16 observers who judged the emotion conveyed. Experiment 1 revealed that the accuracy of emotional judgment varied with singer, emotion, and epoch. Accuracy was highest in the production epoch, however, happiness was well communicated in the pre-production epoch. In Experiment 2, observers judged point-light displays of exaggerated movements. The ratings suggested that the extent of facial and head movements was largely perceived as a gauge of emotional arousal. In Experiment 3, observers rated point-light displays of scrambled movements. Configural information was removed in these stimuli but velocity and acceleration were retained. Exaggerated scrambled movements were likely to be associated with happiness or irritation whereas unexaggerated scrambled movements were more likely to be identified as "neutral." An analysis of singers' facial movements revealed systematic changes as a function of the emotional intentions of singers. The findings confirm the central role of facial expressions in vocal emotional communication, and highlight individual differences between singers in the amount and intelligibility of facial movements made before, during, and after vocalization.

Keywords: singing, emotional communication, point-light displays, face motion

INTRODUCTION

Emotional communication has been investigated in many different modalities including facial expressions (Elfenbein and Ambady, 2002), tone of voice (Johnstone and Scherer, 2000), music (Juslin and Laukka, 2003; Gabrielsson and Lindström, 2010), and gestures associated with music performance (Davidson, 1993; Thompson et al., 2005; Vines et al., 2006). Perceivers are sensitive to the information contained in these channels of communication and can decode emotional signals produced by individuals within and across cultures (Russell et al., 2003; Thompson and Balkwill, 2010).

In music, emotions are encoded in a range of acoustic attributes, including contour, modality, pitch height, intensity, tempo, and rhythm (for a review, see Juslin and Sloboda, 2010). Music performers often supplement these attributes with visual signals of emotion to enhance the clarity or impact of emotional communication. The facial expressions and gestures of performers are known to influence the perception of expressiveness (Davidson, 1993, 1995), tension (Vines et al., 2006), timbre (Saldaña and Rosenblum, 1993), dissonance (Thompson et al., 2005), note duration (Schutz and Lipscomb, 2007), interval size (Thompson and Russo, 2007), phrase structure (Ceaser et al., 2009), and emotion (Dahl and Friberg, 2007; Thompson et al., 2008). Ensemble musicians also use gestures and eye contact to facilitate coordinated action, particularly in sections that introduce new or important material (Williamon and Davidson, 2002).

Studies that have used video recordings have demonstrated that facial expressions can communicate a range of information associated with music performance. Facial expressions used in guitar performances by B.B. King, for example, appear to signal technical difficulty whereas other facial expressions appear to reflect current levels of dissonance associated with a musical passage (Thompson et al., 2005). A case study of the pianist Lang Lang revealed that his facial expressions closely mirrored the musical structure and the underlying meaning of a programmatic musical work (Davidson, 2012). Wöllner (2008) found that expressiveness ratings for audio-visual presentations of orchestral music were more closely correlated with ratings of the conductor's facial expressions than with ratings of the conductor's arms or blurred body movements. Similarly, if auditory information is held constant across renditions but paired with different visual gestures, performance judgments differ (Behne and Wöllner, 2011). A recent meta-analysis revealed a moderate but reliable effect size of the visual domain on perceptions of expressiveness, overall quality, and liking (Platz and Kopiez, 2012).

Musicians can also communicate discrete emotional states such as "happy" and "sad" through the use of facial expressions (Thompson et al., 2005). A sounded major third is judged to be sadder when combined with facial expressions made while singing a minor third, and a sounded minor third is judged to be happier when combined with facial expressions made while singing a major third (Thompson et al., 2008). Dahl and Friberg (2007) found that the emotional intentions of happiness, sadness, and anger were communicated well by the body and head movements of musicians, such that viewers did not even need auditory information to determine the intended emotion.

Music performances are inherently dynamic and emotional responses may change over time (Schubert, 2004). Early work was largely restricted to examinations of static images. Examining the visual information available from complex dynamic motion in facial expressions and body movements was a challenge, particularly in isolating the core dynamic features that were used by perceivers to decode emotion. One method used to examine the contribution of motion to perception was through the use of point-light displays (PLDs). PLDs present the visual information in a reduced form. Before motion capture technology was developed, PLDs were achieved by placing reflective or white markers on dark clothing or a face that had been darkened with make-up. In this method, the form information from a single static image is difficult to identify and unique features are often lost. The addition of dynamic information allows viewers to easily identify biological motion (Blake and Shiffrar, 2007). Using PLDs, participants are able to decode emotion from facial expressions (Bassili, 1978), and even through the gait of point-light walkers (Halovic and Kroos, 2009). Participants are also better able to identify musicians' expressive intentions when presented with the body movements of performers (no sound) than when presented with the sounded performance without visual information (Davidson, 1993). Currently, motion capture allows researchers to record movement, quantitively analyse this movement, and develop PLD videos. Motion capture also allows for the manipulation of features in the point-light display (e.g., only showing particular features or developing non-biological control stimuli). A second method to understand the influence of movement on viewers' perception is to use full-video recordings. Full-video has often been used to examine the visual influence in music. To understand the specific features of interest, researchers sometimes occlude parts of the performer (e.g., Dahl and Friberg, 2007; Thompson et al., 2010) or use filtering methods so that specific features are difficult to identify (e.g., Wöllner, 2008).

Humans appear to be extremely sensitive to motion and emotional information such that the full apex of an emotional expression is not needed to decode emotion. Fiorentini et al. (2012) showed participants images of emotional expressions that developed over time and found that viewers perceived emotions well before the full emotional configuration was reached. One interpretation of these findings is that viewers make use of individual features that emerge early in the formation of a facial expression, such as lip and eyebrow movements. Such features are then used to make probabilistic judgments of an intended emotion.

In music, facial expressions and gestures often occur outside the boundaries of sounded music, for example, in moments of silence that occur before and after musical phrases are vocalized. These ancillary gestures are not a direct consequence of the physical constraints of vocal production but, rather, act to signal emotional, social, and other communicative goals (Davidson, 1995; Palmer, 2012). In some cases, facial expressions reinforce communicative goals that may be ambiguous in the sounded performance, clarifying the structural or emotional characteristics of the music.

Supporting this idea, Livingstone et al. (2009) reported that singers exhibited emotional facial expressions well before they were expected to sing. Musicians watched a model singer express a musical phrase communicating happiness, sadness, or no expression. They were then asked to sing back this phrase and their movements were recorded with motion capture. The results showed that musicians surrounded their vocalizations with meaningful facial expressions. Intended emotions were reflected in facial expressions before, during, and after vocalizations. These findings suggest that musicians hint at the emotional information that is forthcoming in a musical phrase, and sustain those emotional expressions after the cessation of that phrase. Such supra-production expressions may benefit audience members by optimizing their capacity to extract communicative intentions (see also Wanderley et al., 2005).

We used motion capture to examine the facial expressions of seven musicians as they sang phrases with each of four emotional intentions: happiness, sadness, irritation, and no emotion. Irritation was used instead of anger to convey a subtler version of the latter emotion. Facial expressions were captured and analyzed in three epochs: before the musicians began singing (pre-production), during singing (production), and once they had completed singing (post-production). Point-light displays of these facial expressions (without sound) were then presented to independent perceivers who judged their emotional content in the first experiment. In subsequent experiments, we presented the same facial movements to participants along with exaggerated forms (facial movements were algorithmically manipulated to contain a larger range of movements) and in scrambled forms (randomized the direction of marker movements, keeping range of motion constant). The scrambled condition showed the initial marker positions but as the motion started, the direction of the marker trajectory was randomly determined while keeping the range, velocity and acceleration constant. These manipulations allowed us to better understand the nature of the cues used by perceivers to decode emotional intentions.

EXPERIMENT 1

The goal of Experiment 1 was to examine the ability of perceivers to decode the emotional dynamic facial expressions and head movements observed in point-light displays of seven singers. We expected that emotional decoding would be highest in the production phase, when musicians are most likely to be focusing on their communicative intentions. Although musicians may be more focused on communicating the emotion in the production phase, production constraints associated with singing might limit the capacity of singers to express emotion through movements of the mouth. The findings of Livingstone et al. (2009) suggest that the pre- and post-production epochs contain important movement information that singers use to communicate emotion through facial expressions made before and after singing. Perceivers appear to mimic the emotional expressions of singers (see also Chan et al., 2013) but it is unclear whether perceivers can use this information to accurately decode the intended emotion based solely on the motion information conveyed in point-light displays.

It was expected that some emotions would be better decoded depending on the epoch. For example, Bassili (1979), who used PLDs, found that anger is communicated through eyebrow movements and frowns, whereas happiness is communicated through mouth movements (which presumably do not occur in pre- and post-singing epochs). A study of singing using full-video found that happiness was not well communicated during singing, in contrast anger and sadness were communicated during singing (Scotto di Carlo and Guaitella, 2004). Thus, it was expected that the emotion of happiness may not be as well communicated in the production epoch as in the pre- or post-production epochs. In contrast, irritation and sadness were expected to be decoded equally well in each of the epochs.

Finally, we also expected individual differences between singers in their ability to communicate specific emotions, and their tendency to express emotions in facial expressions before and after vocalizations. Although emotional encoding and decoding occurs universally in static facial expressions (Ekman and Friesen, 1971), social norms influence the expression of certain emotions (Scherer et al., 2003) and there are individual differences in the ability to communicate emotionally in music (Davidson, 1993, 2012; Juslin, 2000; Wanderley et al., 2005; Dahl and Friberg, 2007; Timmers and Ashley, 2007). Wanderley et al. (2005) observed that clarinettists differed from each other in the use of idiosyncratic gestures such as knee bending, vertical shoulder movement, and circular movements of the clarinet bell. Similarly, Davidson (2012) observed variability in the body movements used by flautists and clarinettists. Despite such individual differences in performance gestures, perceivers are still able to decode emotional intentions. Consistent with Brunswik's lens model (1956; see also Juslin, 2000), emotional decoding is possible because there are several redundant cues associated with any one emotion, and perceivers evaluate such emotional cues probabilistically. A probabilistic decoding strategy allows perceivers to adapt to idiosyncratic strategies of communicating emotion. In the current study, while all singers were trained musicians, some had more experience as singers whereas others had more experience as instrumentalists. As such, we examined the ability of perceivers to decode emotional facial expressions for each singer separately.

METHODS

Musicians

Seven singers participated in the motion capture session. They were recruited through advertisements to local music theatre groups, drama societies, and choirs. Singers were selected on the following basis: (a) they were actively involved in music-making, (b) they were able to use facial expressions to communicate emotion, and (c) they were able to sing the melody in tune. Two judges determined whether an individual was a possible candidate for the session: One judge was a recording engineer with experience in music education and made decisions regarding the quality of the auditory information. The other judge was a researcher with experience in facial expressions and determined the quality of information conveyed through the visual domain.

All singers were currently involved in music. Most had been singing since childhood and had received extensive musical training. They had an average age of 29 years ($SD = 12.64$); an average of 9.83 ($SD = 6.73$; range = 3–20) years of formal music training; and an average of 22.83 ($SD = 11.39$; range = 5–45) years of active involvement in music. All were paid for their participation.

Motion capture equipment

Figure 1 illustrates the facial positions of 28 of the 29 Vicon markers that were placed on musicians using double-sided hypoallergenic tape. The musicians were asked to wear dark clothing and to avoid wearing make-up or sunscreen for the experimental session. Three markers were positioned on each eyebrow, two were positioned under each eye, six outlined the lips and three outlined the cheeks. One marker was placed on each of the following: chin, forehead, left and right temple, tip of the nose, nasion, and the shoulder as a reference point. The marker on the shoulder was excluded from the animated stimuli. The markers on the temples, shoulder and forehead were 9 mm in diameter and the remaining markers were 4 mm in diameter. The musicians were recorded with eight Vicon MX+ infrared cameras at a frame rate of 200 frames per second. Musicians stood in the middle of an 8-foot capture space (surrounded by the eight cameras).

Stimulus materials

Singers were asked to sing the text phrase to an experimental melody (**Figure 2**) that was presented to them through headphones in a piano timbre. This melody was neutral with respect to its musical mode, which is known to influence emotional judgments (e.g., Hevner, 1935), and was synchronized to a metronome at a tempo of 500 ms per beat. Singers were instructed to sing one syllable of the scripted phrase on each beat.

Four text phrases were created, designed to be semantically neutral or ambiguous in terms of their emotional connotation ("The orange cat sat on a mat and ate a big, fat rat," "The girl and boy walked to the fridge to fetch some milk for lunch," "The broom is in the closet and the book is on the desk," "The small green frog sat on a log and caught a lot of flies").

On each trial, the textual phrase and one of four specific emotions were projected simultaneously on a screen located approximately four meters in front of the singers. The singers were asked to express one of four emotions (irritation, happiness, sadness

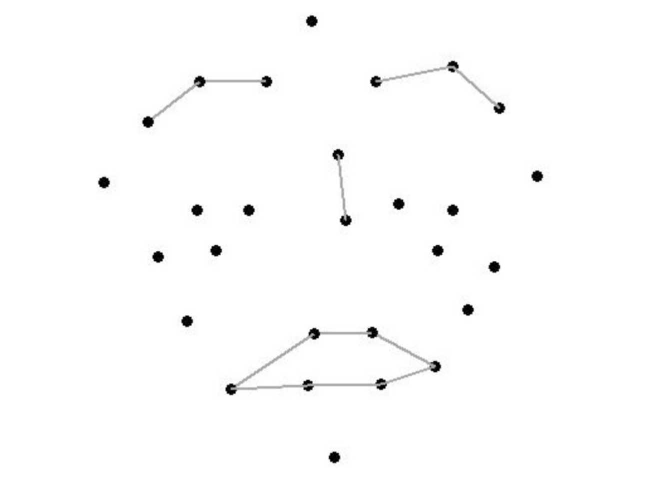


FIGURE 1 | The position of the markers outlining the major features of the face; lines indicate eyebrows, nose, and lips.

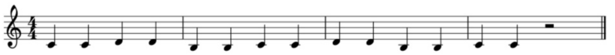


FIGURE 2 | The melody sung by performers.

and neutral/no emotion). Then a recording of the melody was played, followed by four metronome beats that signaled to the singers to begin singing the scripted phrase. Each motion capture recording was initiated when the experimental melody ended and the first metronome beat began. The motion capture recording ended four to five beats after the singing ceased. In total, there were 112 recordings (7 musicians × 4 emotions × 4 phrases).

Point-light stimulus creation

All motion capture stimuli were gap-filled and cleaned to ensure that marker trajectories appeared natural. The shoulder marker was removed from the data set. The spatial trajectories of the remaining 28 markers were smoothed to reduce measurement noise. Smoother trajectories were estimated from the original data using Functional Data Analysis (FDA; Ramsay and Silverman, 2005). This analysis method converts the discrete measurements into continuous functions based on b-splines with a roughness penalty λ set to 10^{-12} applied to the second derivative (acceleration). All recordings were numerically centered by making the origin equivalent to the approximate center of head rotation (located in the neck). The six independent head motion parameters (three translational, three rotational) were estimated from three markers (nasion, right temple, left temple), which were assumed to have moved only due to rigid head motion with no or very little interference from non-rigid skin motion. The standard estimation algorithm based on Procrustes Analysis (Gower, 1975) showed small residuals confirming that the markers were largely unaffected by skin movements.

Data for the three epochs were extracted from the full recordings in the following way. First, two researchers independently determined the onset of the first sung syllable, based on acoustic inspection. In most cases, the judgments were based on the acoustic signal. In a few instances, however, the acoustic signal was missing and the onset and offset of facial singing movements had to be visually approximated and so provided the only criterion for a decision. The average difference between the raters in start times was 10 frames (=50 ms) and for end times was 33 frames (=165 ms).

For the pre-production epoch, data samples from 1.5 s before the onset of the singing were selected. For the production epoch, samples corresponding to a duration of 1.5 s centered on the midpoint of the sung phase were selected. For the post-production epoch, data samples starting with the offset of the singing and extending to 1.5 s beyond this point were selected. The marker data was turned into video clips of point-light displays without any other modifications. Each marker was represented by a black dot moving in front of a white background. A frontal perspective was chosen to reduce the three-dimensional data to the two dimensions of the video clip. The perspective coincided with the x-axis of the Vicon coordinate system and coincided with the direction of an assumed audience during the motion capture session. The movement range across all trials was determined beforehand and the display limit was set accordingly to keep the point-lights visible at all times.

To ensure that the stimulus was recognized as a face, a brief anchor stimulus was added to the beginning of every clip. It consisted of a static point-light face generated from the reference sample (before any emotion was expressed), but with gray lines inserted between selected markers so as to emphasize salient facial features (see **Figure 1**). Three anatomical structures were emphasized: the mouth, by connecting lip markers; the eyebrows, by connecting medial to lateral eyebrow markers; and the nose, by connecting the nasion and the nose tip marker. The final clip consisted of the following sequence: a blank (white) screen for 0.4 s; the static anchor face for 1 s; another blank screen for 0.4 s; the point-light motion stimulus (without connecting lines) for a duration of 1.5 s; and a final blank screen for 0.4 s.

The entire processing described above was accomplished through custom-written Matlab (The MathWorks) routines. To achieve the desired video frame rate of 25 fps, the motion data were down-sampled. For each data sample, a video frame was created in the form of a Matlab figure that was subsequently added to a Quicktime movie using the Matlab Quicktime toolbox written by Slaney (1999).

Analysis of movement data (PCA)

The motions of the singers were assessed to quantitatively examine the changes in facial motion over time. A principal components analysis (PCA) of facial movements and head movements was conducted, using stimuli from both Experiment 1 (normal movements) and Experiment 2 (exaggerated movements). Combining stimuli from the two experiments provided us with enough observations for a robust PCA with 27 variables. The movements of the musicians were first quantified by their displacement (relative to the positions of the neutral expression at the beginning of each trial), velocity and acceleration for the points associated with the lip corners, eyebrows, front-back head movement, lateral head movement, up-down head movement, and the rotational movements of pitch, roll and yaw. PCA is an appropriate analysis because many of these motion variables were highly correlated. Before the analysis was performed, the movement variables were standardized to have the same variance. Five components emerged with eigenvalues greater than 1 (which we used as cut-off criterion). The five components accounted for 82 percent of the variation in the data.

Table 1 shows the correlation between each component and the motion variable of interest. Component 1 is associated with changes in the mouth region, Component 2 is associated with head displacement and head velocity, Component 3 is most strongly associated with head movements and rotations from side to side, Component 4 is associated with head acceleration, and Component 5 is associated with eyebrow movement.

Differences between epochs

The average component scores for each epoch are shown in **Figure 3**. The graph shows that, not surprisingly, there were higher scores in the production epoch for every component as compared to the pre- and post-production epochs. This reflects the larger movements that were used by singers during singing. The figure also shows that there was less movement in the post-production epoch than the pre-production epoch—particularly for the 1st and 5th components, which are associated with mouth and eyebrow movements respectively.

Individual differences between singers

An analysis of differences in the use of movements by singers, as reflected by component scores, was performed. A multivariate analysis of variance with singer (7) as the independent variable and the 5 components as the dependent variables showed that singers may have used somewhat different strategies to encode their emotional intentions. There were significant differences between singers in each of the five components, all *F*s > 11.46, *p*'s < 0.001. **Figure 4** illustrates the average principal component (PC) values for each singer and indicates individual differences in

Table 1 | The principal component scores from the rotated component matrix.

Variable	Rotated component matrix				
	Component				
	1 Mouth	2 Head displacement/velocity	3 Side head motion	4 Head acceleration	5 Eyebrows
Mouth corner displacement	**0.724**	0.246	0.248	0.110	0.208
Eyebrow displacement	0.130	0.319	0.153	−0.115	**0.757**
Mouth area	**0.806**	0.201	0.184	0.113	−0.003
Mouth corner velocity	**0.914**	0.140	0.171	0.111	0.198
Eyebrow velocity	0.282	0.149	0.223	0.138	**0.877**
Mouth area velocity	**0.928**	0.154	0.165	0.104	0.081
Mouth corner acceleration	**0.867**	0.044	0.133	0.201	0.275
Eyebrow acceleration	0.227	−0.037	0.202	0.321	**0.793**
Mouth area acceleration	**0.912**	0.113	0.140	0.135	0.135
Head front back displacement	0.119	**0.699**	0.381	−0.141	0.133
Head lateral displacement	−0.017	0.436	0.552	0.311	0.013
Head up down displacement	0.088	**0.808**	0.220	0.305	0.025
Head front back velocity	0.262	**0.638**	0.569	0.080	0.154
Head lateral velocity	0.148	0.399	**0.690**	0.351	0.058
Head up down velocity	0.152	**0.652**	0.321	0.598	−0.016
Head front back acceleration	0.431	0.327	0.545	0.410	0.272
Head lateral acceleration	0.370	0.234	**0.720**	0.366	0.220
Head up down acceleration	0.257	0.358	0.330	**0.749**	0.066
Head rotation roll displacement	0.197	**0.709**	0.303	0.136	0.126
Head rotation pitch displacement	0.193	**0.774**	0.108	0.234	0.188
Head rotation yaw displacement	0.143	0.427	**0.740**	0.108	0.203
Head rotation roll velocity	0.233	**0.664**	0.429	0.382	0.132
Head rotation pitch velocity	0.289	**0.634**	0.242	0.505	0.211
Head rotation yaw velocity	0.245	0.298	**0.841**	0.103	0.206
Head rotation roll acceleration	0.337	0.443	0.519	0.487	0.219
Head rotation pitch acceleration	0.431	0.269	0.237	**0.674**	0.340
Head rotation yaw acceleration	0.361	0.150	**0.810**	0.114	0.251

Large correlations >0.6 are in bold.

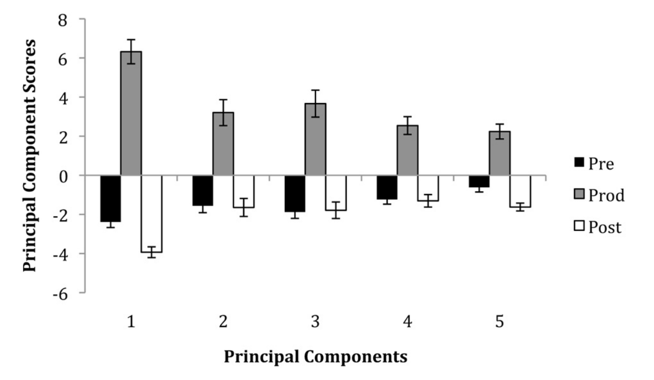


FIGURE 3 | The average principal component scores for each epoch. Error bars represent standard errors.

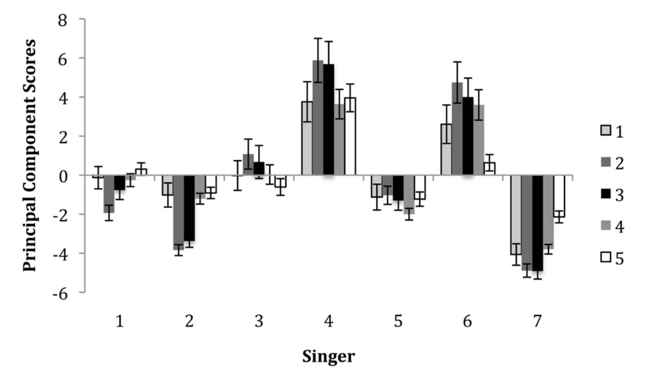


FIGURE 4 | The average principal component scores for each singer. Error bars represent standard errors.

facial movement across features. The averaging over the five principal components gives an indication of overall movement across features. Generally, Singers 4 and 6 used more extensive movements than other musicians. Singer 4 showed prominent eyebrow movement (Component 5), mouth movement (Component 1) and head movement (Components 2–4) when compared to other singers. In contrast, Singer 6 used more extensive head movement (Components 2–4) than the other singers. Singer 7 used smaller facial and head movement than the other singers, with the exception of Singer 2, who used very little head movement. The analysis of the motion data revealed that there were several aspects of motion associated with the expression of emotion by the singers. Singers used facial expressions (mouth and eyebrow movement) and head movement to express emotion. Individual singers also varied in their overall use of motion and in the specific movements that they employed.

EMOTIONAL DECODING

Participants

Sixteen members of the Macquarie University community including researchers, graduate students and post-doctoral fellows (11 females and 5 males) participated in Experiments 1–3, during which they provided ratings of 336 stimuli. There were 1344 conditions (7 musicians × 4 emotions × 3 epochs × 2 exaggeration × 2 scrambled × 4 phrases) but each participant only rated one phrase. The average age of the participants was 37.75 ($SD = 15.16$; range = 21–62) years. Although each experiment was not independent (the same viewers participated), the analyses between variables are reported separately to allow for ease of interpretation.

Materials and procedure

The point-light stimuli were presented on an Apple Macintosh iMac12.2 with an integrated 27 inch monitor that had 2560 × 1440 pixel resolution and was situated in a quiet room. The participants were seated with their face approximately 60 cm away from the monitor, such that the stimulus area subtended a visual angle of roughly 11 degrees. Stimuli were presented in six blocks, with different epochs (pre-production, production, post-production) and scrambling mode (see Experiment 3) presented in separate blocks. To reduce the length of the experiment, the 16 participants were randomly and independently assigned in sets of four to stimuli containing only one of the four text phrases. The exaggerated stimuli (Experiment 2) were presented in the same blocks as the normal stimuli, as these stimuli met the expectations for biological motion.

Custom-written software was programmed in Python and a web-based framework was used to show the movie clips and obtain the ratings from the participants. For each trial, there were four slider scales labeled “Happiness,” “Irritation,” “Sadness,” and “Neutral” ranging from 1 (“*not at all*”) and 7 (“*very much*”). The four sliders appeared horizontally stacked underneath the area where the movie was displayed. The stack order was randomized across blocks. The participants were instructed to first watch the movie and then rate the perceived strength of the emotion expressed by the point-light face by moving the sliders with the computer mouse to a position between 1 and 7. In the pre-production epoch, participants were instructed to rate the extent to which the singer moved toward conveying a particular emotion (i.e., from neutral to some emotion). In the production epoch, participants were instructed to rate the extent to which the singer conveyed a particular emotion. In the post-production epoch, participants were instructed to rate the extent to which the singer moved away from conveying a particular emotion (i.e., from an emotion toward neutral). The participants were able to use more than one scale to indicate a mixture of perceived emotions and were made aware of this option. Once they were satisfied with their ratings they continued to the next trial. There was no audio associated with any of the stimuli.

RESULTS

Three hundred and thirty-six conditions were analyzed in a mixed-design analysis (4 emotions × 4 phrases × 7 singers × 3 epochs), with 84 trials rated per viewer (one phrase). The exaggerated and scrambled conditions were assessed in Experiments 2 and 3. To assess the accuracy of emotional decoding, the emotion ratings were first converted to correct/incorrect responses. The response was considered “correct” if the highest rating of the four emotional ratings matched the emotion communicated and “incorrect” otherwise. For example, if the intended emotion was assigned a rating of "2" and the remaining options were assigned ratings of "1," the intended emotion was still considered correct as this option had the highest rating relative to the incorrect options. Cases in which participants rated two emotions equally high (one matching the intended emotion and the other not matching the intended emotion) were coded as incorrect ($n = 48$).

Correct responses by epoch, singer and emotion

In all three experiments, decoding accuracy did not differ between phrases, therefore these conditions were combined. A GLM analysis including the factors of epoch, singer, emotion and all interactions was performed. **Figures 5A–D** show the mean ratings by emotion, epoch, and singer. Overall, the mean correct responses ($M = 37.43$; $SE = 7.26$) indicated that emotions were decoded at above chance levels. There was a main effect of emotion, $F_{(3,\ 1245)} = 27.44$, $p < 0.001$. This reflected the finding that neutral and happiness were decoded more accurately than irritation and sadness. There was also a main effect of singer, $F_{(6,\ 1245)} = 3.20$, $p = 0.004$. Generally, this showed that Singer 4 was most able to communicate expressively across emotions as compared to the other singers. There was also a significant emotion x singer interaction, $F_{(18,\ 1245)} = 3.903$, $p < 0.001$, which showed that some singers were better at communicating particular emotions than other singers. For example, happiness was best decoded when expressed by Singers 4 and 6, irritation was best decoded when expressed by Singer 4, and sadness was best decoded when expressed by Singers 1 and 7.

Although there was no significant main effect of epoch, $F_{(2,\ 1245)} = 1.29$, $p = 0.279$, there were significant interactions of epoch with other variables: between epoch × emotion, $F_{(6,\ 1245)} = 2.520$, $p = 0.020$; and epoch x singer, $F_{(12,\ 1245)} = 2.208$, $p = 0.010$. The 2-way interaction for epoch x emotion showed that happiness was generally better decoded in the pre-production epoch ($M = 45.53$, $SD = 50.02$) than the post-production epoch ($M = 28.57$, $SD = 45.37$), $t_{(15)} = 2.83$, $p < 0.014$. The epoch by singer interaction showed that overall, Singer 1 was best able to express emotion in the pre-production epoch as compared to the production and post-production epochs and Singer 4 was marginally better at communicating emotions in both the pre-production and production epochs as compared to the post-production epochs.

Finally, there was a 3-way interaction with epoch x singer x emotion, $F_{(36,\ 1245)} = 1.781$, $p = 0.003$. Tests of simple effects with Bonferroni correction showed that there were no significant differences across epochs for Singer 2, Singer 6 and Singer 7. Singer 3 and Singer 4 were better able to express happiness in the pre-production epoch as compared to the production epoch, $t_{(15)} = 3.16$, $p < 0.005$ and $t_{(15)} = 3.65$, $p < 0.001$, respectively. Singer 4 was better able to communicate happiness in the production epoch, $t_{(15)} = 2.76$, $p < 0.017$, as compared to the post-production epoch. Singer 4 also communicated irritation better in the pre-production epoch than the post-production epoch, $t_{(15)} = 2.76$, $p < 0.017$. Singer 1 was best able to communicate sadness in the pre-production epoch as compared to the post-production epoch, $t_{(15)} = 2.76$, $p < 0.017$, while Singer 5 was better able to express sadness in the pre-production epoch

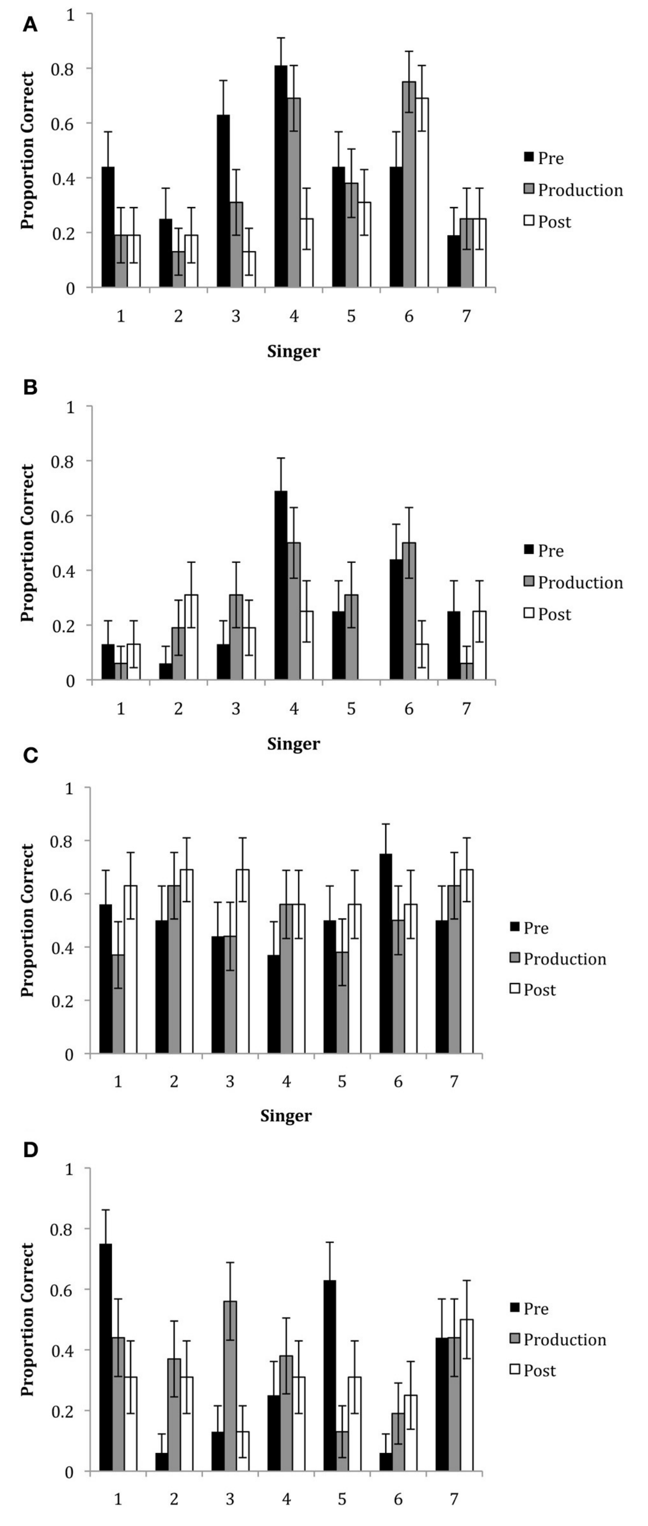


FIGURE 5 | The proportion of correct responses for the emotions of (A) happiness, (B) irritation, (C) neutral and (D) sadness for each of the seven singers in each of the three epochs. Note that missing bars indicate that no participant accurately decoded the emotional intention.

as compared to the production epoch, $t_{(15)} = 3.16$, $p < 0.005$. Singer 3 was best able to express sadness in the production epoch as compared to the pre- and post-production epochs, $t_{(15)} = 2.76$, $p < 0.017$.

DISCUSSION

The findings of Experiment 1 showed that expressions of happiness and neutral were more likely to be perceived by viewers from point-light displays of singers' facial features compared to expressions of irritation and sadness. Although anger and sadness may be communicated in full-video (Dahl and Friberg, 2007), previous work using PLDs has shown that the emotions of anger and sadness may not be as well communicated as happiness in PLDs (Bassili, 1979). The results also showed that emotional decoding was dependent on the singer and epoch. Perceivers were better able to decode emotions in the pre-production and production epochs, as compared to the post-production epoch. Generally, happiness was more clearly decoded in the pre-production epoch than the production epoch. This is consistent with previous findings, suggesting that happiness is a difficult emotion to convey during singing because facial areas signaling happiness are being recruited (Scotto di Carlo and Guaitella, 2004). For some singers (4, 5, 6), perceivers decoded irritation better in the pre-production epoch as compared to the post-production epoch. Similarly, perceivers were better able to decode sadness when communicated by Singer 1 and Singer 5 in the pre-production epoch as compared to the post-production and production epochs respectively. Cues to anger and sadness might be found higher in the face in the form of a frowning motion or raised eyebrows (Bassili, 1979). Due to the restrictions involved in singing, singers conveyed some of the cues just before singing, while other cues, such as eyebrow movements and head movements could be used during singing.

We did not find a strong effect of post-production lingering, at least with regard to emotional decoding. We might infer that from the perspective of the viewing participants, once singers had completed singing, there was not much available evidence for participants to determine the emotion. These findings at first seem to contrast with those of Livingstone et al. (2009), who found that both with motion capture and with EMG, musicians "lingered" or maintained the displacement from the production phase into the post-production phase. However, one important difference between these studies is that Livingstone et al. focused on the production of emotional singing and did not examine emotional decoding. It is possible that musicians in our study did emotionally "linger" or prepare but this may not have been sufficient for perceiving participants to determine the emotional intention in PLDs.

EXPERIMENT 2

The findings of Experiment 1 showed that happiness and neutral were more likely to be decoded by viewers than irritation and sadness. Importantly, several singers expressed the emotion of happiness through facial expressions even before they began singing. Given the modest levels with which the emotional intentions were decoded, Experiment 2 was designed to evaluate whether emotional cues were present but were too subtle for perceivers, based on facial (visual) cues. That is, singers may have encoded the emotion in facial expressions but such movements may not have been sufficiently clear to perceivers, especially when presented as PLDs.

To evaluate this possibility, the PLDs in Experiment 2 were manipulated so that facial movements were exaggerated twofold. This manipulation was performed to assess whether the relevant emotional information was present in facial movements but not adequately detected by perceivers. We expected that exaggerated movements would be more accurately decoded than non-exaggerated movements, because exaggerated movements should convey greater emotional intensity (Pollick et al., 2003). Indeed, a comparison of performance movements for deadpan and expressive performances revealed that the movements used in expressive performances are similar to, but larger than the movements used in deadpan performances (Davidson, 1994; Wanderley et al., 2005). That is, exaggerated movements may enhance the expressiveness of facial movements, leading to increased decoding accuracy. However, exaggerating the temporal and dynamic characteristics of the motion may actually lead to reduced decoding accuracy for some emotions that may rely on slower movements (e.g., sadness; Kamachi et al., 2001; Sato and Yoshikawa, 2004; Recio et al., 2013).

METHODS

Participants

The participants were the same 16 individuals from Experiment 1. Technically, Experiments 2 and 3 might be considered separate conditions rather than experiments; however, these conditions were separated to make interpretation clearer. A caveat of this approach is that participants were exposed to all the stimuli, which might have biased responses in various conditions. That is, participants' ratings may have been influenced by stimuli to which they had previously been exposed.

Point-light creation

We created exaggerated stimuli by multiplying the original head motion parameters and face motion trajectories by a factor of two. This doubled the distance of each marker trajectory while keeping the time constant. However, the velocity of the marker movement was also increased. This level of exaggeration was selected with the aim of maximizing the impact of the resultant movements without appearing to be biologically impossible. The reference positions for each trajectory were subtracted from the entire trajectory before being exaggerated. As each trial always commenced with a neutral facial expression (which was used as an anchor before participants saw the dynamic PLDs), these expressions determined the reference positions. In this analysis, only the exaggerated stimuli were considered.

RESULTS

Emotional decoding

Three hundred and thirty-six conditions were analyzed (4 emotions × 4 phrases × 7 singers × 3 epochs); each viewer rated 84 trials (one phrase) for the exaggerated movement stimuli. As before, trials were considered incorrect when participants rated two emotions equally high (one matching the intended emotion and the other not matching the intended emotion; $n = 55$). Overall, the mean correct responses ($M = 38.91$; $SE = 7.63$) indicated emotions were decoded above chance levels. A GLM analysis including the factors of epoch, singer, emotion and all interactions was performed on the percent correct values. **Figures 6A–D** show the mean responses by epoch, singer, and emotion. The results showed that there was a significant main effect of epoch, $F_{(2,\ 1245)} = 6.172$, $p = 0.002$. This finding showed that emotional decoding was better in the pre-production and production epochs than in the post-production epoch. There was also a main effect of emotion, $F_{(3,\ 1245)} = 11.08$, $p < 0.001$. Overall, happiness and irritation were better decoded than neutral and sadness. There was no significant effect of singer, $F_{(6,\ 1245)} = 1.559, p = 0.156$.

There were significant 2-way interactions between epoch x emotion, $F_{(6,\ 1245)} = 3.520$, $p = 0.002$; emotion × singer, $F_{(18,\ 1245)} = 6.718$, $p < 0.001$; but not epoch × singer, $F_{(12,\ 1245)} = 0.742, p = 0.711$. The two-way interactions revealed that again, happiness was better decoded in the pre-production, $t_{(15)} = 4.91$, $p < 0.001$, and production epochs, $t_{(15)} = 3.26$, $p = 0.003$, as compared to the post-production epoch. Irritation was better decoded in the production epoch as compared to the post-production epoch, $t_{(15)} = 2.53$, $p = 0.033$. The singer by emotion interaction revealed that with the exception of Singers 2 and 7, most were able to communicate happiness. Singer 4 was best able to communicate irritation. Singers 1 and 7 were best able to communicate sadness.

There was also a significant 3-way interaction with epoch × singer × emotion, $F_{(36,\ 1245)} = 1.597$, $p = 0.014$. Tests of simple effects with Bonferroni correction showed that there were no significant differences across epochs for Singer 1 and Singer 6. The findings showed that Singer 4 was marginally better at communicating happiness in the pre-production and production epochs as compared to the post-production epoch, $t_{(15)} = 2.37, p = 0.053$. Singer 2 was marginally better at communicating happiness in the pre-production epoch than the production epoch and Singer 3 was also marginally better at communicating happiness in the pre-production epoch as compared to the post-production epoch, $t's_{(15)} = 2.37, p = 0.053$. Singer 4 was better able to communicate irritation in the pre-production, $t_{(15)} = 3.16, p = 0.005$, and the production epochs, $t_{(15)} = 2.37$, $p = 0.053$, as compared to the post-production epoch. Singer 7 was better able to communicate sadness in the pre-production, $t_{(15)} = 2.76$, $p = 0.017$, and production epochs, $t_{(15)} = 2.37$, $p = 0.053$ as compared to the post-production epochs.

Comparison of viewers' emotion ratings with PC movement analysis

Each of the principal components was used as a predictor of viewers' emotion ratings in multiple regression analyses. This analysis assessed the association between the singers' facial motion cues and the viewers' decoding of emotion. Presumably, emotional judgments should be associated with specific cues signaling emotion. **Table 2** shows the results of multiple regression analyses that predicted viewers' ratings in each emotion/epoch condition from the five principal component values of the facial movements. In the production epoch of each emotion, Component 1,

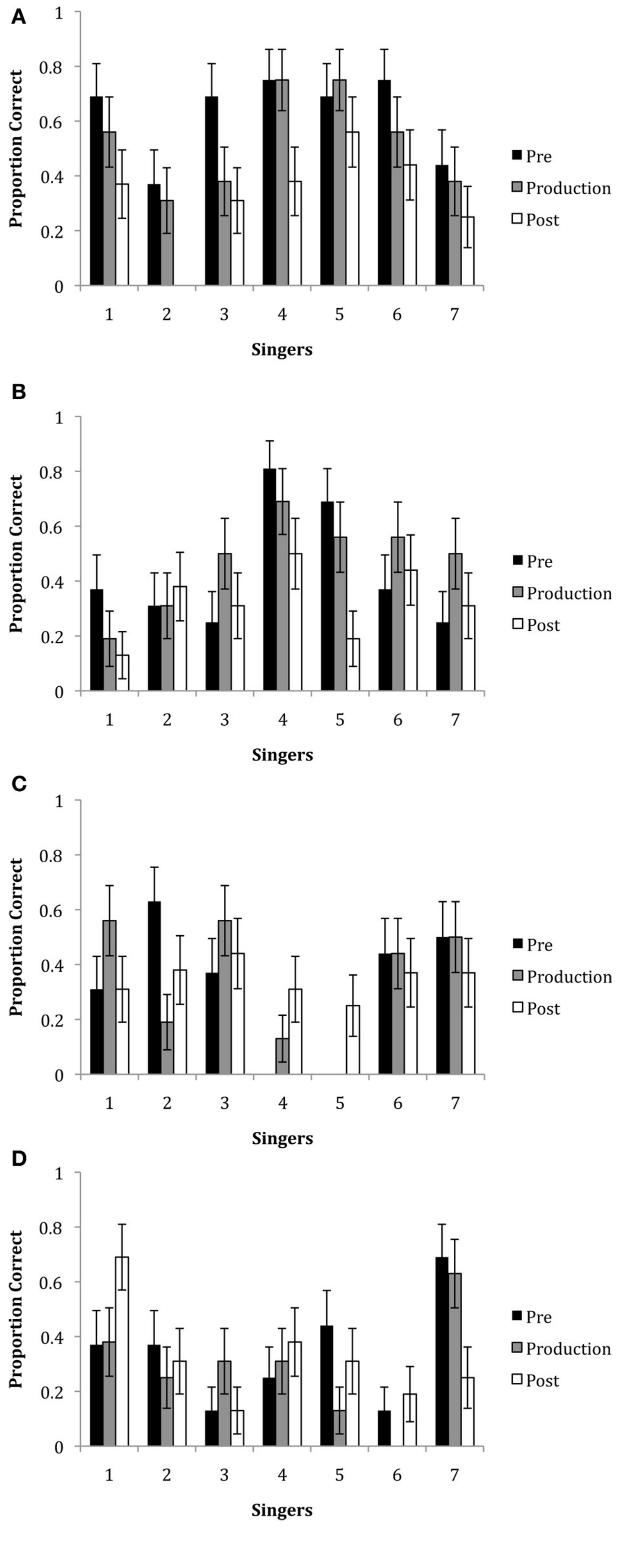


FIGURE 6 | The proportion of correct responses for the emotions of (A) happiness, (B) irritation, (C) neutral and (D) sadness for each of the seven singers in each of the three epochs using the exaggerated stimuli. Note that missing bars indicate that no participant accurately decoded the emotional intention.

which was associated with singers' mouth movements, predicted viewers' ratings of emotion. Expressions of happiness were generally associated with increased mouth movements and head displacement. Irritation was associated with increased mouth and eyebrow movements (Components 1 and 5) and this was particularly significant in the pre- and post-production epochs. Neutral expressions were generally associated with reduced movements in most of components. Finally, sadness was associated with some eyebrow movements, likely an upward movement of the inner eyebrow, head displacement and head rotation. An examination of the PLDs showed that head displacement reflected a downward motion of the head accompanied by a slight rotation.

Overall, perceivers were sensitive to specific cues in singers' facial movements for decoding emotion. The high decoding of happiness in the pre-production epoch may have occurred because there was considerable movement of the mouth corner in the pre-production epoch when compared to the production epoch. Perceivers associated eyebrow movements with the expression of irritation. The expression of neutral was associated with less movement overall. Sadness was associated with eyebrow movement and head displacement.

DISCUSSION

The findings of Experiment 2 suggest that emotional information can be conveyed in the PLDs of exaggerated face and head movements. The overall accuracy of emotional decoding appeared to be similar between Experiments 1 and 2. In Experiment 2, the emotions of happiness and irritation were particularly well decoded. The exaggerated movements may mainly disambiguate high-arousal emotional intentions (i.e., happiness and irritation) from low-arousal emotional intentions (i.e., neutral and sadness). Consistent with this interpretation, the high arousal emotions of irritation and happiness were generally well identified. In contrast, sadness was not well decoded.

There appeared to be differences in the decoding of exaggerated facial movements depending on the epoch. In particular, perceivers were able to decode happiness in several singers before they began singing. Perceivers were able to decode emotion from specific singers when they expressed irritation (Singer 4) and sadness (Singer 7) in the pre-production and production epochs as compared to the post-production epoch. These findings suggest that the expressive intentions of some singers were perceptible outside of the timeframe of singing when the movements were exaggerated.

EXPERIMENT 3

Experiment 2 demonstrated that facial movements associated with high-arousal emotions (happiness and irritation) were decoded accurately by viewers when they were algorithmically exaggerated in range of motion. Experiment 3 addressed whether the extent of motion, rather than the specific motion configuration associated with singing, communicates important emotional information. Although unlikely, it is possible that the decoding of emotion might have been based solely on the magnitude of motion information rather than the particular expressive movements. That is, it is possible that the coherence or configuration of the marker information may not have been necessary to decode emotion.

Experiment 3 presented "scrambled" motion configurations of the same facial movements used in previous experiments. The direction of marker movement commenced from a randomly determined position. The marker trajectory could be in any 360 degree direction. That is, the marker appeared in the neutral starting position, and then moved with the same acceleration, velocity, and distance as in the original stimulus but in a randomly determined trajectory that was independent of other markers. The manipulation was introduced to all stimuli from Experiments 1 and 2, and included both the original and exaggerated stimuli. It was predicted that perceivers would be sensitive to the overall amount of displacement, velocity, and acceleration of point-light movements, which may contain information about the level of arousal of the intended emotion. However, such a manipulation removes configural information (e.g., features associated with a smile or a frown), and motion coherence, which may be important for differentiating emotions that are similar to each other in their level of arousal. Previous work has demonstrated that the relative positions and timing of markers in PLDs are needed for accurate perception (e.g., Bertenthal and Pinto, 1994). In the absence of configuration information, we expected emotions to be less accurately decoded. However, because randomized movements should reflect overall arousal levels, we reasoned that viewers might be most accurate at decoding low arousal emotions such as sadness and neutral when the scrambled stimuli were not exaggerated, and most accurate at decoding high arousal emotions when the scrambled stimuli were exaggerated.

METHODS

Scrambled stimuli were created by randomly changing the orientation of movement as it appeared in the two-dimensional image plane. This occurred for the trajectories of all markers without respect to the rigidity or non-rigidity of the movements. The location of the first sample of each trajectory of each trial was used as the center of rotation and determined the randomly selected direction that the marker would travel. The first frame of each trial showed an unscrambled face, while in the following frames, the face tended to immediately disintegrate or jitter, due to the markers moving in random directions with the same amount of displacement, speed, and acceleration of the original trajectories. All other values associated with the movement were retained, including the maximum, minimum and standard deviations associated with individual marker movement. In many cases, the stimulus no longer resembled a moving face and head because individual marker movements no longer conformed to the configuration found in a face. The scrambling was applied to both types of trials—the original motion and the exaggerated motion.

RESULTS

Six hundred and seventy-two conditions were considered in this analysis (4 emotions × 4 phrases × 7 singers × 3 epochs × 2 exaggeration), with each viewer rating 168 trials (one phrase). There were more conditions in this analysis than in Experiments 1 or 2 because participants rated both the original stimuli and

Table 2 | The coefficient of determination, *F*-value and regression coefficients associated with the regression analyses with the components as predictors of emotion in each of the three epochs.

Emotion	Epoch	R^2	*F*	Components				
				1 Mouth	2 Head displacement/velocity	3 Side head motion	4 Head acceleration	5 Eyebrows
Happiness	Pre	0.22	11.99	0.15*	0.13*	−0.07	−0.09	0.01
	Production	0.31	19.96	0.08**	0.09*	−0.04	−0.09	0.03
	Post	0.20	10.59	0.17**	0.03	−0.01	−0.04	−0.10*
Irritation	Pre	0.16	8.12	−0.08	0.08	0.03	−0.13*	0.21***
	Production	0.24	13.87	0.09***	0.00	0.00	−0.04	0.01
	Post	0.14	7.33	0.03	0.02	−0.04	0.01	0.08*
Neutral	Pre	0.34	22.89	−0.11*	−0.18**	0.06	0.30***	−0.18**
	Production	0.34	22.45	−0.12***	−0.12**	0.06	0.17**	−0.04
	Post	0.31	19.78	−0.16*	−0.12**	0.05	0.24***	−0.18**
Sadness	Pre	0.10	4.54	−0.02	0.00	−0.03	−0.09	0.07
	Production	0.21	11.45	−0.05*	0.04	0.00	−0.04	−0.01
	Post	0.10	4.61	−0.07	0.12**	−0.12*	−0.09	0.21***

$^*p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

the exaggerated stimuli. As before, trials on which participants rated two emotions equally high (one matching the intended emotion and the other not matching the intended emotion) were considered incorrect ($n = 145$). **Table 3** displays the mean percent correct for each condition. Overall, the mean correct responses ($M = 28.23$, $SE = 6.37$) indicated that emotions were decoded at chance levels. A GLM analysis including the factors of epoch, emotion, and exaggeration with all interactions was performed. The results showed that there was a significant main effect of emotion, $F_{(2,\ 2664)} = 16.12$, $p = 0.001$, such that stimuli expressing happiness and neutral were better decoded than stimuli expressing neutral and sadness. There was a main effect of epoch, $F_{(2,\ 2664)} = 4.16$, $p = 0.02$, showing that stimuli in the production epoch were better decoded than the stimuli in the post-production epoch. There was no main effect of exaggerated stimuli, $F_{(1,\ 2664)} = 0.017$, $p = 0.89$, suggesting that exaggeration alone did not suggest any one particular emotion. However, there was a significant interaction between emotion and exaggeration, $F_{(3,\ 2664)} = 24.91$, $p = 0.001$. When happiness and irritation were exaggerated, these emotions were better decoded than the emotions of neutral and sadness. There was no significant interaction between epoch and emotion, $F_{(6,\ 2664)} = 0.68$, $p = 0.66$, and between epoch and exaggeration, $F_{(2,\ 2664)} = 0.62$, $p = 0.54$. The 3-way interaction between epoch, emotion and exaggerated was marginally significant, $F_{(36,\ 2664)} = 1.95$, $p = 0.07$. This reflected the trend that exaggerated stimuli expressing happiness were better decoded in the production epoch as compared to the post-production epoch.

DISCUSSION

The results of Experiment 3 show that when the relative global relationships between point-lights are removed and only motion information is maintained, perceivers had difficulty decoding the emotional expression. This finding suggests that range, velocity, and acceleration of motion (which were the same for scrambled and biological movements) were not the sole determinants of viewers' emotional ratings of singers' facial movements. Instead, specific configural information and motion coherence about facial movements, which was lost in the scrambled versions, guided viewers to more accurate ratings of the biological facial movements. The manipulation also helped to clarify some of the strategies used by participants as they attempted to decode emotional intentions. In the original condition, participants tended to assign high ratings of neutral expression, possibly because they found the movements to be ambiguous or uninterpretable. However, when movement was exaggerated, viewers were likely to perceive the emotions of happiness and irritation in the production epoch. This finding suggests that the arousal level of an emotion can be conveyed in the absence of relative global information, but only when the magnitude of the motion information is obvious and over a longer duration (as in the production epoch). Despite the presentation of the experimental conditions being counterbalanced, one possibility is that because the same observers participated in the experiments they were aware that greater motion was associated with happiness and irritation. We do not believe that this is likely due to the higher accuracy for happiness and irritation in only the exaggerated condition. If participants were primed, then we would have expected higher decoding accuracy in the original condition for these emotions.

GENERAL DISCUSSION

In three experiments, we examined the communication of singers' emotions based on facial movements before, during, and immediately after singing. The findings suggest that singers use facial expressions and head movements in ways that correlate with the intended emotion. Perceivers, in turn, interpret the movements used by singers and can decode intended emotions. However, accurate decoding depends on the intended emotion, the epoch, and the singer. The emotional connotations of certain movements can be clarified when the recorded movements are exaggerated,

Table 3 | The average accuracy ratings for each emotion and epoch for the original and exaggerated scrambled stimuli.

Emotion	Epoch					
	Pre		Production		Post	
	Spatial type		Spatial type		Spatial type	
	Original	Exaggerated	Original	Exaggerated	Original	Exaggerated
Happiness	25.89 (44.00)	**39.28** (49.01)	25.89 (44.00)	**46.42** (50.09)	26.78 (44.48)	28.57 (45.37)
Irritation	25.00 (43.49)	25.89 (44.00)	20.53 (40.57)	**38.39** (48.85)	11.61 (32.18)	30.35 (46.18)
Neutral	**41.07** (49.42)	23.21 (49.41)	**52.68** (50.15)	27.67 (44.94)	**49.99** (50.22)	21.42 (41.22)
Sadness	22.32 (41.83)	20.53 (40.57)	23.21 (42.41)	19.64 (39.91)	18.75 (39.21)	19.64 (39.91)

Standard deviations are in parentheses. Values in bold indicate above chance decoding.

especially for high-arousal emotions. Exaggerating the movements associated with a low-arousal emotion, however, can suggest a high-arousal emotion, leading to lower decoding accuracy. Removal of configural cues (through randomizing movement) leads to low decoding rates, suggesting the importance of configural cues; however, overall arousal information may be preserved in randomized movement. Finally, significant individual differences were observed: singers differed from each other in their use of facial and head movements in pre-production, production, and post-production epochs, leading to differential decoding rates for each singer at different temporal epochs. These findings will be discussed in turn.

First, our results corroborate a growing body of evidence that singers use facial expressions and head movements in ways that correlate with expressive intentions, including emotional intentions (Davidson, 1993, 1994, 1995; Thompson et al., 2005, 2008, 2010; Livingstone et al., 2009). The findings from Experiments 1 and 2 showed that perceivers were sensitive to eyebrow and lip movements. Inspection of the videos showed that singers frowned or raised their eyebrows to signal irritation and sadness respectively, and smiled to signal happiness. For all singers, the amount of displacement, velocity, and acceleration varied as a function of the intended emotion. Happiness was associated with higher values on all these motion variables whereas sadness was associated with lower values. These commonalities between individual musicians allow perceivers to use consistent strategies when decoding emotional intentions. This finding is consistent with Brunswik's Lens model in that some cues must be common amongst all senders for receivers to be able to decode the senders' intentions.

Head movements were also used to express emotion although perceiver's judgments seemed not to significantly associate head motion with any one particular emotion. Head movements make performances more natural, expressive, and signal cycles of tension and relaxation (Wanderley et al., 2005; Busso et al., 2007; Castellano et al., 2008). Moreover head movements alone have been found to communicate emotion (Dahl and Friberg, 2007).

Second, we observed that perceivers could interpret emotional information from face and head movements not only during singing, but prior to the onset of singers' vocalization. In Experiments 1 and 2, viewers were able to decode happiness before singing commenced. This effect, however, did not reliably extend to the other emotional intentions and was not evident in the post-production epoch.

Currently, more research is needed to better understand the phenomenon of emotional preparation and lingering. Our findings are consistent with Livingstone et al. (2009), in that across epochs, musicians used movements as a form of expression. The current study found that perceivers could not meaningfully use the information in the post-production epoch to decode emotions. There are a few possibilities for this outcome. The first is that singers used movements in the post-production epoch but not to the same extent as in the pre-production and production epochs, particularly mouth and eyebrow movements. A second is that moving "away" from an emotion is unnatural for perceivers to decode. Some evidence for this possibility comes from the finding that participants are poorer at decoding emotion from full-video sequences that are shown backwards as compared to forwards (Cunningham and Wallraven, 2009; Experiment 4). Firm comparisons between Livingstone et al. (2009) and our work are difficult to make due to fundamental differences in methods (i.e., the use of point-light displays vs. videos; assessment of emotional decoding vs. emotional mimicry in viewers).

Third, the results showed that exaggerating movements sometimes assisted in emotional decoding although the manipulation may have distorted the emotional expression. In Experiment 2, the emotions of happiness and irritation were well decoded as compared to neutral and sadness. The high decoding of these emotions may be attributable to high arousal emotions being associated with greater displacement, velocity, and acceleration than low arousal emotions. These findings are also consistent with past work demonstrating that exaggerated movements lead to higher ratings of emotional intensity (Pollick et al., 2003). The data suggest that exaggeration of the motion did not benefit the decoding of sadness. When movements expressing sadness were exaggerated in Experiment 2, the movements were often no longer consistent with the expression of this emotion in some singers. Accurate decoding of sadness may rely on information that is consistent with the expected motion information. For example, the expression of sadness unfolds more slowly than other emotions and when its development speeds up, it is no longer perceived as natural (Kamachi et al., 2001; Sato and Yoshikawa, 2004). In our stimuli, sadness was only well decoded for musicians who showed minimal movement and for whom the exaggeration manipulation would not have affected as strongly.

Fourth, the poor decoding accuracy observed in Experiment 3 confirms that configural cues and information associated with the direction of movement for individual markers are important for accurate decoding. Motion data alone (displacement/distance travelled, velocity and acceleration) may help viewers to differentiate emotional vs. non-emotional stimuli, but these data do not appear to provide sufficient information for accurate decoding. Our results seemed to show that greater degrees of movement were associated with higher arousal—analogous to findings for inverted point-light biological motion (Dittrich et al., 1996; Clarke et al., 2005). Scrambling movements may have other unintended effects though. It is possible that without the configural information and coherence between individual marker movements, participants were not able to effectively make use of motion information. That is, while the distance travelled, velocity and acceleration for each marker was the same in the scrambled conditions as in the biological conditions, participants may be even more disadvantaged by motion not being biologically possible or coherent. The interaction between motion and form information is still an area of debate. Thirkettle et al. (2009), using PLDs, found that both form and motion information were important to discrimination of human motion. Future work comparing various control conditions may reveal whether or not scrambling motion is an effective control condition or has other unintended effects (Hiris, 2007; Thirkettle et al., 2009).

Fifth, there were individual differences in emotional expression. For example, Singers 4 and 6 were generally able to express irritation and happiness more clearly than other musicians. Yet interestingly, viewers did not seem to be able to decode the emotion of sadness when expressed by these singers. This may be due to their use of larger movements. Singer 4 also exhibited highly expressive eyebrows and control over corrugator supercilli and procerus muscles—those involved in frowning (Ekman and Friesen, 1976). In contrast, perceivers tended to identify sadness more clearly when communicated by Singers 1, 3, 5, and 7. However, this depended on the epoch. One common factor was that these singers used reduced movements and showed specific facial or head cues, such as raised eyebrows. These findings show individual differences in emotional decoding. The strategies adopted by some individuals may have enhanced their ability to express some emotions at the expense of others. With many signals present, the cues were used probabilistically but perceivers may have had difficulty ignoring idiosyncratic movements when decoding emotion.

The modest decoding accuracy in the pre- and post-production epochs might be contrasted with the rich auditory and visual information musicians are able to use in performances. Point-light displays of a second and a half in duration are highly impoverished relative to full videos (even if considerable information is conveyed; Blake and Shiffrar, 2007). Decoding emotions from full-face, synthesized dynamic motions may take as long as 2.5 to 3 s for happiness and disgust, respectively (Gutiérrez-Maldonado et al., 2014). Controlled exposures to static images reveal that happiness can be decoded after 25 ms but that other emotions require more time. When free responses are measured, at least a second is required to accurately decode emotion from static images (Calvo and Lundqvist, 2008). Furthermore, comparisons of emotional decoding for PLDs are generally much lower than would be expected for full static images showing the facial expression (Bassili, 1979).

To conclude, musicians used facial and head movements to communicate emotions, and viewers were generally sensitive to these signals. There are idiosyncratic patterns in the use of these movements, and their development over time. Musicians can use pre- and post-production facial movements to supplement and surround the acoustic channel to support emotional communication. These expressions may be especially important given that movements during vocalization are heavily constrained by production. However, the influence of facial movements may vary from study to study and between individual musicians. When movements were artificially exaggerated, high arousal emotions were better expressed but low arousal emotions were more poorly expressed. Again, there were exceptions to this rule. Perceivers interpret overall movements in terms of general levels of arousal, while configural cues may provide detailed information about specific emotional intentions. The facial expressions of musicians are combined with the auditory domain to provide a rich audiovisual experience for listeners. This audiovisual expression of emotion may act to facilitate social interaction in daily life, but in music, it may highlight the emotional, expressive, and musical goals of the performer.

ACKNOWLEDGMENTS

This project was funded by a Discovery Grant DP0987182 from the Australian Research Council awarded to William F. Thompson and Caroline Palmer. We would like to thank Rachel Bennetts, Julia Case, Alex Chilvers, Catherine Greentree, Shaun Halovic, and Bojan Neskovic for technical support.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: http://www.frontiersin.org/journal/10.3389/fpsyg.2014.00262/abstract

REFERENCES

Bassili, J. N. (1978). Facial motion in the perception of faces and of emotional expression. *J. Exp. Psychol. Hum. Percept. Perform.* 4, 2049–2058. doi: 10.1037/0096-1523.4.3.373

Bassili, J. N. (1979). Emotion recognition: the role of facial movement and the relative importance of upper and lower areas of the face. *J. Pers. Soc. Psychol.* 37, 2049–2058. doi: 10.1037/0022-3514.37.11.2049

Behne, K., and Wöllner, C. (2011). Seeing or hearing the pianists? A synopsis of an early audiovisual perception experiment and a replication. *Musicae Sci.* 15, 324–342. doi: 10.1177/1029864911410955

Bertenthal, B. I., and Pinto, J. (1994). Global processing of biological motions. *Psychol. Sci.* 5, 221–225. doi: 10.1111/j.1467-9280.1994.tb00504.x

Blake, R., and Shiffrar, M. (2007). Perception of human motion. *Annu. Rev. Psychol.* 58, 47–74. doi: 10.1146/annurev.psych.57.102904.190152

Brunswik, E. (1956). *Perception and the Representative Design of Psychological Experiments, 2nd Edn.* Berkeley, CA: University of California Press.

Busso, C., Deng, Z., Grimm, M., Neumann, U., and Narayanan, S. (2007). Rigid head motion in expressive speech animation: analysis and synthesis. *IEEE Trans. Audio Speech Lang. Process.* 15, 1075–1087. doi: 10.1109/TASL.2006.885910

Calvo, M. G., and Lundqvist, D. (2008). Facial expressions of emotion (KDEF): identification under different display conditions. *Behav. Res. Methods* 40, 109–115. doi: 10.3758/BRM.40.1.109

Castellano, G., Mortillaro, M., Camurri, A., Volpe, G., and Scherer, K. (2008). Automated analysis of body movement in emotionally expressive piano performances. *Music Percept.* 26, 103–119. doi: 10.1525/mp.2008.26.2.103

Ceaser, D. K., Thompson, W. F., and Russo, F. A. (2009). Expressing tonal closure in music performance: auditory and visual cues. *Can. Acoust.* 37, 29–34.

Chan, L., Russo, F. A., and Livingstone, S. (2013). Automatic facial mimicry of emotion during perception of song. *Music Percept.* 30, 361–367. doi: 10.1525/mp.2013.30.4.361

Clarke, T. J., Bradshaw, M. F., Field, D. T., Hampson, S. E., and Rose, D. (2005). The perception of emotion from body movement in point-light displays of interpersonal dialogue. *Perception* 34, 1171–1180. doi: 10.1068/p5203

Cunningham, D. W., and Wallraven, C. (2009). Dynamic information for the recognition of conversational expressions. *J. Vis.* 9, 1–17. doi: 10.1167/9.13.7

Dahl, S., and Friberg, A. (2007). Visual perception of expressiveness in musicians' body movements. *Music Percept.* 24, 433–454. doi: 10.1525/mp.2007.24.5.433

Davidson, J. W. (1993). Visual perception and performance manner in the movements of solo musicians. *Psychol. Music* 21, 103–113. doi: 10.1177/030573569302100201

Davidson, J. W. (1994). Which areas of a pianist's body convey information about expressive intention to an audience? *J. Hum. Mov. Stud.* 26, 279–301.

Davidson, J. W. (1995). "What does the visual information contained in music performances offer the observer? Some preliminary thoughts," in *Music and the Mind Machine: Psychophysiology and Psychopathology of the Sense of Music,* ed R. Steinberg (Heidelberg: Springer), 105–114. doi: 10.1007/978-3-642-79327-1_11

Davidson, J. W. (2012). Bodily movement and facial actions in expressive music performance by solo and duo instrumentalists: two distinctive case studies. *Psychol. Music* 40, 595–633. doi: 10.1177/0305735612449896

Dittrich, W. H., Troscianko, T., Lea, S. E., and Morgan, D. (1996). Perception of emotion from dynamic point-light displays represented in dance. *Perception* 25, 727–738. doi: 10.1068/p250727

Ekman, P., and Friesen, W. V. (1971). Constants across cultures in the face and emotion. *J. Pers. Soc. Psychol.* 17, 124–129. doi: 10.1037/h0030377

Ekman, P., and Friesen, W. V. (1976). Measuring facial movement. *Environ. Psychol. Nonverbal Behav.* 1, 56–75. doi: 10.1007/BF01115465

Elfenbein, H. A., and Ambady, N. (2002). On the universality and cultural specificity of emotion recognition: a meta-analysis. *Psychol. Bull.* 128, 203–235. doi: 10.1037/0033-2909.128.2.203

Fiorentini, C., Schmidt, S., and Viviani, P. (2012). The identification of unfolding facial expressions. *Perception* 41, 532–555. doi: 10.1068/p7052

Gabrielsson, A., and Lindström, E. (2010). "The role of structure in the musical expression of emotions," in *Music and emotion*: *Theory and Research, 2nd Edn.,* eds P. Juslin and J. Sloboda (Oxford: Oxford University Press), 368–400.

Gower, J. C. (1975). Generalized procrustes analysis. *Psychometrika* 40, 33–51. doi: 10.1007/BF02291478

Gutiérrez-Maldonado, J., Rus-Calafell, M., and González-Conde, J. (2014). Creation of a new set of dynamic virtual reality faces for the assessment and training of facial emotion recognition ability. *Virtual Real.* 18, 61–71. doi: 10.1007/s10055-013-0236-7

Halovic, S., and Kroos, C. (2009). "Facilitating the perception of anger and fear in male and female walkers," in *Symposium on Mental States, Emotions and their Embodiment* (Edinburgh).

Hevner, K. (1935). The affective character of the major and minor modes in music. *Am. J. Psychol.* 47, 103–118. doi: 10.2307/1416710

Hiris, E. (2007). Detection of biological and nonbiological motion. *J. Vis.* 7, 1–16 doi: 10.1167/7.12.4

Johnstone, T., and Scherer K. R. (2000). "Vocal communication of emotion," in *The Handbook of Emotions,* eds M. Lewis and J. Haviland (New York, NY: Guildford Press), 226–235.

Juslin, P. N. (2000). Cue utilization in communication of emotion in music performance: relating performance to perception. *J. Exp. Psychol. Hum. Percept. Perform.* 26, 1797–1813. doi: 10.1037/0096-1523.26.6.1797

Juslin, P. N., and Laukka, P. (2003). Communication of emotions in vocal expression and music performance: different channels, same code? *Psychol. Bull.* 129, 770–814. doi: 10.1037/0033-2909.129.5.770

Juslin, P. N., and Sloboda, J. (2010). *Music and Emotion: Theory and Research, 2nd Edn.* Oxford: Oxford University Press.

Kamachi, M., Bruce, V., Mukaida, S., Gyoba, J., Yoshikawa, S., and Akamatsu, S. (2001). Dynamic properties influence the perception of facial expressions. *Perception* 30, 875–887. doi: 10.1068/p3131

Livingstone, S. R., Thompson, W. F., and Russo, F. A. (2009). Facial expressions and emotional singing: a student of perception and production with motion capture and electromyography. *Music Percept.* 26, 475–488. doi: 10.1525/MP.2009.26.5.475

Palmer, C. (2012). "Music performance: movement and coordination," in *The Psychology of Music, 3rd Edn.,* ed D. Deutsch (Amsterdam: Elsevier Press), 405–422.

Platz, F., and Kopiez, R. (2012). When the eye listens: a meta-analysis of how audio-visual presentation enhances appreciation of music performance. *Music Percept.* 30, 71–83. doi: 10.1525/mp.2012.30.1.71

Pollick, F. E., Hill, H., Calder, A., and Paterson, H. (2003). Recognising facial expression from spatially and temporally modified movements. *Perception* 32, 813–826. doi: 10.1068/p3319

Ramsay, J., and Silverman, B. W. (2005). *Functional Data Analysis*, *2nd Edn.* New York, NY: Springer-Verlag.

Recio, G., Schacht, A., and Sommer, W. (2013). Classification of dynamic facial expressions of emotion presented briefly. *Cogn. Emot.* 27, 1486–1494. doi: 10.1080/02699931.2013.794128

Russell, J. A., Bachorowski, J., and Fernandez-Dols, J. (2003). Facial and vocal expressions of emotion. *Annu. Rev. Psychol.* 54, 329–349. doi: 10.1146/annurev.psych.54.101601.145102

Saldaña, H. M., and Rosenblum, L. D. (1993). Visual influences on auditory pluck and bow judgments. *Percept. Psychophys.* 54, 406–416. doi: 10.3758/BF03205276

Sato, W., and Yoshikawa, S. (2004). The dynamic aspects of emotional facial expressions. *Cogn. Emot.* 18, 701–710. doi: 10.1080/02699930341000176

Scherer, K. R., Johnstone, T., and Klasmeyer, G. (2003). "Vocal expression of emotion," in *The Handbook of the Affective Sciences,* eds R. J. Davidson, K. R. Scherer, and H. Goldsmith (New York, NY: Oxford University Press), 433–456.

Schubert, E. (2004). Modeling perceived emotion with continuous musical features. *Music Percept.* 21, 561–585. doi: 10.1525/mp.2004.21.4.561

Schutz, M., and Lipscomb, S. (2007). Hearing gestures, seeing music: vision influences perceived tone duration. *Perception* 36, 888–897. doi: 10.1068/p5635

Scotto di Carlo, N., and Guaitella, I. (2004). Facial expressions of emotion in speech and singing. *Semiotica* 49, 37–55. doi: 10.1515/semi.2004.036

Slaney, M. (1999). MakeQTMovie-Create QuickTime movies in Matlab. *Interval Technical Report,* 1999–066.

Thompson, W. F., and Balkwill, L.-L. (2010). "Cross-cultural similarities and differences," in *Music and Emotion: Theory and Research, 2nd Edn.,* eds P. Juslin and J. Sloboda (Oxford: Oxford University Press), 755–788.

Thompson, W. F., Graham, P., and Russo, F. A. (2005). Seeing music performance: Visual influences on perception and experience. *Semiotica* 156, 177–201. doi: 10.1515/semi.2005.2005.156.203

Thompson, W. F., and Russo, F. A. (2007). Facing the music. *Psychol. Sci.* 18, 756–757. doi: 10.1111/j.1467-9280.2007.01973.x

Thompson, W. F., Russo, F. A., and Livingstone, S. L. (2010). Facial expressions of singers influence perceived pitch relations. *Psychon. Bull. Rev.* 17, 317–322. doi: 10.3758/PBR.17.3.317

Thompson, W. F., Russo, F. A., and Quinto, L. (2008). Audio-visual integration of emotional cues in song. *Cogn. Emot.* 22, 1457–1470. doi: 10.1080/02699930701813974

Thirkettle, M., Benton, C. P., and Scott-Samuel, N. E. (2009). Contributions of form, motion and task to biological motion perception. *J. Vis.* 9, 1–11. doi: 10.1167/9.3.28.

Timmers, R., and Ashley, R. (2007). Emotional ornamentation in performances of a Handel sonata. *Music Percept.* 25, 117–134. doi: 10.1525/mp.2007.25.2.117

Vines, B. W., Krumhansl, C. L., Wanderley, M. M., and Levitin, D. J. (2006). Cross-modal interactions in the perception of music performance. *Cognition* 101, 80–113. doi: 10.1016/j.cognition.2005.09.003

Wanderley, M. M., Vines, B. W., Middleton, N., McKay, C., and Hatch, W. (2005). The musical significance of clarinetists' ancillary gestures: an exploration of the field. *J. New Music Res.* 34, 97–113. doi: 10.1080/09298210500124208

Williamon, A., and Davidson, J. W. (2002). Exploring co-performer communication. *Musicae Sci.* 1, 53–72. doi: 10.1177/102986490200600103

Wöllner, C. (2008). Which part of the conductor's body conveys most expressive information? A spatial occlusion approach. *Musicae Sci.* 12, 249–272.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 04 September 2013; accepted: 11 March 2014; published online: 29 April 2014.

Citation: Quinto LR, Thompson WF, Kroos C and Palmer C (2014) Singing emotionally: a study of pre-production, production, and post-production facial expressions. Front. Psychol. ***5****:262. doi: 10.3389/fpsyg.2014.00262*

This article was submitted to Cognitive Science, a section of the journal Frontiers in Psychology.

frontiers in
PSYCHOLOGY

ORIGINAL RESEARCH ARTICLE
published: 07 March 2014
doi: 10.3389/fpsyg.2014.00156

The influence of vocal training and acting experience on measures of voice quality and emotional genuineness

Steven R. Livingstone[1,2] *, Deanna H. Choi[3] and Frank A. Russo[1,2]

[1] Department of Psychology, Ryerson University, Toronto, ON, Canada
[2] Toronto Rehabilitation Institute, Toronto, ON, Canada
[3] Department of Psychology, Queen's University, Kingston, ON, Canada

Edited by:
Bruno Gingras, University of Vienna, Austria

Reviewed by:
Victoria Williamson, Goldsmiths, University of London, UK
Dan Bowling, University of Vienna, Austria

***Correspondence:**
Steven R. Livingstone, Department of Psychology, Ryerson University, 350 Victoria Street, Toronto, ON M5B 2K3, Canada
e-mail: steven.livingstone@ryerson.ca

Vocal training through singing and acting lessons is known to modify acoustic parameters of the voice. While the effects of singing training have been well documented, the role of acting experience on the singing voice remains unclear. In two experiments, we used linear mixed models to examine the relationships between the relative amounts of acting and singing experience on the acoustics and perception of the male singing voice. In Experiment 1, 12 male vocalists were recorded while singing with five different emotions, each with two intensities. Acoustic measures of pitch accuracy, jitter, and harmonics-to-noise ratio (HNR) were examined. Decreased pitch accuracy and increased jitter, indicative of a lower "voice quality," were associated with more years of acting experience, while increased pitch accuracy was associated with more years of singing lessons. We hypothesized that the acoustic deviations exhibited by more experienced actors was an intentional technique to increase the genuineness or truthfulness of their emotional expressions. In Experiment 2, listeners rated vocalists' emotional genuineness. Vocalists with more years of acting experience were rated as more genuine than vocalists with less acting experience. No relationship was reported for singing training. Increased genuineness was associated with decreased pitch accuracy, increased jitter, and a higher HNR. These effects may represent a shifting of priorities by male vocalists with acting experience to emphasize emotional genuineness over pitch accuracy or voice quality in their singing performances.

Keywords: singing, emotion, emotional genuineness, acting, training, individual differences, voice quality, linear mixed models

The goals of a singer are varied and many: accurate pitch reproduction, desired voice quality, clear intelligibility, precise timing, and intended emotional inflection; these factors are not independent, and how they are prioritized may reflect differences in the training and experience of a performer (Ostwald, 2005; Bunch, 2009). Two types of training that may differentially affect vocal acoustic goals are singing training and acting experience. Numerous studies have investigated the acoustics of the expert singing voice (Sundberg, 2003), and the effects of short-term training on singing acoustics (Smith, 1963; Brown et al., 2000; Awan and Ensslen, 2010). The acoustic qualities of the trained actor's speaking voice have also been studied, though less extensively (Nawka et al., 1997; Bele, 2006), as have the effects of short-term acting training on speech acoustics (Timmermans et al., 2005; Walzak et al., 2008). To the authors' knowledge, there has only been one study that has considered the influence of acting training on acoustic measures of voice quality (Walzak et al., 2008). In addition, there are no studies of which we are aware that have compared the relative amounts of singing training and acting experience on the acoustics or perception of the singing voice. This is peculiar given the popularity of opera and musical theater, which often require both singing and acting experience. Amongst vocalists with a high level of acting experience, there may be a reprioritization of vocal goals toward emotional genuineness over pitch accuracy or voice quality. In contrast, vocalists with more years of singing training may instead prioritize pitch accuracy and voice quality. In this paper we sought to examine the relationship between acting experience and singing training on the acoustics and perception of the male singing voice.

Pitch accuracy may be considered one of the most salient perceptual dimensions on which we rate the quality of the singing voice. In a national survey of singing pedagogues, intonation, the ability to sing in tune, was regarded as the most important factor in assessing singing talent (Watts et al., 2003). Trained singers are able to reproduce known melodies with a high degree of pitch accuracy, varying between 30 to 42 cents on average (Larrouy-Maestri et al., 2013). Pitch accuracy in the general population has received considerable interest within the last 10 years (for a review, see Hutchins and Peretz, 2012). Although untrained singers can be quite accurate in terms of pitch when singing familiar and unfamiliar tunes (Dalla Bella et al., 2007; Pfordresher et al., 2010), they fare worse than trained singers when producing single pitches; deviating on average by 1.3 semitones from the target pitch compared to 0.5 semitones for trained singers (Ternstrom et al., 1988; Amir et al., 2003; Hutchins and Peretz, 2012). Non-musicians have also been characterized as being "imprecise," as their fundamental frequency (F_0) for a given pitch can vary across repeated productions (Pfordresher et al., 2010). Thus, the effect of singing training on pitch accuracy appears to depend on the musical context; that is, melodies vs. single pitches.

Where inaccurate pitch production occurs is likely to vary with the structure of the melody. One likely candidate though is the first note of the melody. In a study of untrained child vocalists and trained adult singers, Howard and Angus (1997) found that children were most inaccurate in the pitch of the first note of the melody. In the present study we also examine pitch measures of the first note. How pitch inaccuracy is quantified is an important methodological decision. During vocalization, the rapid opening and closing of the glottis produces a dynamic F_0 contour that varies over time (Fujisaki, 1983). While mean F_0 is often reported, this measure does not capture the range of vocalized F_0. In this study we examine the mean, minimum (floor), and maximum (ceiling) F_0 of the first note in an effort to capture the true range of pitch accuracy. What causes inaccurate pitch production is not fully understood, though it is thought that issues related to voice training, such as poor air support, vocal tension, lack of energy, and poor voice placement are determining factors and that pitch accuracy improves through singing training (Telfer, 1995; Willis and Kenny, 2008). However it remains unclear whether other forms of artistic experience, specifically acting experience, have an effect on singing pitch accuracy. One phenomenon in which acting experience may play a role is through the reprioritization of pitch accuracy during *phrasing*.

In musical theater, phrasing has been described as "the singer's personal stamp on the song," where "one performer may sing the lyric with absolute fidelity to the song as written, singing it pitch for pitch, . . . while another singer may absolutely transform the same song through her variations" (Deer and Dal Vera, 2008, p. 226). Taylor (2012, p. 34) writes that "performers are not completely circumscribed by the musical text in the meanings and emotions they communicate, as intonation, dynamic range and pitch are relative concepts that are stylistically interpreted." Thus, phrasing has been suggested to include changes to the intonation, intensity, and pitch from that of the notated score, with the effect of tailoring the meaning and emotions communicated to the individual desires of the singer. As vocalists gain greater acting experience, they may work to refine or emphasize their individuality, which may lead to an increase in deviations from the notated score. Thus, vocalists with a high level of acting experience may deviate more from the notated score than vocalists with less acting experience. Where in the melody these intentional deviations may occur is unknown. However, the first note of the melody is again a likely candidate, as any such deviation at this point would be particularly salient to the listener and may set up expectations about the quality or nature of the ensuing performance.

Artistic phrasing may encompass a broader range of perturbations than pitch and intensity, and include factors related to the perception of "voice quality." Two acoustic measures that are thought to index the perception of voice quality are jitter (Juslin and Laukka, 2001) and harmonics-to-noise ratio (HNR). The set of acoustic measures thought to capture vocal quality is debated (Raphael et al., 2011). Other perceptual qualities, such as harshness, tenseness, and creakiness have also been implicated in affecting voice quality (Gobl and Nì Chasaide, 2003). Jitter refers to fine-scale perturbations in F_0 caused by variations in the glottal pressure cycle (Lieberman, 1961; Scherer, 1989). HNR is a measure of the amount of noise in phonation, and refers to the ratio of energy contained at harmonics of F_0 compared to energy that is not (noise; Yumoto et al., 1982). Jitter and HNR are used to assess vocal pathology, with older and pathologically "rough" voices characterized by higher jitter and lower HNR values (Wilcox and Horii, 1980; Ferrand, 2002). HNR has also been associated with the perception of vocal attractiveness (Bruckert et al., 2010). Our investigation examined these spectral features in male vocalists. Previous research suggests that the presence or absence of the "singer's formant," a characteristic peak near 3 kHz in the vocal energy spectrum, varies across genders and may be absent in higher female voices (Bartholomew, 1934; Sundberg, 1974; Weiss et al., 2001). As these differences may have added additional variance to our spectral measures, our investigation focused on male vocalists. We operationalize phrasing as deviations from the notated score (e.g., F_0 accuracy, intonation), as well as spectral perturbations of the voice that relate to voice quality.

How a performer's use of phrasing may affect the perception of the singing voice is unknown, though one candidate is emotional genuineness (Krumhuber and Kappas, 2005; Langner et al., 2010; Scherer et al., 2013). Genuineness refers to the degree to which a listener or observer thinks or feels the vocalist's expression is a truthful reflection of the vocalist's physiological, mental, and emotional state. This quality is of particular importance to actors, who use the pejorative term *indicating* to refer to a non-truthful performance. Katselas (2008, p. 109) writes that "to indicate is to show, I repeat, *show* the audience emotion, character through external means . . . without really feeling or experiencing the moment. It's a token, a symbol, an indication, the shell of the thing without internal connection or actual experience." We hypothesize that vocalists with greater acting experience may sacrifice accurate singing production and voice quality, as measured through increased F_0 deviations, more jitter, and a lower HNR, to achieve greater levels of emotional genuineness.

In this paper we report two experiments that examined the relationships between the relative amounts of acting and singing experience on the acoustics and perception of the singing voice. The first experiment involved acoustical analyses of short phrases that were sung with different emotions and intensities. We expected that vocalists with more years of acting experience would show decreased pitch accuracy, with an F_0 (mean, floor, ceiling) further from the target note pitch, and lower voice quality (increased jitter, lower HNR), relative to vocalists with fewer years of acting training. We also expected that vocalists with more years of singing training would exhibit increased pitch accuracy, with an F_0 (mean, floor, ceiling) closer to the target note pitch, and potentially higher voice quality (higher average HNR, decreased jitter), relative to vocalists with fewer years of singing training. The second experiment examined listeners' perception of emotional genuineness from vocalist's singing performances. Listeners rated the emotional genuineness of recordings that were used in Experiment 1. We expected that vocalists with more years of acting experience would be rated as more emotionally genuine, and that these ratings would be associated with increased F_0 deviations, more jitter, and a lower HNR.

In both experiments we examined these relationships using repeated measures linear mixed models (LMMs). This form of analysis is particularly suited to a repeated measures design where covariates are of interest, as the use of repeated measures in traditional multiple regression violates the assumption of independence (Bland and Altman, 1994). LMMs also offer advantages over linear regression and analyses of covariance, allowing for the specification of random intercepts, with the fitting leading to independent intercepts for each vocalist or listener.

EXPERIMENT 1

Participants were required to sing short statements with five different emotional intentions (calm, happy, sad, angry, and fearful) and two intensities (normal, strong) while having their vocal productions recorded. We predicted that vocalists with more years of acting experience would produce a less pitch-accurate performance, have a lower HNR and more jitter – indicative of lower voice quality – relative to vocalists with fewer years of acting experience. We also predicted that more highly trained singers, as indexed by their years of singing lessons, would produce a more pitch-accurate performance, a higher HNR, and less jitter – indicative of higher voice quality – relative to vocalists with fewer years of singing training. We selected years of acting experience over acting lessons, as actors' primary form of training in our sample was through active drama performance.

METHOD

Participants

Twelve male vocalists (mean age = 26.3, SD = 3.8) with varying amounts of private or group singing lessons ($M = 4.8$, SD = 3.7), and varying levels of acting experience ($M = 10.8$, SD = 4.0), were recruited from the Toronto acting community. A correlation of vocalists' years of singing lessons with their years of acting experience was not significant $r(10) = 0.07$, $p = 0.84$, indicating there was no relationship between extent of training in the two domains of interest. Normality of the data were also confirmed with Shapiro–Wilk tests on age ($p > 0.05$), years of acting experience ($p > 0.05$), and years of singing lessons ($p > 0.05$). Participants were native English speakers, and were paid $50 CAD for their participation.

Stimuli and apparatus

Two neutral English statements were used ("Kids are talking by the door," "Dogs are sitting by the door"). Statements were seven syllables in length and were matched in word frequency and familiarity using the MRC psycholinguistic database (Coltheart, 1981). Two isochronous melodies were used; one for the positively valenced emotions, calm and happy (F3, F3, A3, A3, F3, E3, F3), and one for the negatively valenced emotions, sad, angry, and fearful (F3, F3, A^{b}3, A^{b}3, F3, E3, F3). Both melodies used piano MIDI tones of fixed acoustic intensity, consisting of six eighth notes (300 ms) and ending with a quarter note (600 ms), and were encoded at 16 bit/48 kHz (wav format). Positively and negatively valenced melodies were in the major and minor modes respectively (Dalla Bella et al., 2001).

The stimulus timeline consisted of three main epochs: Task presentation (4500 ms), Count-in (2400 ms), and Vocalization (4800 ms). In the task presentation epoch, the statement and emotion to be produced by the vocalist were presented on screen as text for 4500 ms. Once the text had been on screen for 1000 ms, the melody to be used by the vocalist was sounded (2400 ms). The count-in epoch presented a visual count-in timer ("1," "2," "3," "4") at an IOI of 600 ms. The start of the vocalize epoch was signaled with a green circle that was displayed for 2400 ms. The stimulus timeline was preceded by an auditory beep (500 ms) and 1000 ms of silence, and ended with an auditory beep (500 ms). Temporal accuracy of the presentation software was confirmed with the Black Box Toolkit (Plant et al., 2004).

Stimuli were presented visually on a 15 inch Macbook Pro running Windows XP SP3 and auditorily over KRK Rocket 5 speakers, controlled by Matlab, 2009b and the Psychophysics Toolbox (3.0.8 SVN 1648, Brainard, 1997). Recordings were performed in a sound-attenuated recording studio equipped with sound baffles. Vocal output was recorded with an AKG C414 B-XLS cardioid microphone with a pop filter, positioned 30 cm from the vocalist, and digitized on a Mac Pro computer with Pro Tools at 16 bit/48 kHz, and a Digidesign 003 mixing workstation.

Design and procedure

The experimental design was a 5 (Emotion: calm, happy, sad, angry, fearful) × 2 (Statement: kids, dogs) × 2 (Intensity: normal, strong) × 2 (Repetition) within-subjects design, with 40 trials per participant. A dialog script was used with vocalists. Each emotion was described, along with a vignette describing a scenario involving that emotion. Trials were blocked by emotion. Two presentations orders of emotion were used, and counterbalanced across participants (calm, happy, sad, angry fearful, or sad, angry, fearful, calm, happy). Within emotion blocks, trials were blocked by statement and counterbalanced across participants. For all vocalists, strong intensity productions followed normal intensity productions. An intensity factor was included to capture a broader range of emotional expression (Diener et al., 1985; Sonnemans and Frijda, 1994), which has been shown to affect the acoustics of vocal emotional productions (Banse and Scherer, 1996; Juslin and Laukka, 2001). It was emphasized that vocalists were to produce genuine expressions of emotion, and that they were to prepare themselves physiologically using method acting or emotional memory techniques so as to induce the desired emotion prior to recording. Time was provided between each emotion to allow vocalists to reach the intended emotional state. This form of induction procedure has been used previously in the creation of emotional stimuli (Bänziger et al., 2012). The concept of indicating was also explained, and vocalists were instructed not to produce an indicated performance. Vocalists were told to sing the basic notated pitches, but that they were free to vary acoustic characteristics in order to convey the desired emotion in a genuine manner. Vocalists were standing during all productions. Vocalists were allowed to repeat a given trial until they were comfortable with their production. The final two productions were used in subsequent analyses.

Analyses

Recordings were edited using Adobe Audition CS6. Vocal intensity was peak-normalized within each vocalist to retain acoustic intensity variability across the emotions. Recording levels were adjusted across vocalists to prevent clipping, given the range in vocal intensity across participants[1]. Acoustic recordings were analyzed with Praat (Boersma and Weenink, 2013). Fundamental frequency (F_0 mean, floor, and ceiling), HNR, and jitter (local) were extracted[2]. To assess pitch accuracy, F_0 of the first note of the melody was examined ($M_{\text{duration}} = 225.3$ ms, SD = 85.35 ms). Three measures of pitch accuracy in the first note were examined: F_0 mean is the average pitch of the first note; F_0 floor is the minimum pitch value during the first note, while F_0 ceiling is the maximum pitch value during the first note. Pitch contours of the first note were converted to cents to provide a normalized measure of inaccuracy from the intended pitch (F3 = 174.614 Hz); a value of 0 cents would indicate perfect accuracy (174.614 Hz), 100 cents would indicate a sharp performance of 1 semitone above the target pitch (184.997 Hz), and −100 cents would indicate a flat performance of 1 semitone below the target pitch (164.814 Hz). Note onsets and offsets were marked in Praat with respect to characteristic changes in the spectrogram, acoustic intensity, and pitch contours. Ten percent of the samples were checked by a second rater (mean inter-rater boundary time difference = 2.1 ms, SD = 2.2 ms). HNR and jitter measures were taken across the voiced portions of the entire utterance.

Statistical analyses

Linear mixed models were fitted using the MIXED function in SPSS 22.0. In Experiment 1, all models were fitted with a diagonal covariance structure for the repeated covariance type, which is the default structure for repeated measures in SPSS 22.0. In Experiment 1, analogous models were also fitted using AR(1) and ARH(1), more suited to longitudinal repeated measures, and the more conservative unstructured covariance matrix (Field, 2009). Models fitted with AR(1) and ARH(1) yielded poorer fits, while models fitted with unstructured covariance could not be assessed as the number of parameters to be fitted exceeded the number of observations. Random effects were fitted with a variance components (VC) covariance structure, as is suggested for random intercept models (Field, 2009). All other statistical tests were carried out in Matlab, 2013b or SPSS 22.0.

RESULTS

Separate repeated measures LMMs were conducted to assess how vocal experience predicted acoustic measures of the singing voice. Five acoustic measures were examined: F_0 (mean, floor, and ceiling), Jitter, and HNR. Repeated measures LMMs were used as each vocalist was recorded singing 40 times, with Vocalist (12) entered as a random effect (intercept), and Emotion (5 levels), Intensity (2), Statement (2), Repetition (2), Singing Lessons (continuous), and Acting Experience (continuous) entered as fixed effects. LMMs were built using a "step-up" strategy, starting with an unconditional means model with only intercepts for fixed and random effects, and then adding in random coefficients (Singer, 1998; Snijders and Bosker, 1999; Raudenbush and Bryk, 2002; Twisk, 2006). For each step, changes to the model fit were assessed with likelihood tests using maximum likelihood (ML) estimation (Twisk, 2006). Factors which significantly improved the model fit were retained. Adding the effect of Repetition (all p-values > 0.236), or any of its interactions with Emotion, Intensity, and Statement (all p-values > 0.163) were not found to significantly improve model fits for any acoustic parameter and was not included in the final model. Similarly, the interaction of Statement × Intensity did not significantly improve model fits for any acoustic parameter and was not included in the final model (all p-values > 0.395). While Statement × Emotion only improved the model fit for F_0 (ceiling), the interaction was retained to facilitate comparisons between models (Cheng et al., 2009).

Outcomes for the final models are described in **Table 1**. For F_0 (floor), main effects were reported for Statement, Emotion, and Intensity, indicating that vocalists varied their minimum F_0 depending on their emotional intent or statement. Pairwise comparisons with Bonferroni correction confirmed that Calm ($M = -229.21$, SE = 15.66) exhibited a lower F_0 floor than Happy ($M = -161.41$, SE = 11.81), Angry ($M = -164.19$, SE = 12.44), and Fearful ($M = -124.85$, SE = 15.66), but not Sad ($M = -184.21$, SE = 15.66). Normal intensity emotions ($M = -191.54$, SE = 9.98) had a lower F_0 floor than strong intensity emotions ($M = -154.17$, SE = 10.12). Importantly, vocal experience was found to have a significant effect on vocalists' F_0 floor, where vocalists with more years of acting experience exhibited a lower F_0 floor, $b = -9.21$, $t(8.84) = -4.15$, $p = 0.003$; illustrated in **Figure 1**. Conversely, vocalists with more years of singing training exhibited a higher F_0 floor in their first note, $b = 6.40$, $t(8.84) = 2.64$, $p = 0.027$. To further examine these effects, we took median splits based on years of Acting Experience: $F_{0\ \text{Floor}-\text{ActingLow}} = -145.42$ cents, SD = 112.34 ($N = 8$), and $F_{0\ \text{Floor}-\text{ActingHigh}} = -234.0$ cents, SD = 170.94 ($N = 4$), and on years of Singing Lessons: $F_{0\ \text{Floor}-\text{SingingLow}} = -209.8$ cents, SD = 152.92 ($N = 6$) and $F_{0\ \text{Floor}-\text{SingingHigh}} = -140.1$ cents, SD = 118.17 ($N = 6$). These results suggest that vocalists with greater acting experience, and vocalists with less singing training, exhibited an F_0 floor that was further from the target pitch. The relationship between the categorical fixed factors and F_0 floor, when controlling for vocal experience, showed significant variance in the intercepts across vocalists var(u_{0j}) = 2686.33, $\chi^2(1) = 52.21$, $p < 0.01$.

For F_0 mean, main effects were reported for Statement, Emotion, and Intensity, indicating that vocalists also varied their mean F_0 depending on their emotional intent or statement. Pairwise

[1] Prior to recording, participants were asked to sing and speak several test sentences with a very angry emotional intention. Very angry was selected as this emotion was often the loudest during the audition pre-screening sessions. Recording levels were adjusted based on the loudest of these test productions. For occasional instances of clipping or "popping" during the recording sessions, the actor was asked to repeat the trial.

[2] Pitch contour was extracted with an autocorrelation algorithm (ac) in Praat, with the following settings: pitch floor 70 Hz, pitch ceiling 420 Hz, very accurate pitch contour tracking, maximum periodicity candidates 15, silence threshold 0.08, voicing threshold 0.45, octave cost 0.01, octave-jump cost 0.4, voiced/unvoiced cost 0.14, time step 0.004. Jitter (local) was extracted using a periodic cross-correlation algorithm in Praat (periodic, cc).

Table 1 | Summary of results from linear mixed models in Experiment 1 comparing the effects of vocal experience on acoustic parameters of the voice.

Acoustic parameter	Fixed effects							Random effects
	Continuous		Categorical					Intercept
	Acting	Singing	Statement	Emotion	Intensity	S × E	E × I	Vocalist
F_0 Floor	$F_{(1,8.84)} = 17.22$, **$p = 0.003$**	$F_{(1,8.84)} = 6.98$, **$p = 0.027$**	$F_{(1,323.28)} = 10.48$, **$p = 0.001$**	$F_{(4,139.46)} = 12.26$, **$p < 0.001$**	$F_{(1,302.34)} = 14.75$, **$p < 0.001$**	$F_{(4,145.77)} = 0.97$, $p = 0.424$	$F_{(4,138.68)} = 2.56$, **$p = 0.041$**	$\text{var}(u_{0j}) = 2686.33$, $\chi^2(1) = 52.21$ **$p < 0.01$**
F_0 Mean	$F_{(1,10.47)} = 8.59$, **$p = 0.014$**	$F_{(1,10.47)} = 1.08$, $p = 0.321$	$F_{(1,209.1)} = 9.96$, **$p = 0.002$**	$F_{(4,86.48)} = 34.6$, **$p < 0.001$**	$F_{(1,182.9)} = 83.26$, **$p < 0.001$**	$F_{(4,85.21)} = 0.37$, $p = 0.828$	$F_{(4,86.98)} = 10.59$, **$p < 0.001$**	$\text{var}(u_{0j}) = 734.63$, $\chi^2(1) = 54.21$ **$p < 0.01$**
F_0 Ceiling	$F_{(1,10.47)} = 0.67$, $p = 0.43$	$F_{(1,10.47)} = 0.09$, $p = 0.776$	$F_{(1,214.76)} = 9.96$, $p = 0.303$	$F_{(4,116.29)} = 40.27$, **$p < 0.001$**	$F_{(1,216.51)} = 62.63$, **$p < 0.001$**	$F_{(4,111.53)} = 2.437$, $p = 0.051$	$F_{(4,86.98)} = 9.2$, **$p < 0.001$**	$\text{var}(u_{0j}) = 2583.64$, $\chi^2(1) = 47.77$ **$p < 0.01$**
Jitter	$F_{(1,11.61)} = 5.0$, **$p = 0.046$**	$F_{(1,11.61)} = 0.033$, $p = 0.86$	$F_{(1,245.44)} = 20.18$, **$p < 0.001$**	$F_{(4,99.53)} = 82.01$, **$p < 0.001$**	$F_{(1,217.39)} = 80.03$, **$p < 0.001$**	$F_{(4,98.179)} = 2.38$, $p = 0.057$	$F_{(4,86.98)} = 10.59$, **$p < 0.001$**	$\text{var}(u_{0j}) = 3.61 \times 10^{-6}$, $\chi^2(1) = 91.68$ **$p < 0.01$**
HNR	$F_{(1,11.80)} = 2.29$, $p = 0.156$	$F_{(1,11.80)} = 0.172$, $p = 0.686$	$F_{(1,360.63)} = 37.28$, **$p < 0.001$**	$F_{(4,122.16)} = 294.62$, **$p < 0.001$**	$F_{(1,339.71)} = 141.46$, **$P < 0.001$**	$F_{(4,112.73)} = 0.369$, $p = 0.83$	$F_{(4,124.91)} = 26.29$, **$p < 0.001$**	$\text{var}(u_{0j}) = 2.17$, $\chi^2(1) = 126.64$ **$p < 0.01$**

The significance of the fixed effects was assessed with Type III SS F-tests on the final multivariate model. Changes in model fit for fixed effects were assessed with ML estimation. Variance estimates for random effects are reported using REML estimation (Twisk, 2006). Statistically significant p-values are highlighted with bold typeface. Fixed effects and interactions that did not significantly improve model fits in any of the acoustic parameters were not included in the final models.

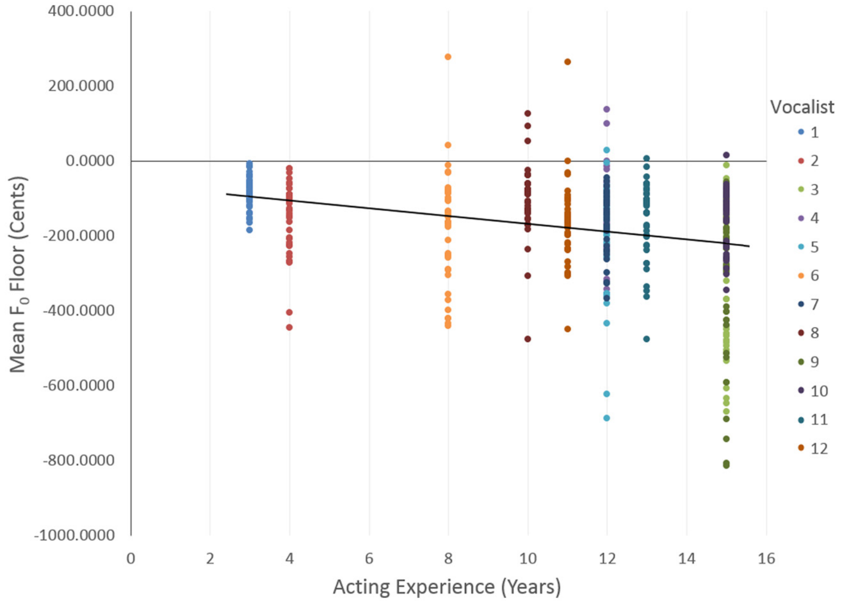


FIGURE 1 | Vocalists' years of acting experience and F_0 floor of their first note for all trials. Solid black line indicates a line of best fit, with a linear regression solution: F_0 floor = –62.66 – 10.37 × Acting experience.

comparisons with Bonferroni correction confirmed that Calm ($M = -46.36$, SE = 7.15) exhibited a lower F_0 mean than Happy ($M = -6.2$, SE = 7.56), Sad ($M = -23.26$, SE = 7.94), Angry ($M = 35.71$, SE = 10.6), and Fearful ($M = 25.59$, SE = 9.92). Normal intensity emotions ($M = -28.16$, SE = 6.72) also had a lower F_0 mean than strong intensity emotions ($M = 22.35$, SE = 7.82).

Importantly, acting experience was found to have a significant effect on vocalists' F_0 mean, where vocalists with more years of acting experience exhibited a lower mean F_0, $b = -4.92$, $t(10.47) = -2.93$, $p = 0.014$. To further examine these pitch differences, we took median splits on Acting Experience: $F_{0\ \text{Mean-ActingLow}} = 10.79$ cents, SD = 90.52 ($N = 8$) and $F_{0\ \text{Mean-ActingHigh}} = -21.7$ cents, SD = 82.69 ($N = 4$). These results suggest that vocalists with more years of acting experience were more flat on the first note. The mean absolute pitch of the first note across all vocalists was 55.2 cents (SD = 70.06). These results suggest that vocalists in general sang the first note of the melody within half a semitone of the target pitch. The relationship between the categorical fixed factors and F_0 mean, when controlling for vocal experience, also showed significant variance in the intercepts across vocalists, $\text{var}(u_{0j}) = 734.63$, $\chi^2(1) = 54.21$, $p < 0.01$.

For F_0 ceiling, main effects were reported for Emotion and Intensity, indicating that vocalists varied their F_0 ceiling depending on their emotional intent. Pairwise comparisons with Bonferroni correction confirmed that Calm ($M = 75.78$, SE = 13.46) had a lower F_0 ceiling than Happy ($M = 142.48$, SE = 14.72), Sad ($M = -153.0$, SE = 18.52), Angry ($M = 216.2$, SE = 21.47), and Fearful ($M = 237.75$, SE = 19.62). Normal intensity emotions ($M = 119.28$, SE = 13.74) also had a lower F_0 ceiling than strong intensity emotions ($M = 210.80$, SE = 15.59). No relationship was reported between vocal experience and F_0 ceiling. The relationship between the categorical fixed factors and F_0 ceiling also showed significant variance in the intercepts across vocalists, $\text{var}(u_{0j}) = 2583.64$, $\chi^2(1) = 47.77$, $p < 0.01$.

For Jitter, main effects were reported for Statement, Emotion, and Intensity, indicating that the level of jitter in vocalists' voices varied depending on their emotional intent or statement. Pairwise comparisons with Bonferroni correction confirmed that Calm ($M = 0.011$, SE = 4.21×10^{-4}) had less jitter than Happy ($M = 0.014$, SE = 4.43×10^{-4}), Sad ($M = 0.013$, SE = 4.57×10^{-4}), Angry ($M = 0.017$, SE = 5.46×10^{-4}), and Fearful ($M = 0.017$, SE = 5.62×10^{-4}). Normal intensity emotions ($M = 0.013$, SE = 4.09×10^{-4}) had less jitter than strong intensity emotions ($M = 0.015$ SE = 4.47×10^{-4}). These findings are important as they demonstrate that the level of jitter in a vocalist's voice can be affected by both lexical and emotional goals. Following from this, Acting Experience was found to have a significant effect on vocalists' jitter levels, where vocalists with more years of acting experience exhibited a higher level of vocal jitter, $b = 2.31 \times 10^{-4}$, $t(11.61) = 2.24$, $p = 0.046$. To further examine this effect, we took median splits based on years of Acting Experience: $\text{Jitter}_{\text{ActingLow}} = 1.37\% \times 10^{-2}\%$, SD = 5.0×10^{-3} ($N = 8$), and $\text{Jitter}_{\text{ActingHigh}} = 1.49\% \times 10^{-2}\%$, SD = 4.1×10^{-2} ($N = 4$). These results suggest that vocalists with more years of acting experience had higher levels of vocal jitter. No relationship was reported between jitter and years of Singing lessons. The relationship between our fixed factors and jitter also showed significant variance in the intercepts across vocalists, $\text{var}(u_{0j}) = 3.61 \times 10^{-6}$, $\chi^2(1) = 91.68$, $p < 0.01$.

For HNR, main effects were reported for Statement, Emotion, and Intensity, indicating that vocalists varied the HNR in their voice depending on their emotional intent or statement. Pairwise comparisons with Bonferroni correction confirmed that Calm ($M = 18.58$, SE $= 0.37$) had a higher HNR than Happy ($M = 15.99$, SE $= 0.38$), Sad ($M = 17.41$, SE $= 0.38$), Angry ($M = 13.0$, SE $= 0.38$), and Fearful ($M = 14.19$, SE $= 0.37$). Normal intensity emotions ($M = 16.57$, SE $= 0.36$) also had a higher HNR than strong intensity emotions ($M = -15.1$, SE $= 0.36$). As with jitter, this is an important finding as it confirms that the HNR in a vocalist's voice is not fixed. No relationships were found between HNR and Acting experience, and HNR and Singing lessons. The relationship between the fixed factors and HNR was also found to show significant variance in the intercepts across vocalists, $\text{var}(u_{0j}) = 2.17$, $\chi^2(1) = 126.64$, $p < 0.01$.

DISCUSSION

The results of Experiment 1 confirmed that different types of vocal training, in the form of years of acting experience and years of singing lessons, produced differences in the acoustics of the singing voice. Vocalists with more years of acting experience exhibited a lower F_0 floor, with the most experienced actors singing on average up to 234 cents below the target pitch, a deviation of more than 2 semitones (E^b3 instead of F3). In contrast, vocalists with more years of singing lessons exhibited a F_0 floor that was closer to the target pitch relative to less trained singers. Vocalists with more years of acting experience also sang the first note flat, with a lower F_0 mean relative to vocalists with fewer years of acting experience. Overall, vocalists' mean pitch for the first note varied within half a semitone of the target pitch. No relationships were reported for F_0 ceiling and vocal training. On measures of voice quality, vocalists with more years of acting experience exhibited higher levels of jitter. No relationship was found between vocal experience and HNR. Importantly, both jitter and HNR varied consistently across emotion, intensity, and statement, confirming that like emotional speech (Dupuis and Pichora-Fuller, in press), these spectral aspects of the emotional singing voice are not fixed within a vocalist. These results partially support our hypothesis, and suggest that vocalists with more years of acting experience sung with a lower voice quality, as indexed by greater pitch inaccuracy and higher levels of jitter. No effects were reported between singing training and measures of voice quality, and so our hypotheses regarding these acoustic measures was not supported. Significant random intercepts were reported in all acoustic features, indicating a consistent tendency by some vocalists to exhibit higher or lower levels of these acoustic measures than other vocalists, even when controlling for the effects of their vocal experience background. These results support the use of LMMs in the analysis of Experiment 1, by accounting for additional variance within acoustic parameters across the vocalists.

Collectively, these results suggest that the type and amount of vocal training a singer receives may have a significant effect on acoustic measures of their singing voice. In particular, vocalists with more years of acting experience sung with a lower voice quality and greater pitch inaccuracy. We theorize that such deviations may have been intentional so as to increase the perception of emotional genuineness during their performances. To assess this relationship we conducted a second experiment in which listeners' evaluated the emotional genuineness of vocalists' singing performances.

EXPERIMENT 2

Experiment 2 examined listeners' perception of emotional genuineness from vocalists' singing recordings. In Experiment 1, vocalists with more years of acting experience exhibited increased pitch inaccuracy and higher levels of vocal jitter. We theorized these deviations were an intentional singing technique by more experienced actors to increase the genuineness of their performances. We hypothesized that vocalists with more years of acting experience would be rated by listeners as possessing higher levels of emotional genuineness. We further expected that acoustic measures of the voice would also be associated with listeners' perception of genuineness. We hypothesized that recordings with a lower F_0 floor and increased jitter would be rated as more genuine. While no effect was reported between vocal training and HNR in Experiment 1, based on our original theoretical predictions we hypothesized that recordings with a lower HNR would be rated as more emotionally genuine.

METHOD

Participants

Fourteen adults (7 female, mean age = 29.29, SD = 7.49) were recruited from the Ryerson university community. The experiment took approximately 30 min. No participant from Experiment 1 took part in Experiment 2.

Stimulus and apparatus

A subset of acoustic recordings from Experiment 1 were used as stimuli in Experiment 2. Ten recordings were used for each vocalist, one for each emotional category and emotional intensity level. The statement used was "Kids are talking by the door." Stimuli were presented acoustically with a Macbook Pro laptop and Logitech X-140 powered external speakers.

Design and procedure

The experimental design was a 12 (Vocalist) × 5 (Emotion: calm, happy, sad, angry, fearful) × 2 (Intensity: normal, strong) within-subjects design, with 120 trials per participant. Trials were presented in random order. On each trial, participants were asked to rate the genuineness of the vocalist's production using a 5-point scale (1 = not at all genuine to 5 = very genuine). Prior to the experiment, the concept of emotional genuineness was explained to participants as follows: "Emotional genuineness concerns whether you believe that the vocalist was truly experiencing the emotion they were portraying. Emotional genuineness should not be confused with the intensity or clarity of the portrayed emotion." Loudness was adjusted to a comfortable level, and was held constant across presentations.

Analyses

The relationships between listeners' genuineness ratings and vocalists' years of acting experience and singing lessons were assessed with LMMs. The statistical procedures described in Experiment 1 were reused in Experiment 2. As in Experiment 1, analogous models were fitted using AR(1) and ARH(1), and the more conservative unstructured covariance matrix (Field, 2009). Models fitted with AR(1) and ARH(1) yielded poorer fits, while models fitted with unstructured covariance failed to converge. Random effects were again fitted with a VC covariance structure.

RESULTS

A three-level repeated measures LMM was conducted to assess how vocal experience predicted listeners' ratings of emotional genuineness. A repeated measures LMM was used as each vocalist was presented 10 times to each of the 14 listeners. Vocalist (12) was entered as a random effect, and was further added as a random effect nested within Listener (14). The variables Emotion (5 levels), Intensity (2), Singing Lessons (continuous), and Acting Experience (continuous) were entered as fixed effects. Based on Experiment 1 results, we did not expect Singing lessons to have a significant effect on listeners' ratings of genuineness. However for completeness, its effect on the model was examined. Singing Lessons did not significantly improve the model fit ($p = 0.542$), and was not included in the final model.

Outcomes of the final model are described in **Table 2**. Main effects were reported for Emotion and Intensity, as was an interaction between Emotion and Intensity, illustrated in **Figure 2**. Pairwise comparisons with Bonferroni correction confirmed that Calm ($M = 2.985$, SE $= 0.12$) was rated as significantly more genuine than Happy ($M = 2.637$, SE $= 0.12$) and Fearful ($M = 2.604$, SE $= 0.12$), but not Angry ($M = 2.807$, SE $= 0.12$) or Sad ($M = 2.851$, SE $= 0.12$). Less intense emotions ($M = 2.858$, SE $= 0.12$) were also rated as more genuine than more intense emotions ($M = 2.695$, SE $= 0.12$). Less intense emotions were rated as more genuine for all emotions except angry, suggesting a role in the interaction.

Importantly, vocalists' acting experience was found to have a significant effect on listeners' ratings of emotional genuineness, where vocalists with more years of acting experience were rated as more emotionally genuine, $b = 0.035$, $t(150.45) = 4.46$, $p < 0.001$, illustrated in **Figure 3**. This result supports our main hypothesis that vocalists with more years of acting experience would be rated as more emotionally genuine. The relationship between the categorical fixed factors and Genuineness, when controlling for acting experience, showed significant variance in the intercepts across Listener $\text{var}(u_{0j}) = 0.172$, $\chi^2(1) = 203.27$, $p < 0.01$, and in the intercepts across Vocalist within Listener, $\text{var}(u_{0j}) = 0.066$, $\chi^2(1) = 31.47$, $p < 0.01$.

To determine if a relationship existed between the acoustic features examined in Experiment 1 and listeners' ratings of emotional genuineness, we ran a LMM with Emotion (5 levels), Intensity (2), F_0 Floor (continuous), F_0 Mean (continuous), F_0 Ceiling (continuous), F_0 Jitter (continuous), and HNR (continuous) entered as fixed effects. Adding the effects of F_0 mean ($p = 0.94$) and F_0 ceiling ($p = 0.258$) were not found to significantly improve model fits for emotional genuineness, and were not included in the final model. The main effect of Intensity was significant until the addition of the final acoustic parameter HNR, after which it was no longer significant ($p = 0.386$). To facilitate a comparison with previous models, this effect was retained in the final model.

Outcomes of the final model are described in **Table 3**. Main effects were reported for Emotion as was an interaction between Emotion and Intensity. Importantly, three of the five acoustic parameters examined were found to affect listeners' ratings of emotional genuineness. Recordings with a lower F_0 floor were rated as more emotionally genuine, $b = -5.97 \times 10^{-4}$, $t(1125.03) = -2.75$, $p = 0.006$, illustrated in **Figure 4A**.

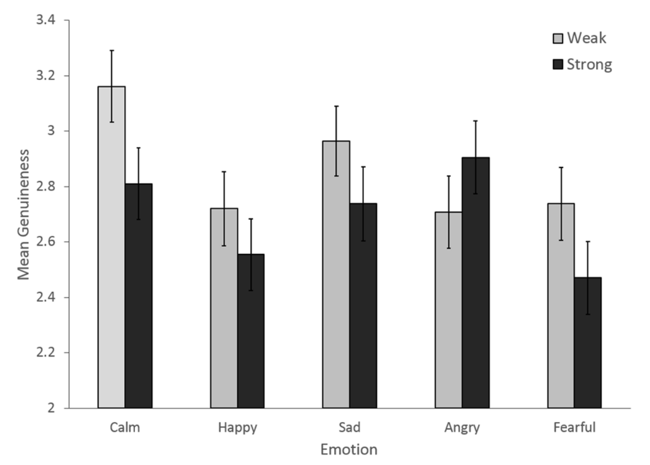


FIGURE 2 | Mean genuineness ratings showing the Emotion by Intensity interaction in Experiment 2. Error bars denote the standard error of the means.

Table 2 | Summary of results from the linear mixed model in Experiment 2 comparing listeners' ratings of emotional genuineness with vocalist training background of the vocalist.

Perceptual parameter	**Fixed effects**				**Random effects**	
	Continuous	**Categorical**			**Intercepts**	
	Acting	**Emotion**	**Intensity**	**E × I**	**Listener**	**Listener × Vocalist**
Genuineness	$F_{(1,151.44)} = 20.03$, $\boldsymbol{p < 0.001}$	$F_{(4,574.23)} = 9.44$, $\boldsymbol{p < 0.001}$	$F_{(1,1498.97)} = 12.53$, $\boldsymbol{p < 0.001}$	$F_{(4,574.23)} = 4.22$, $\boldsymbol{p = 0.002}$	$\text{var}(u_{0j}) = 0.172$, $\chi^2(1) = 203.27$, $\boldsymbol{p < 0.01}$	$\text{var}(u_{0j}) = 0.066$, $\chi^2(1) = 31.47$, $\boldsymbol{p < 0.01}$

The significance of the fixed effects was assessed with Type III SS F-tests on the final multivariate model. Changes in model fit for fixed effects were assessed with ML estimation. Variance estimates for random effects are reported using REML estimation (Twisk, 2006). Statistically significant p-values are highlighted with bold typeface. Fixed effects that did not significantly improve the model fit were not included in the final model.

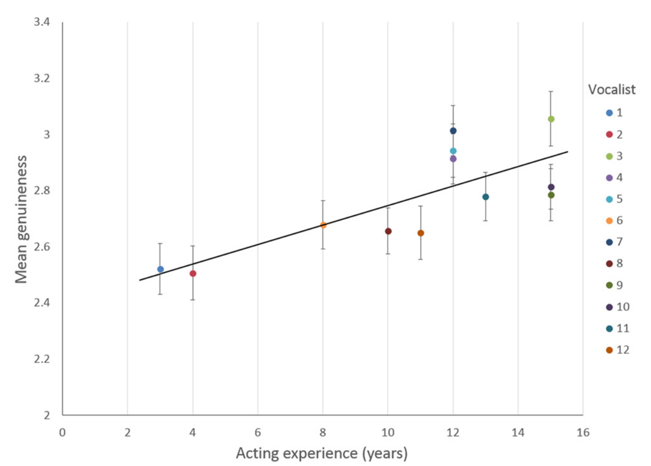


FIGURE 3 | Listeners' mean genuineness ratings for each vocalist, and vocalists' years of acting experience in Experiment 2. Solid black line indicates a line of best fit, with a linear regression solution: Genuineness = 2.4 + 0.035 × Acting experience. Error bars denote the standard error of the means.

Recordings with increased jitter were also rated as more emotionally genuine, $b = 16.93$, $t(950.72) = 2.05$, $p = 0.041$. Finally, recordings with increased HNR were also rated as more emotionally genuine, $b = 0.095$, $t(932.06) = 5.02$, $p < 0.001$, illustrated in **Figure 4B**. The model continued to show significant variance in the intercepts across Listener $\text{var}(u_{0j}) = 0.170$, $\chi^2(1) = 203.27$, $p < 0.01$, and in the intercepts across Vocalist within Listener, $\text{var}(u_{0j}) = 0.080$, $\chi^2(1) = 31.47$, $p < 0.01$.

DISCUSSION

The results of Experiment 2 confirmed that listeners' ratings of emotional genuineness were related to the level of acting experience of the vocalist, and to the acoustic features of the voice for: F_0 floor, Jitter, and HNR. Vocalists with more years of acting experience were rated as more emotionally genuine relative to vocalists with fewer years of acting experience, supporting our main hypothesis. No relationship was reported between years of singing lessons and emotional genuineness, as was expected based on findings from Experiment 1. The experimental factors Emotion and Intensity were also both found to affect listeners' perception of genuineness. Calm productions were overall rated as the most genuine, while fearful productions were rated as the least genuine. Interestingly, less intense emotions were rated as more genuine than strongly intense emotions. This suggests that vocalists' emotional displays were more believable when their expressions were less intense. However, the interaction between emotion intensity suggested that while this was the case for most emotions, strongly intense anger appeared to be rated as more genuine than less intense anger. Significant random intercepts were also reported for ratings of genuineness, both for individual listeners and for vocalists within listeners, indicating a consistent tendency by listeners to rate the genuineness of recordings more or less between one another, and for some vocalists over others. These results support the use of LMMs in the analysis of Experiment 2, by accounting for additional variance within genuineness ratings across listeners.

Importantly, three of the five acoustic measures examined were found to be significantly related to listeners' ratings of emotional genuineness. Recordings with a lower F_0 floor were rated as more emotionally genuine, as were recordings with increased jitter, both of which supported our hypothesis. HNR was also associated with listeners' ratings of emotional genuineness. However, counter to our hypothesis, recordings with a higher HNR were rated as more emotionally genuine. Thus, our hypothesis regarding HNR was only partially supported, as while HNR was associated with listeners' perception of emotion, the direction of the relationship was opposite to our predictions.

It is unclear why recordings with a higher mean HNR were rated as more emotionally genuine. A tentative explanation is that genuineness ratings may have been influenced by factors related to vocal attractiveness. Voices with a higher average HNR tend to be judged as more attractive (Bruckert et al., 2010). Consistent with the "halo effect" (Zebrowitz et al., 1996), participants are more willing to ascribe positive attributes, such as likability, to voices that are judged to be attractive (Zuckerman and Driver, 1989).

GENERAL DISCUSSION

Two experiments provided converging evidence that different types of vocal training affect the acoustics of the male singing voice in divergent ways, which concomitantly affect listeners' perception of emotional genuineness. Vocalists' exhibited differences in their fundamental frequency (F_0 floor, mean), and levels of jitter that were related to their years of vocal experience. Vocalists with more years of acting experience exhibited increased pitch inaccuracy with a lower minimum F_0 and a lower mean F_0 relative to the target pitch of the first note, and increased vocal jitter. In contrast, vocalists with more years of singing training exhibited a higher F_0 floor that was closer to the target pitch (less flat). No relationship was found between vocal training and HNR. Collectively, these results suggested that vocalists with more years of acting experience sung with a lower voice quality. It was theorized that vocalists' reduction in voice quality was an intentional phrasing technique – particularly amongst vocalists with a lot of acting experience – to increase the perception of their emotional genuineness. Findings from the perceptual experiment supported this hypothesis. Vocalists with more years of acting experience were rated as more genuine. No relationship was found between the amount of singing training and the perception of genuineness. As hypothesized, recordings with a lower F_0 floor and increased vocal jitter were rated as more emotionally genuine. As hypothesized, HNR was also associated with listeners' perception of genuineness, however the direction of the effect went against our expectations as voices containing a higher HNR were rated as more genuine. This latter finding may reflect a moderating role of vocal attractiveness in judgments of emotional genuineness (Zuckerman and Driver, 1989; Bruckert et al., 2010). Overall, these findings support our two main hypotheses that different types of vocal training affect the acoustics of the male singing voice in unique ways, which in turn affect listeners' perception of emotional genuineness.

An important outcome of this investigation was the identification of acoustic measures that affected listeners' perception of emotional genuineness. All three acoustic features varied consistently with the emotional category and intensity of the vocalist,

Table 3 | Summary of results from the linear mixed model in Experiment 2 comparing listener ratings of emotional genuineness with acoustic measures of the voice.

Perceptual parameter	Fixed effects						Random effects	
	Continuous			Categorical			Intercepts	
	***F*₀ (floor)**	**Jitter**	**HNR**	**Emotion**	**Intensity**	**E × I**	**Listener**	**Listener × Vocalist**
Genuineness	$F_{(1,151.44)} = 7.56$, $\boldsymbol{p = 0.006}$	$F_{(4,574.23)} = 4.19$, $\boldsymbol{p = 0.041}$	$F_{(1,1498.97)} = 25.25$, $\boldsymbol{p < 0.001}$	$F_{(4,778.63)} = 7.38$, $\boldsymbol{p < 0.001}$	$F_{(1,1610.45)} = 0.75$, $p = 0.39$	$F_{(4,700.78)} = 6.69$, $\boldsymbol{p < 0.001}$	$\mathrm{var}(u_{0j}) = 0.170$, $\chi^2(1) = 203.27$, $\boldsymbol{p < 0.01}$	$\mathrm{var}(u_{0j}) = 0.080$, $\chi^2(1) = 31.47$, $\boldsymbol{p < 0.01}$

*The significance of the fixed effects was assessed with Type III SS F-tests on the final multivariate model. Changes in model fit for fixed effects were assessed with ML estimation. Variance estimates for random effects are reported using REML estimation (Twisk, 2006). Statistically significant p-values are highlighted with bold typeface. Fixed effects that did not significantly improve the model fit and were not included in the previous model (**Table 2**), were not included in the final model.*

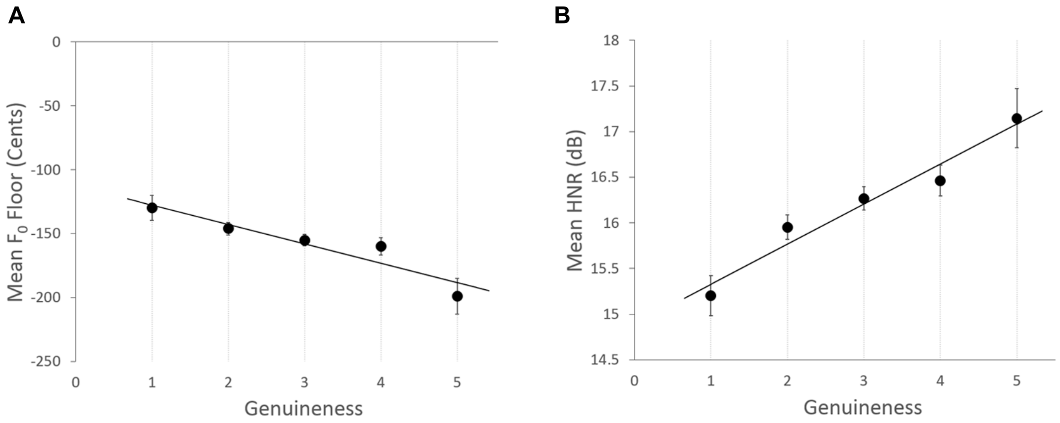


FIGURE 4 | Relationships between vocalists' acoustic features and in Experiment 2 for (A) Mean F_0 floor for each genuineness rating category. Solid black line indicates a line of best fit, with a linear regression solution: F_0 floor = −112.49 − 15.21 × Genuineness. **(B)** Mean HNR for each genuineness rating category in Experiment 2. Solid black line indicates a line of best fit, with a linear regression solution: HNR = 14.89 + 0.44× Genuineness. For both figures, error bars denote the standard error of the means.

confirming that the spectral qualities jitter and HNR of the voice are not fixed for a given vocalist. While F_0 is generally under the conscious control of the vocalist, it is unclear whether the same is true of jitter or HNR. Thus an interesting avenue for future research would be to examine if vocalists can be trained to consciously control the levels of jitter and HNR in the voice. These outcomes would be relevant to vocal pedagogy in those performers seeking to increase their emotional genuineness with listeners. The findings would also be relevant to vocal attractiveness research, where increased HNR is thought to influence the perception of vocal attractiveness (Bruckert et al., 2010).

The results of the present study indicated that vocalists with more years of acting experience sung with a lower voice quality. We theorized that these performers were seeking to put their "personal stamp on the song" (Deer and Dal Vera, 2008, p. 226), where the use of stylistic deviations may function to enhance the individual uniqueness or emotionality of the performance. The connection between stylistic deviations and performer uniqueness has been reported previously. Repp (1992) examined the expressive timing deviations of 24 international concert pianists in their performances of Schubert's Träumerei. While all pianists exhibited characteristic tempo changes matching the structure of the work, large individual differences were reported in which performers deviated extensively from the expected timing curve, and particularly for two of the more famous performers. In a follow-up study involving graduate piano students' performances of the same work, Repp (1995) found that the students also exhibited similar timing patterns, but that their deviations were much more homogeneous than those of the concert pianists. These findings suggest that *phrasing*, as it is referred to in the acting world, may be a general artistic phenomenon in which more experienced performers seek to differentiate themselves with their own unique style. Thus, while phrasing may involve the deviation or degradation of a typical performance, it may be done so purposefully and should not be considered erroneous. We believe that the acoustic deviations exhibited by vocalists with a large amount of acting experience in this study should be viewed in this light. In the present study these relationships were examined using performers who had varying levels of singing and acting experience. In the future these effects could be examined more directly with participants who were more closely matched on their years of singing and acting experience.

The importance of genuineness in emotion research has received increasing attention over the last decade. Differences in the production and perception of genuine versus simulated emotions is a topic of intense debate (Russell et al., 2003; Scherer, 2003; Vogt and André, 2005). The use of induction procedures is also gaining use amongst researchers who require ecologically valid stimuli. (Douglas-Cowie et al., 2007; Bänziger et al., 2012). In this study an induction procedure was used in an attempt to induce the physiological and mental correlates of the emotion being expressed. Likewise, researchers are increasingly assessing observers' beliefs about the genuineness of their stimuli (Langner et al., 2010). The results of the present study suggest that vocal training type and the duration of experience may serve as useful predictors of a vocalist's emotional genuineness, and that these factors should be considered in future genuineness studies.

CONCLUSION

The goals of a vocal performer are varied and many: accurate pitch reproduction, desired voice quality, clear intelligibility, precise timing, and intended emotional inflection. The findings of the present study confirm that these factors are not independent, and that performers may prioritize different aspects of their performance due to differences in their vocal training. These acoustic changes have important consequences on listeners' evaluation of emotion, and highlight the nuanced quality of individual differences in singing performance.

ACKNOWLEDGMENTS

This research was supported by AIRS (Advancing Interdisciplinary Research in Singing), a Major Collaborative Research Initiative funded by the Social Sciences and Humanities Research Council of Canada [www.airsplace.ca]. The authors thank Rena Sharon, Darryl Edwards, and Charlene Santoni for helpful discussions regarding the role of stylistic constraints and vocal pedagogy on emotional expressiveness in song.

REFERENCES

Amir, O., Amir, N., and Kishon-Rabin, L. (2003). The effect of superior auditory skills on vocal accuracy. *J. Acoust. Soc. Am.* 113, 1102. doi: 10.1121/1.1536632

Awan, S. N., and Ensslen, A. J. (2010). A comparison of trained and untrained vocalists on the Dysphonia Severity Index. *J. Voice* 24, 661–666. doi: 10.1016/j.jvoice.2009.04.001

Banse, R., and Scherer, K. R. (1996). Acoustic profiles in vocal emotion expression. *J. Pers. Soc. Psychol.* 70, 614–636. doi: 10.1037/0022-3514.70.3.614

Bänziger, T., Mortillaro, M., and Scherer, K. R. (2012). Introducing the Geneva multimodal expression corpus for experimental research on emotion perception. *Emotion* 12, 1161. doi: 10.1037/a0025827

Bartholomew, W. T. (1934). A physical definition of "good voice-quality" in the male voice. *J. Acoust. Soc. Am.* 6, 25–33. doi: 10.1121/1.1915685

Bele, I. V. (2006). The speaker's formant. *J. Voice* 20, 555–578. doi: 10.1016/j.jvoice.2005.07.001

Bland, J. M., and Altman, D. G. (1994). Correlation regression and repeated data. *BMJ: Br. Med. J.* 308, 896.

Boersma, P., and Weenink, D. (2013). *Praat: Doing Phonetics by Computer* (*Version 5.3.49*) [*Computer Program*]. Available at: http://www.praat.org/ (accessed March 19, 2013).

Brainard, D. H. (1997). The psychophysics toolbox. *Spat. Vis.* 10, 433–436. doi: 10.1163/156856897X00357

Brown, W. S. J., Rothman, H. B., and Sapienza, C. M. (2000). Perceptual and acoustic study of professionally trained versus untrained voices. *J. Voice* 14, 301–309. doi: 10.1016/S0892-1997(00)80076-4

Bruckert, L., Bestelmeyer, P., Latinus, M., Rouger, J., Charest, I., Rousselet, G., et al. (2010). Vocal attractiveness increases by averaging. *Curr. Biol.* 20, 116–120. doi: 10.1016/j.cub.2009.11.034

Bunch, M. D. (2009). *Dynamics of the Singing Voice*, 5th Edn. Vienna: Springer-Verlag/Wien.

Cheng, J., Edwards, L. J., Maldonado-Molina, M. M., Komro, K. A., and Muller, K. E. (2009). Real longitudinal data analysis for real people: building a good enough mixed model. *Stat. Med.* 29, 504–520. doi: 10.1002/sim.3775

Coltheart, M. (1981). The MRC psycholinguistic database. *Q. J. Exp. Psychol.* 33A, 497–505. doi: 10.1080/14640748108400805

Dalla Bella, S., Giguère, J.-F., and Peretz, I. (2007). Singing proficiency in the general population. *J. Acoust. Soc. Am.* 121, 1182. doi: 10.1121/1.2427111

Dalla Bella, S., Peretz, I., Rousseau, L., and Gosselin, N. (2001). A developmental study of the affective value of tempo and mode in music. *Cognition* 80, B1–B10. doi: 10.1016/S0010-0277(00)00136-0

Deer, J., and Dal Vera, R. (2008). *Acting in Musical Theatre: A Comprehensive Course*, 1st Edn. Abingdon, Oxon.

Diener, E., Larsen, R. J., Levine, S., and Emmons, R. A. (1985). Intensity and frequency: dimensions underlying positive and negative affect. *J. Pers. Soc. Psychol.* 48, 1253–1265. doi: 10.1037/0022-3514.48.5.1253

Douglas-Cowie, E., Cowie, R., Sneddon, I., Cox, C., Lowry, O., McRorie, M., et al. (2007). The HUMAINE database: addressing the collection and annotation of naturalistic and induced emotional data. *Affect. Comput. Intell. Interact.* 488–500.

Dupuis, K., and Pichora-Fuller, M. K. (in press). Intelligibility of emotional speech in younger, and older adults. *Ear Hear.*

Ferrand, C. T. (2002). Harmonics-to-noise ratio: an index of vocal aging. *J. Voice* 16, 480–487. doi: 10.1016/S0892-1997(02)00123-6

Field, A. (2009). *Discovering Statistics Using SPSS*, 3rd Edn. London: Sage publications.

Fujisaki, H. (1983). "Dynamic characteristics of voice fundamental frequency in speech and singing," in *The Production of Speech*, ed. P. MacNeilage (New York: Springer), 39–55.

Gobl, C., and Nì Chasaide, A. (2003). The role of voice quality in communicating emotion, mood and attitude. *Speech Commun.* 40, 189–212. doi: 10.1016/S0167-6393(02)00082-1

Howard, D. M., and Angus, J. A. (1997). A comparison between singing pitching strategies of 8 to 11 year olds and trained adult singers. *Logoped. Phoniatr. Vocol.* 22, 169–176. doi: 10.3109/14015439709075331

Hutchins, S. M., and Peretz, I. (2012). A frog in your throat or in your ear? Searching for the causes of poor singing. *J. Exp. Psychol. Gen.* 141, 76–97. doi: 10.1037/a0025064

Juslin, P. N., and Laukka, P. (2001). Impact of intended emotion intensity on cue utilization and decoding accuracy in vocal expression of emotion. *Emotion* 1, 381–412. doi: 10.1037/1528-3542.1.4.381

Katselas, M. (2008). *Acting Class: Take a Seat*. Beverly Hills: Phoenix Books, Inc.

Krumhuber, E., and Kappas, A. (2005). Moving smiles: the role of dynamic components for the perception of the genuineness of smiles. *J. Nonverbal Behav.* 29, 3–24. doi: 10.1007/s10919-004-0887-x

Langner, O., Dotsch, R., Bijlstra, G., Wigboldus, D. H., Hawk, S. T., and van Knippenberg, A. (2010). Presentation and validation of the Radboud Faces Database. *Cogn. Emot.* 24, 1377–1388. doi: 10.1080/02699930903485076

Larrouy-Maestri, P., Magis, D., and Morsomme, D. (2013). The effect of melody and technique on the singing voice accuracy of trained singers. *Logoped. Phoniatr. Vocol.* 1–4. doi: 10.3109/14015439.2013.777112

Lieberman, P. (1961). Perturbations in vocal pitch. *J. Acous. Soc. Am.* 33, 597–603. doi: 10.1121/1.1908736

Nawka, T., Anders, L. C., Cebulla, M., and Zurakowski, D. (1997). The speaker's formant in male voices. *J. Voice* 11, 422–428. doi: 10.1016/S0892-1997(97)80038-0

Ostwald, D. F. (2005). *Acting for Singers: Creating Believable Singing Characters*. New York, NY: Oxford University Press.

Pfordresher, P. Q., Brown, S., Meier, K. M., Belyk, M., and Liotti, M. (2010). Imprecise singing is widespread. *J. Acoust. Soc. Am.* 128, 2182–2190. doi: 10.1121/1.3478782

Plant, R. R., Hammond, N., and Turner, G. (2004). Self-validating presentation and response timing in cognitive paradigms: how and why? *Behav. Res. Methods Instrum. Comput.* 36, 291–303. doi: 10.3758/BF03195575

Raphael, L. J., Borden, G. J., and Harris, K. S. (2011). *Speech Science Primer: Physiology, Acoustics, and Perception of Speech*, 5th Edn. Philadelphia: Lippincott Williams & Wilkins.

Raudenbush, S., and Bryk, A. (2002). *Hierarchical Linear Models: Applications and Data Analysis Methods (Advanced Quantitative Techniques in the Social Sciences)*, 2nd Edn. Thousand Oaks, CA: Sage Publications, Inc.

Repp, B. H. (1992). Diversity and commonality in music performance: an analysis of timing microstructure in Schumann's "Träumerei". *J. Acoust. Soc. Am.* 92, 2546. doi: 10.1121/1.404425

Repp, B. H. (1995). Expressive timing in Schumann's "Träumerei": an analysis of performances by graduate student pianists. *J. Acoust. Soc. Am.* 98, 2413–2427. doi: 10.1121/1.413276

Russell, J. A., Bachorowski, J.-A., and Fernández-Dols, J.-M. (2003). Facial and vocal expressions of emotion. *Annu. Rev. Psychol.* 54, 329–349. doi: 10.1146/annurev.psych.54.101601.145102

Scherer, K. R. (1989). "Vocal correlates of emotional arousal and affective disturbance," in *Handbook of Social Psychophysiology* H. Wagner and A. Manstead (London: Wiley), 165–197.

Scherer, K. R. (2003). Vocal communication of emotion: a review of research paradigms. *Speech Commun.* 40, 227–256. doi: 10.1016/S0167-6393(02)00084-5

Scherer, K. R., Schaufer, L., Taddia, B., and Prégardien, C. (2013). "The singer's paradox: on authenticity in emotional expression on the opera stage," in *The Emotional Power of Music: Multidisciplinary Perspectives on Musical Arousal, Expression, and Social Control*, eds T. Cochrane, B. Fantini, and K. R. Scherer (New York: Oxford University Press), 55–74. doi: 10.1093/acprof:oso/9780199654888.003.0005

Singer, J. D. (1998). Using SAS PROC MIXED to fit multilevel models, hierarchical models, and individual growth models. *J. Educ. Behav. Stat.* 23, 323–355. doi: 10.2307/1165280

Smith, R. B. (1963). The effect of group vocal training on the singing ability of nursery school children. *J. Res. Music Educ.* 11, 137–141. doi: 10.2307/3344153

Snijders, T. A. B., and Bosker, R. J. (1999). *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling*. Newbury Park: Sage publications.

Sonnemans, J., and Frijda, N. H. (1994). The structure of subjective emotional intensity. *Cogn. Emot.* 8, 329–350. doi: 10.1080/02699939408408945

Sundberg, J. (1974). Articulatory interpretation of the "singing formant." *J. Acoust. Soc. Am.* 55, 838. doi: 10.1121/1.1914609

Sundberg, J. (2003). Research on the singing voice in retrospect. *TMH-QPSR* 45, 11–22.

Taylor, M. (2012). *Musical Theatre, Realism and Entertainment.* Surrey: Ashgate Publishing, Ltd.

Telfer, N. (1995). *Successful Warm-Ups* (*Book 1, Conductor's Edition Ed.*). San Diego, CA: Neil A. Kjos Music Company.

Ternstrom, S., Sundberg, J., and Collden, A. (1988). Articulatory *F*0 perturbations and auditory feedback. *J. Speech Lang. Hear. Res.* 31, 187–192. doi: 10.1044/jshr.3102.187

Timmermans, B., De Bodt, M. S., Wuyts, F. L., and Van de Heyning, P. H. (2005). Analysis and evaluation of a voice-training program in future professional voice users. *J. Voice* 19, 202–210. doi: 10.1016/j.jvoice.2004.04.009

Twisk, J. W. (2006). *Applied Multilevel Analysis: A Practical Guide.* Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511610806

Vogt, T., and André, E. (2005). "Comparing feature sets for acted and spontaneous speech in view of automatic emotion recognition," in *Proceedings of the IEEE International Conference on Multimedia and Expo, ICME 2005*, Amsterdam, 474–477.

Walzak, P., McCabe, P., Madill, C., and Sheard, C. (2008). Acoustic changes in student actors' voices after 12 months of training. *J. Voice* 22, 300–313. doi: 10.1016/j.jvoice.2006.10.006

Watts, C., Barnes-Burroughs, K., Andrianopoulos, M., and Carr, M. (2003). Potential factors related to untrained singing talent: a survey of singing pedagogues. *J. Voice* 17, 298–307. doi: 10.1067/S0892-1997(03)00068-7

Weiss, R., Brown, W. S., and Moris, J. (2001). Singer's formant in sopranos: fact or fiction? *J. Voice* 15, 457–468. doi: 10.1016/S0892-1997(01)00046-7

Wilcox, K. A., and Horii, Y. (1980). Age and changes in vocal jitter. *J. Gerontol.* 35, 194–198. doi: 10.1093/geronj/35.2.194

Willis, E. C., and Kenny, D. T. (2008). Effect of voice change on singing pitch accuracy in young male singers. *J. Interdiscip. Music Stud.* 2, 111–119.

Yumoto, E., Gould, W. J., and Baer, T. (1982). Harmonics-to-noise ratio as an index of the degree of hoarseness. *J. Acoust. Soc. Am.* 71, 1544–1550. doi: 10.1121/1.387808

Zebrowitz, L. A., Voinescu, L., and Collins, M. A. (1996). "Wide-Eyed" and "Crooked-Faced": determinants of perceived and real honesty across the life span. *Pers. Soc. Psychol. Bull.* 22, 1258–1269. doi: 10.1177/01461672962212006

Zuckerman, M., and Driver, R. E. (1989). What sounds beautiful is good: the vocal attractiveness stereotype. *J. Nonverbal Behav.* 13, 67–82. doi: 10.1007/BF00990791

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 12 September 2013; accepted: 08 February 2014; published online: 07 March 2014.

Citation: Livingstone SR, Choi DH and Russo FA (2014) The influence of vocal training and acting experience on measures of voice quality and emotional genuineness. Front. Psychol. ***5****:156. doi: 10.3389/fpsyg.2014.00156*

This article was submitted to Cognitive Science, a section of the journal Frontiers in Psychology.

frontiers in
PSYCHOLOGY

ORIGINAL RESEARCH ARTICLE
published: 04 February 2014
doi: 10.3389/fpsyg.2014.00025

Simulating and stimulating performance: introducing distributed simulation to enhance musical learning and performance

Aaron Williamon[1]*, *Lisa Aufegger*[1] *and Hubert Eiholzer*[2]

[1] Centre for Performance Science, Royal College of Music, London, UK
[2] Department of Research and Development, Conservatory of Southern Switzerland, Lugano, Switzerland

Edited by:
Bruno Gingras, University of Vienna, Austria

Reviewed by:
Dianna T. Kenny, The University of Sydney, Australia
Chia-Jung Tsay, University College London, UK

Correspondence:
Aaron Williamon, Centre for Performance Science, Royal College of Music, Prince Consort Road, London SW7 2BS, UK
e-mail: aaron.williamon@rcm.ac.uk

Musicians typically rehearse far away from their audiences and in practice rooms that differ significantly from the concert venues in which they aspire to perform. Due to the high costs and inaccessibility of such venues, much current international music training lacks repeated exposure to realistic performance situations, with students learning all too late (or not at all) how to manage performance stress and the demands of their audiences. Virtual environments have been shown to be an effective training tool in the fields of medicine and sport, offering practitioners access to real-life performance scenarios but with lower risk of negative evaluation and outcomes. The aim of this research was to design and test the efficacy of simulated performance environments in which conditions of "real" performance could be recreated. Advanced violin students ($n = 11$) were recruited to perform in two simulations: a solo recital with a small virtual audience and an audition situation with three "expert" virtual judges. Each simulation contained back-stage and on-stage areas, life-sized interactive virtual observers, and pre- and post-performance protocols designed to match those found at leading international performance venues. Participants completed a questionnaire on their experiences of using the simulations. Results show that both simulated environments offered realistic experience of performance contexts and were rated particularly useful for developing performance skills. For a subset of 7 violinists, state anxiety and electrocardiographic data were collected during the simulated audition and an actual audition with real judges. Results display comparable levels of reported state anxiety and patterns of heart rate variability in both situations, suggesting that responses to the simulated audition closely approximate those of a real audition. The findings are discussed in relation to their implications, both generalizable and individual-specific, for performance training.

Keywords: music education, distributed simulation, performance anxiety, performance science, virtual reality

INTRODUCTION

Exceptional musical performances require an ability to execute complex physical and mental skills on stage under intense pressure and public scrutiny. While many of these skills can be honed through deliberate practice (Ericsson et al., 1993), opportunities to gather experience on stage, which is simultaneously rich in contextual complexity and yet safe to allow musicians to experiment and develop artistically, are rare. Here, we report a new educational and training initiative aimed at creating realistic, interactive performance scenarios using simulation.

A high quality virtual environment should offer a three dimensional visual experience of weight, height, and depth and provide "an interactive experience visually in full real-time motion with sounds and possibly with tactile and other forms of feedback" (Roy, 2003, p. 177). The more elaborate the quality of the simulation, the more the feeling of immersion and perception of "reality" that is experienced by users. This also depends on the level of interactivity, dimensionality, accuracy, fidelity, and sensory input and output (Satava, 1993). When these features are addressed, simulation can be used as a performance tool to explore and study specific behaviors and as an educational tool to acquire and practice skills (Gallagher et al., 2005; Axelrod, 2007).

Thus far, successful application of virtual training environments has been shown in studies addressing fears of heights, flying, spiders, and public speaking. Slater et al. (2006), for instance, showed significant increased signs of anxiety in people with phobias when speaking in front of a neutrally behaving virtual audience. By comparing somatic and cognitive features, including the Personal Report of Confidence as a Speaker (PRCS) as well as heart rate measurements before, during, and after the performance, results suggest that such exposure can be an effective tool for treatment. A follow-up study by Pertaub et al. (2006) employed virtual audiences who could respond neutrally, positively, and negatively. Their verbal responses included expressions such as "I see" or "That's interesting," moving to more evaluative statements like "That's absolute nonsense." The audience was also able to provide non-verbal cues such as shifts in facial expression, changes in posture, and short animations including yawning, turning their heads, or walking out of the room. The results of comparing subjective measurements (e.g., PRCS) before and after virtual exposure showed that the way the audience responded directly affected participants' confidence as public speakers, in that positive audience reactions elicited higher confidence levels while the negative response caused a significant reverse effect.

Virtual environments have also been shown to be an effective tool for training elite performers such as pilots, athletes, and surgeons, especially when they have relatively little exposure to real-world performance contexts or when failure carries career- and/or life-threatening risks. In surgery, for example, such training, compared with more traditional methods, has been shown to reduce the number of errors in surgical procedures, enhance the rate and extent of skill acquisition, and improve planning strategies (e.g., Grantcharov et al., 2004; Xia et al., 2011). It not only helps trainees who lack experience in real-world surgical contexts but also advanced surgeons who require exposure to new procedures and technologies (Sutherland et al., 2006).

To a limited extent, virtual environments have been employed in music, where studies have tested their use in managing music performance anxiety. To date, the effects of exposure in these settings on psychological and physiological responses to performance stress are mixed. On the one hand, exposure to performing conditions using simulation has been shown to decrease state anxiety significantly compared with a control group, particularly for those musicians who score higher on *trait* anxiety measures (Bissonette et al., 2011). On the other, Orman (2003, 2004) found no discernible, consistent patterns of change in either self-reported anxiety levels or heart rate for musicians taking part in an intervention offering graded exposure to stressful performance situations. Such inconsistent results can also be found in other domains, for instance in studies examining surgical performance by Munz et al. (2004) and Torkington et al. (2001), who compared simulated training against more conventional training and no training at all. This emphasizes the importance of the backgrounds of individual participants, their levels of immersion (i.e., how realistically they experience the environment), and the quality of the interface on the overall effectiveness and range of uses of a virtual environment.

In terms of the studies conducted in music, it is possible that the inconsistent findings are due to differences in the fidelity and interactivity of the simulations used, but in addition, several methodological limitations should be highlighted in the extant research. Firstly, participants' use of anxiolytic medication and other substances that may have affected their perceptions of or physiological response to stress was not controlled. Secondly, while heart rate was measured in all studies, there were limitations with how the data were reported (Orman, 2003, 2004), or the results were not reported at all (Bissonette et al., 2011). Concerning the former, stressful events have been shown to influence temporal fluctuations of the peak-to-peak times in the electrocardiogram (ECG), the R-to-R (RR) interval, known as heart rate variability (HRV) (Berntson and Cacioppo, 2004). While the simplest measures of HRV are the mean of the RR time series and the standard deviation about its mean, both of these statistics are based on the *absolute* magnitude of the RR interval. However, in many applications, *relative* measures such as the power ratio of the low- and high-frequency components of HRV allow for more reliable comparisons between participants (for further information, see "Data treatment and analysis" and Williamon et al., 2013). Finally, the existing studies test the hypothesis that performance exposure using virtual environments can ameliorate psychological and physiological symptoms of performance anxiety; while this seems plausible on the surface, the inter- and intra-individual variability in how people experience and interpret such symptoms can be extremely large (Williamon, 2004). Given the multifaceted nature and impact of performance anxiety on musicians and the personal significance it holds even for highly experienced performers (Kenny, 2011), mere exposure to performance situations can be seen as the first of many steps in identifying and managing pernicious anxiety-related problems. Rather, it is possible that anxiety management will be only one of many possible uses of simulation training for enhancing musicians' learning and performance.

The aim of our research was to design, test, and explore possible uses of new interactive, simulated environments that provide salient cues from real-life performance situations—in this case, a recital and an audition. The environments were designed on the principles of "distributed simulation," in which only a *selective abstraction* of environmental features are provided. Distributed simulation has been tested and applied widely in the field of surgical education, where fully-immersive virtual operating theatres are expensive to build, maintain, and run. The findings suggest that, indeed, a simulated environment only requires few environmental cues above and beyond the interactive simulation of an injury, wound, etc. in order to produce significant advancements in learning surgical techniques compared with more common training methods (Kassab et al., 2011). For instance, these may include a scaled-down operating lamp, background sounds played through loud speakers, and even life-sized pictures of machines commonly found in operating theatres; as the surgeons themselves do not operate such equipment and must focus on performing the procedure at hand, the imitated (yet plausible) environmental features often go unnoticed while adding significantly to the level of immersion experienced by participants. This approach has resulted in low-cost, convincing simulations that are portable, widely accessible, and can be used in almost any available space. In our study, advanced violin students performed the same piece in performance settings that employed a selective abstraction of recital and audition environments. The musicians' perceptions and experience of performing in the environments were obtained through questionnaires, and for a subset of participants, ECG data were recorded during their performances.

MATERIALS AND METHODS

PARTICIPANTS

Eleven violinists (6 men, 5 women; mean age = 22.45 years, *SD* = 2.25) were recruited for the study through the Royal

College of Music's (RCM) student email list. They had been playing the violin for 16.18 years ($SD = 2.75$), including their first performance at age 6.50 ($SD = 2.20$). All students performed regularly ($M = 2.32$ times per month, $SD = 1.30$) and practiced on average for 3.93 h per day ($SD = 0.85$).

A subset of 7 participants from the larger sample (4 men, 3 women; mean age = 22.57 years, $SD = 2.50$) were selected based on their availability to take part in an additional performance for a real audition panel. These musicians had been playing for 16.00 years ($SD = 3.21$), with their first performance at age 7.57 ($SD = 2.14$). They performed for audiences on average 2.00 times per month ($SD = 1.15$) and practiced for 3.89 h per day ($SD = 0.93$).

This study was granted ethical approval by the Conservatoires UK Research Ethics Committee and was conducted according to ethical guidelines of the British Psychological Society. Informed consent was obtained from all participants, and no payment was given in exchange for participation.

THE PERFORMANCE SIMULATOR

The performance simulator was developed through a collaboration of the first and third authors and London-based design consultancy Studiohead. The aim was to generate back-stage and on-stage environments using a selective abstraction of key features consistent across a wide range of Western classical performance venues. The selection of these features was informed by interviews with advanced musicians on their experiences and perceptions of performing (see Clark et al., in press) and on further pilot interviewing focused on performance environments. By undertaking user research, interviewing staff and students from the RCM and watching performances from backstage, common features could be identified and then recreated, including interaction with a backstage manager, the ritual of walking on stage, using appropriate lighting and sound cues (see **Figure 1**), and providing realistic, interactive virtual audiences of different types and sizes.

With these features in mind, the simulator was designed along the principles of distributed simulation (i.e., low-cost and portable, with high fidelity) to operate in two modes:

1. a recital with 24 virtual audience members
2. an audition situation with three "expert" virtual judges

In both simulations, pre- and post-performance protocols were employed that matched those found at leading international performance venues—for instance, entrance to a "green room" for warm-up, stage calls at regular intervals, and scripted procedures for entering, bowing, and exiting the stage (see "Procedure"). A short introduction to the simulator is provided in Movie 1 (see **Supplementary Material**).

Recital simulation

To create an interactive audience and capture plausible audience behaviors, 11 concert-goers were filmed individually using green-screen technology sitting still while listening to a Western classical performance (with naturalistic body swaying, fidgeting movements, and coughing) and responding to a successful or unsuccessful performance with a standing ovation, enthusiastic applause, polite applause, or aggressive booing and displays of displeasure. They performed these tasks on the same approximate timeline, achieved by having them watch a video of an actor mounted next to the video camera and asking them to synchronize their behaviors with those of the actor. They were filmed individually so that audiences of different sizes could be compiled. For the purposes of this study, the audience consisted of 24 people seated in two blocks of chairs situated in a small auditorium designed using Adobe After Effects (CS5.5) (**Figure 2**). A Flash interface was created to enable the audience to be manipulated using pre-set control commands from a computer console located in the backstage area.

In terms of environmental cues, the backstage area was equipped with CCTV footage of audience members taking their

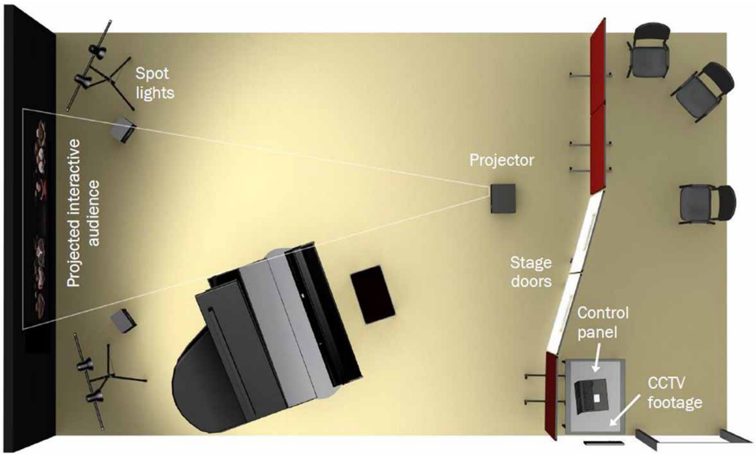


FIGURE 1 | The performance simulator, showing backstage (right) and stage (left) areas. Backstage, CCTV footage of the virtual audience or audition panel is shown on a wall-mounted flat-screen monitor, and controls for operating the audience and audition panel are located on a nearby computer. On stage, a ceiling-mounted beamer projects the life-sized audience or audition panel onto a wall, with spot-lights and loudspeakers on both sides. Stage curtains (not shown) frame the projected image.

seats in an auditorium (**Figure 2**). On stage, there were spot-lights and curtains on both sides of the projected audience. Noise distractions such as coughing, sneezing, and phone ringing were included in the stage area played through loudspeakers. Recordings of these, as well as all applause and booing, were made separately by another group of 16 volunteers in an anechoic chamber. Small recital hall reverberation was added to these recordings and then synchronized and layered on top of the video footage.

Audition simulation

The members of the simulated audition panel were three professional actors who were filmed together sitting behind a table. Starting each audition with a neutral "Hello, please start whenever you are ready," the panel showed typical evaluative behaviors such as making notes, looking pensively, and leaning back while simultaneously portraying positive, neutral, or negative facial expressions and behavioral feedback during the performance (e.g., smiling or frowning; leaning in toward the performer, or folding arms and leaning back). Each mode of listening was presented in loops of 5 min; the research team could change the mode of response after the loop or immediately. At the end of the performance, the audition panel could respond to the performance in a positive, neutral, or negative fashion—for instance, an enthusiastic "Thank you, that was excellent," a polite but non-committal "Thank you very much," a disappointed sigh followed by "Thank you for coming," or a disruptive "Thank you, I think we've heard enough" displaying displeasure and frustration. To enhance the level of interactivity between the panel members, the actors were filmed together in a small room, and then a virtual audition room with black background was added using Adobe After Effects (CS5.5) (**Figure 3**). Similar to the recital simulation, a Flash interface was created to enable the panel to be manipulated using pre-set control commands from a computer console located in the backstage area.

The environmental cues in the audition simulation included CCTV footage displayed backstage of the panel chatting among themselves in hushed voices. On stage, there were spot-lights, loudspeakers, and curtains on both sides of the projected panel.

MEASURES

Simulation evaluation questionnaire

A simulation evaluation questionnaire was based closely on work in surgical education by Kassab et al. (2011). It consisted of

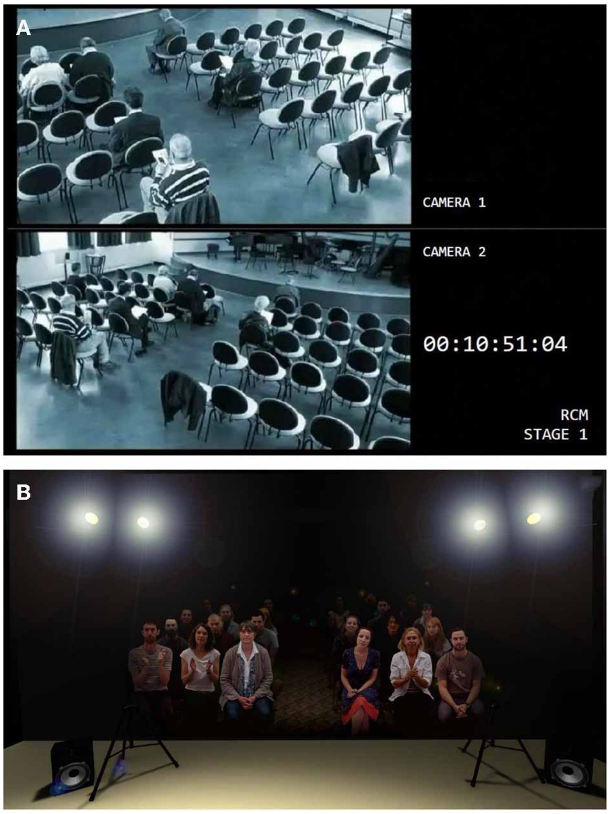


FIGURE 2 | (A) A still from the CCTV footage displayed in the backstage area of the recital simulation. **(B)** A still of the virtual audience projected in the stage area, framed by spot-lights, loudspeakers, and (not shown) stage curtains.

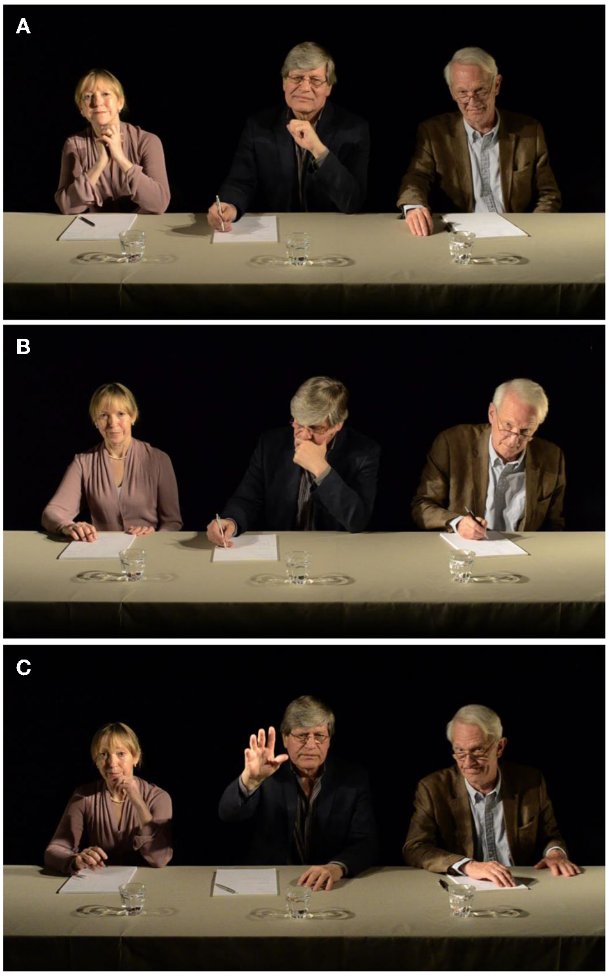


FIGURE 3 | Stills of the virtual audition panel **(A)** responding positively, **(B)** showing displeasure, and **(C)** stopping the audition.

19 statements about (1) general perceptions and experience of performing in the simulator, (2) the quality of the backstage experience, (3) the quality of the on-stage experience, and (4) the potential for using the simulator to develop performance skills. Each statement was rated on a 5-point Likert-type scale from 1 = “strongly disagree” to 5 = “strongly agree.” **Table 1** shows all 19 statements, as well as descriptive statistics (mean, median, standard deviation) for each.

State anxiety inventory

Immediately prior to each performance, participants were asked to complete Form Y1 (state anxiety) of the State-Trait Anxiety Inventory (STAI; Spielberger et al., 1970). This 20-item questionnaire measures one underlying construct showing the “temporal cross-section in the emotional stream of life of a person, consisting of subjective feelings of tension, apprehension, nervousness, and worry, and activation (arousal) of the autonomic nervous system” (Spielberger and Sydeman, 1994, p. 295). Responses to each question are made on a 4-point scale (1 = “almost never” to 4 = “almost always”), and after inversion of positively worded items, a cumulative score is calculated ranging from low (20) to high (80) state anxiety (for further details, see also Spielberger, 1983).

Electrocardiogram

For the subset of 7 participants, ECG data were collected before and during performances in the real and simulated auditions using a wireless Zephyr Bioharness. The device (80 × 40 ×

Table 1 | Descriptive statistics for questions in the Simulation Evaluation Questionnaire completed after performing in the recital and audition simulations.

Question	*Recital*				*Audition*			
	Median	*Mean*	*SD*	*p*	*Median*	*Mean*	*SD*	*p*
GENERAL PERCEPTIONS AND EXPERIENCE								
1. The simulation (including backstage, the audience, spot-lights, etc.) provided a realistic experience	4.0	3.42	1.08	ns	4.0	3.67	0.50	0.005
2. The steps involved in the simulation (i.e., waiting backstage, walking on stage, etc.) closely approximated a real performance situation	4.0	3.75	0.75	0.013	3.5	3.50	1.00	ns
3. I behaved and presented myself in the same way as I do in a real performance	4.0	3.42	1.24	ns	4.0	3.67	0.98	0.046
BACKSTAGE								
4. The interaction with the backstage manager was realistic	4.5	4.00	1.20	0.018	4.0	3.92	1.24	0.029
5. The CCTV footage in the backstage area was realistic	3.0	2.83	1.40	ns	4.0	3.58	1.56	ns
6. The sounds heard in the backstage area were realistic	3.5	3.42	1.08	ns	4.0	4.00	1.12	0.028
7. The decor of the backstage area (including signage and lighting) was realistic	4.0	3.67	0.49	0.005	3.5	3.42	1.08	ns
ON STAGE								
8. The transition from backstage on to stage was realistic	4.0	4.08	0.90	0.008	4.0	3.67	0.98	0.046
9. The interaction with the audience in the performance space was realistic	3.5	3.42	1.08	ns	3.5	3.42	1.08	ns
10. The spot-lights in the performance space were realistic	4.5	4.08	1.08	0.012	3.5	3.42	1.37	ns
11. The curtains in the performance space were realistic	3.0	3.33	0.98	ns	3.0	3.17	0.93	ns
SKILL DEVELOPMENT								
12. The simulation could be used to enhance my musical skills	4.0	4.25	0.86	0.005	4.0	4.25	0.75	0.004
13. The simulation could be used to enhance my technical skills	5.0	4.33	0.98	0.005	4.5	4.33	0.77	0.004
14. The simulation could be used to enhance my communicative/presentational skills	3.0	3.58	1.31	ns	4.0	4.17	0.83	0.006
15. The simulation could be used to help me manage performance anxiety and/or other performance problems	4.0	4.17	0.93	0.007	4.0	4.33	0.77	0.003
16. The simulation could be used to highlight strengths in my performance	4.5	4.33	0.88	0.004	4.0	4.17	0.71	0.004
17. The simulation could be used to highlight weaknesses in my performance	4.5	4.25	0.88	0.006	4.5	4.42	0.66	0.003
18. I would recommend the simulation to people who are interested in developing/refining their performance skills	5.0	4.33	1.10	0.005	4.5	4.42	0.66	0.003
19. I would recommend the simulation to people who are interested in teaching performance skills	5.0	4.42	1.16	0.004	4.5	4.33	0.77	0.004

Ratings for each statement were given from 1 = “strongly disagree” to 5 = “strongly agree.” The significance level p is shown for comparisons of the median rating for each question against a hypothesized median of 3, the scale mid-point, using the Wilcoxon signed-rank test.

15 mm, weight 35 g) snaps onto an elasticated chest belt (width 50 mm, weight 50 g) and has a sampling rate of 250 Hz. Tests of reliability and validity of the Bioharness have been conducted by Johnstone et al. (2012a,b).

PROCEDURE

At the start of the study, each of the 11 participants attended a 20-min induction session during which they were informed they would be giving repeated performances of the "Allemande" (with repeats) from J. S. Bach's *Partita No. 2 in D minor* (BWV 1004), a piece all had previously performed in public. Background information on musical experience was collected, and a preliminary health screening was conducted. On explicit questioning, all participants confirmed that they were not currently taking anxiolytic medication or other substances that may affect their perceptions of or physiological responses to performing. For the subset of 7 participants, the Bioharness was fitted and baseline heart rate data were collected. At the end of this session, participants were instructed not to consume alcohol or caffeinated drinks or to smoke for at least 2 h before their forthcoming performances.

Performance protocol

For each performance, participants were asked to arrive 20–30 min before their scheduled performance time. They were shown to a "green room" where they were fitted with the Bioharness (where applicable, $n = 7$) and allowed to engage in their usual pre-performance routine (e.g., warming up, practicing, stretching, etc.). Stage calls were given by a member of the research team—acting as the "backstage manager"—at 15 and 5 min before the performance.

At 0 min, each violinist was escorted to the backstage area, asked to complete Form Y1 of the STAI, and required to wait a further 5 min while the backstage manager carried out a scripted check that the stage furniture was correctly placed, the auditorium lights were set, and the audience/audition panel was ready for the performance to begin. During this time, the backstage area was dimly lit, and the participants had sight of the relevant CCTV footage and could hear the low murmur of talking from the audience/audition panel. At the end of these scripted checks, the backstage manager turned to the performer, confirmed that s/he was ready to go on stage, opened the stage door for the musician to walk out, and triggered the applause from the audience (recital) or the greeting from the panel (audition).

While on stage, participants bowed (recital) or returned the greeting (audition) and started their performance as soon as they felt ready. As this was the first experiment using the simulator, polite (but neutral) reactions to the performances were set for the virtual audience and audition panel and no deliberate distractions (coughing, sneezing, talking, phone ringing, etc.) were interjected. Following the end of the performance and shortly into the audience's/panel's response, the stage door opened as a sign to the participant to exit the stage.

For the subset of 7 participants, an additional real audition was organized, in which the same procedure was followed. For parity with the simulated audition, three real panel members (two men, one woman) were shown the footage used in the simulated audition and instructed to provide the exact same neutral response (physically and orally) to all performances.

Performances for all participants were scheduled at the same time on separate days. The order of condition (i.e., simulated recital, simulated audition, real audition) was counterbalanced.

DATA TREATMENT AND ANALYSIS

Simulation evaluation questionnaire

Initial inspection of data from the simulation evaluation questionnaire revealed that 13 of the 19 questions did not meet the criterion for normality (Shapiro Wilk). Therefore, responses to this questionnaire have been analyzed using non-parametric tests (SPSS v19).

Electrocardiogram

A key indicator of stress is the temporal fluctuation of the peak-to-peak times in the ECG—the R-to-R (RR) interval—known as heart rate variability (HRV) (Berntson et al., 1997; Berntson and Cacioppo, 2004). This can be studied through time- or frequency-domain analyses. While the time domain can be characterized through a simple calculation of the mean RR or its standard deviation, the frequency domain is studied using power spectral analysis examining HRV's low frequency (LF) and high frequency (HF) components: 0.04–0.15 and 0.15–0.4 Hz, respectively (Berntson and Cacioppo, 2004). The HF element is known to reflect activity of the parasympathetic nervous system (PNS), associated with homeostasis and balance, as well as the respiratory sinus arrhythmia (RSA), naturally occurring variations in heart rate due to breathing. The LF element, although more complex, reflects sympathetic nervous system (SNS) activity, associated with greater arousal. The ratio of LF to HF is widely used as an indicator of the sympatho-vagal balance, with an increase usually reflecting an elevated activation of physical or mental effort (Berntson and Cacioppo, 2004).

In terms of analyses, the mean HRV and its standard deviation are based on the absolute magnitude of the RR interval, whereas in many applications *relative* measures of a process have been shown to exhibit a greater consistency when comparing different individuals, for whom resting heart rate can vary considerably. The relative LF/HF power ratio, although the subject of recent debate as to whether it actually reflects the sympatho-vagal balance (Billman, 2013) and about its accuracy for dynamic stress level assessment (Williamon et al., 2013), has been widely used in the study of stress in performance (Nakahara et al., 2009; for a review, see Billman, 2013).

Here, the ECG data were analyzed using MATLAB (R2013a). Transformation of the RR into a continuous time series with an equivalent sampling frequency of 4 Hz was carried out using cubic spline interpolation as well as a median filter. Bandpass filtering was conducted (0.04–0.4 Hz) via a 4th order Butterworth filter before estimating signal dynamics via overlapping windows using the LF/HF ratio, with LF and HF components obtained via additional 4th order Butterworth filters. In the results below, mean LF/HF ratios are reported as measured at baseline and before and during each performance condition.

RESULTS

PERFORMING IN THE RECITAL AND AUDITION SIMULATIONS

For insight into participants' perceptions and experience of performing in the two simulations, the median rating for each item on the simulation evaluation questionnaire was compared against a hypothesized median of 3, the scale mid-point, using the one sample Wilcoxon signed-rank test (i.e., the non-parametric equivalent of the one-sample t-test). As shown in **Table 1**, the medians for all questions were either 3 or above, and the p-values indicate that the medians were significantly higher than 3 for 12 of 19 statements for the recital simulation (63.2%) and 13 of 19 for the audition simulation (68.4%). With median values significantly higher than 3 on the majority of statements, these results suggest that both simulations offered a high quality, realistic performance experience, although with some variation between them.

Focusing more closely on the skill development statements, the participants reported strong potential for both simulations to be used to enhance their own learning and performance skills, to manage performance anxiety and/or other performance problems, and to teach these skills to others. Indeed, 7 of 8 statements for the recital simulation and 8 of 8 for the audition simulation were significantly greater than 3.

In terms of reported anxiety, participants' mean state anxiety score for the recital simulation was 35.09 ($SD = 11.09$) and for the audition simulation was 34.09 ($SD = 7.09$) [according to Spielberger, 1983, moderate levels of anxiety are represented by scores of 36.47 ($SD = 10.02$) for male students and 38.76 ($SD = 11.95$) for female students]. A paired samples t-test showed no significant difference between state anxiety scores across the two simulations ($t_{10} = 0.33, p > 0.05$), suggesting that the average perceived anxiety before each simulated performance was comparable.

PERFORMING IN THE SIMULATED AND REAL AUDITIONS

For the subset of 7 participants, the LF/HF ratio was calculated from the ECG for (1) the last 5 min of their 20-min induction session (baseline), (2) 5 min immediately before their performance for the simulated and real auditions (pre-performance), and (3) their entire simulated and real auditions (performance). A repeated measures analysis of variance (ANOVA) was run with time of measurement (baseline vs. pre-performance vs. performance) and type of audition (simulated vs. real) as within-subjects variables. The results indicate a significant effect of time [$F_{2,12} = 21.01, p < 0.001$] and audition [$F_{1,6} = 9.94, p < 0.05$] and a significant interaction between time and audition [$F_{2,12} = 8.28, p < 0.01$]. Inspection of **Figure 4** suggests that these significant differences are mostly likely due to the high LF/HF ratio in the pre-performance period for the real audition. For levels of self-reported anxiety, participants' mean state anxiety score for the simulated audition was 32.29 ($SD = 7.39$) and for the real audition was 37.43 ($SD = 7.09$), suggesting moderate levels of perceived state anxiety (Spielberger, 1983). A paired samples t-test found no significant difference between state anxiety scores across the two auditions ($t_6 = -1.65, p > 0.05$), suggesting that the average perceived anxiety before each audition was comparable.

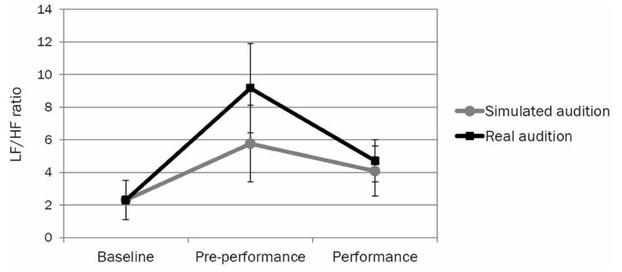


FIGURE 4 | The mean LF/HF ratio (with *SD*) calculated from ECG data collected for the violinists at baseline, pre-performance, and during performance for the simulated and real auditions.

DISCUSSION

The aim of this study was to design, test, and explore the possible uses of two new distributed simulation environments for enhancing musicians' learning and performance. These included a small recital and audition setting, both of which offered key visual, auditory, and other environmental cues commonly found backstage and on stage at international performance venues, as well as scripted protocols for how musicians should be guided through them. The violinists found the simulations to be realistic and convincing, with all median ratings for statements in the simulation evaluation questionnaire at the scale mid-point or above (and a majority of statements rated significantly above). In terms of potential uses of the simulations, the musicians were clear in recognizing their positive value for developing performance skills. Moreover, when examining participants' self-reported anxiety and physiological responses to performance stress in the simulated audition versus a real audition, there were no significant differences in state anxiety between the two conditions; while significant differences emerged in physiological responses for the *pre*-performance period, the direction of response (i.e., an increase in the LF/HF ratio) was the same for both, and there were no apparent differences during the performance. Reasons why the higher LF/HF ratio in the pre-performance period did not correspond to higher state anxiety scores could be manifold; musicians may experience the heightened physiological state which accompanies performance as either enabling or debilitating. Nonetheless, in this study comparable psychological and physiological responses were evoked across both auditions, real and simulated.

The distributed simulations used in this research drew upon a selective abstraction of salient procedural and environmental features. In both cases, an important element was the "backstage manager" who guided each musician through the scripted performance process and served as the gatekeeper for their transitions on and off the stage. The recital simulation benefitted particularly from the decor of the backstage area, including signage and lighting, as well as spot-lights in the performance space, while the simulated audition was enhanced especially by auditory cues played in the backstage area.

Unlike previous research in music (Orman, 2003, 2004; Bissonette et al., 2011), which focused on exposure to virtual environments as a means of managing performance anxiety, our results suggest that simulation training may offer wide ranging benefits for musical learning and performance. While further studies are needed to establish these benefits objectively, feedback elicited as part of the simulation evaluation questionnaire suggests that the simulations provided a path by which musicians could not only highlight their performance strengths but also address their weaknesses. As such, they were seen as potentially useful tools for teaching musical skills more efficiently. More generally, this research adds to the growing literature demonstrating that virtual environments can effectively aid learning, especially as they can be accessed repeatedly and consistently, at controlled levels of risk and with pre-defined outcomes. The extent to which the effectiveness of the present simulations will persist through repeated exposure remains to be tested, but findings from other domains would suggest that the prospects of using virtual environments repeatedly and longitudinally to enhance learning and performing are indeed promising (see Rothbaum et al., 2000, 2002).

This study has focused on responses to simulated recital and audition scenarios with relatively neutral and "well behaved" virtual observers. Subsequent studies should examine the influence of disruptions and distractions (e.g., coughing, sneezing, phone ringing, etc.), as well as more positive and negative observer responses. They should also consider recruiting larger samples of musicians, who can specifically be studied in low and high anxiety groups. In addition, there is scope for building and testing further simulated environments: from changing audience size or modifying the number and type of environmental features to involving more performers or altering the performance task itself. It would also be instructive to test whether the degree of background knowledge about the development of the simulations or prior knowledge of whether the performer will encounter a real or virtual audience before they enter the stage (i.e., a blind experiment) would impact how musicians perceive and interact with each simulation.

In general terms, one could argue that distributed simulation training, as employed in this study, has much to offer musicians. However, difficulty arises in mapping out precisely *how* to use simulation in ways that are meaningful to those at different levels of skill, with varying degrees and types of performance exposure, and with very personal experiences of performance anxiety symptoms. Further experimental work that systematically addresses these parameters, carried out using psychophysiological and questionnaire-based measures extended from the current study, could provide such insight.

In this respect, subsequent investigations are needed that explore both generalizable and individual-specific benefits of simulation training. Research should systematically test the effects of simulation on skill acquisition, planning, and self-regulatory strategies, and techniques for improving communication and presentational skills, alongside interventions for managing anxiety. It should also investigate how to match certain types, lengths, and intensities of training to individual musicians' learning and performance objectives. When on stage, it is precisely these objectives that distinguish one performance from another.

AUTHOR CONTRIBUTIONS

All authors contributed extensively to the work presented in this paper.

ACKNOWLEDGMENTS

This research was supported by grants from the Royal College of Music and from the Swiss State Secretariat for Education, Research and Innovation. We wish to thank our participants for giving generously of their time and for their feedback; Matthew Harrison and Cian Plumbe of Studiohead for their design expertise; Roger Kneebone of Imperial College London for guidance on distributed simulation; Seb Jouan of Arup Acoustics and Stephen Dance of London South Bank University for assistance in recording sounds used in the recital simulation; Robin Browne, William Hoyland, and Norma Jones of the Harry Partnership for playing the panel members in the audition simulation; and Karl Altenburg, Benjamin Cracknell, Michaela Korte, Sylvia Liu, Rosie Perkins, and George Waddell for help with data collection.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: http://www.frontiersin.org/journal/10.3389/fpsyg.2014.00025/abstract

REFERENCES

Axelrod, R. (2007). "Simulation in the social science," in *Handbook of Research on Nature Inspired Computing for Economics and Management*, ed R. Jean-Phillipe (Hershey, PA: Idea Group Pub.), 90–100.

Berntson, G. G., Bigger, J. J. T., Eckberg, D. L., Grossman, P., Kaufmann, P. G., Malik, M., et al. (1997). Heart rate variability: origins, methods, and interpretive caveats. *Psychophysiology* 34, 623–648. doi: 10.1111/j.1469-8986.1997.tb02140.x

Berntson, G. G., and Cacioppo, J. T. (2004). "Heart rate variability: stress and psychiatric conditions," in *Dynamic Electrocardiograph*, eds M. Malik and J. Camm (Hoboken, NJ: Wiley-Blackwell), 56–63.

Billman, G. E. (2013). The LF/HF ratio does not accrately measure cardiac sympatho-vagal balance. *Front. Physiol.* 4:26. doi: 10.3389/fphys.2013.00026

Bissonette, J., Dube, F., Provencher, M. D., and Moreno Sala, M. T. (2011). "The effect of virtual training on music performance anxiety," in *Proceedings of the International Symposium on Performance Science 2011*, eds A. Williamon, D. Edwards, and L. Bartel (Utrecht: European Association of Conservatoires), 585–590.

Clark, T., Lisboa, T., and Williamon, A. (in press). An investigation into musicians' thoughts and perceptions during performance. *Res. Stud. Music Educ.*

Ericsson, K. A., Krampe, R. T., and Tesch-Römer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychol. Rev.* 100, 363–406. doi: 10.1037/0033-295X.100.3.363

Gallagher, A. G., Ritter, E. M., Champion, H., Higgins, G., Fried, M. P., Moses, G., et al. (2005). Virtual reality for the operating room: proficiency-based training as a paradigm shift in surgical skills training. *Ann. Surg.* 241, 364–372. doi: 10.1097/01.sla.0000151982.85062.80

Grantcharov, T. P., Kristiansen, V. B., Bendix, J., Bardram, L., Rosenberg, J., and Funch-Jensen, P. (2004). Randomized clinical trial of virtual reality simulation for laparoscopic skills training. *Br. J. Surg.* 91, 146–150. doi: 10.1002/bjs.4407

Johnstone, J. A., Ford, P. A., Hughes, G., Watson, T., and Garrett, A. T. (2012a). BioHarness multivariable monitoring device. Part I: validity. *J. Sports Med.* 11, 400–408.

Johnstone, J. A., Ford, P. A., Hughes, G., Watson, T., and Garrett, A. T. (2012b). BioHarness multivariable monitoring device. Part II: reliability. *J. Sports Med.* 11, 409–417.

Kassab, E., Tun, J. K., Arora, S., Kind, D., Ahmed, K., Miskovic, D., et al. (2011). Blowing up the barriers in surgical training: exploring and validating the concept of distributed simulation. *Ann. Surg.* 254, 1059–1065. doi: 10.1097/SLA.0b013e318228944a

Kenny, D. T. (2011). *Psychology of Music and Performance Anxiety.* Oxford: Oxford University Press. doi: 10.1093/acprof:oso/9780199586141.001.0001

Munz, Y., Kumar, B. D., Moorthy, K., Bann, S., and Darzi, A. (2004). Laparoscopic virtual reality and box trainers. *Surg. Endosc.* 18, 485–494. doi: 10.1007/s00464-003-9043-7

Nakahara, H., Furuya, S., Obata, S., Masuko, T., and Kinoshita, H. (2009). Emotion-related changes in heart rate and its variability during performance and perception of music. *Ann. N.Y. Acad. Sci.* 1169, 359–362. doi: 10.1111/j.1749-6632.2009.04788.x

Orman, E. K. (2003). Effect of virtual reality graded exposure on heart rate and self-reported anxiety levels of performing saxophonists. *J. Res. Music Educ.* 51, 302–315. doi: 10.2307/3345657

Orman, E. K. (2004). Effect of virtual reality graded exposure on anxiety levels of performing musicians: a case study. *J. Music Ther.* 41, 70–78. doi: 10.1093/jmt/41.1.70

Pertaub, D. P., Slater, M., and Barker, C. (2006). An experiment on public speaking anxiety in response to three different types of virtual audience. *Presence Teleop. Virt. Environ.* 11, 68–78. doi: 10.1162/105474602317 343668

Rothbaum, B. O., Hodges, L., Smith, S., Lee, J. H., and Price, L. (2000). A controlled study of virtual reality exposure therapy for the fear of flying. *J. Consult. Clin. Psychol.* 68, 1020–1026. doi: 10.1037/0022-006X.68. 6.1020

Rothbaum, B. O., Hodges, L., Anderson, P. L., Price, L., and Smith, S. (2002). Twelve-month follow-up of virtual reality and standard exposure therapies for the fear of flying. *J. Consult. Clin. Psychol.* 70, 428–432. doi: 10.1037/0022-006X. 70.2.428

Roy, S. (2003). State of the art of virtual reality therapy (VRT) in phobic disorders. *Psychnol. J.* 2, 176–183.

Satava, R. M. (1993). Virtual reality surgical simulator. *Surg. Endosc.* 7, 203–205. doi: 10.1007/BF00594110

Slater, M., Pertaub, D.-P., Barker, C., and Clark, D. M. (2006). An experimental study on fear of public speaking using a virtual environment. *Cyberpsychol. Behav.* 9, 627–633. doi: 10.1089/cpb.2006.9.627

Spielberger, C. D. (1983). *State-Trait Anxiety Inventory STAI (Form Y)*. Palo Alto, CA: Consulting Psychologists Press, Inc.

Spielberger, C. D., Gorsuch, R. L., and Lushene, R. E. (1970). *Manual for the State-Trait Anxiety Inventory.* Palo Alto, CA: consulting Psychologists Press.

Spielberger, C. D., and Sydeman, S. J. (1994). "State-trait anxiety inventory and state-trait anger expression inventory," in *The Use of Psychological Testing for Treatment Planning and Outcome Assessment*, ed M. E. Maruish (Hillsdale, NJ: Lawrence Erlbaum Associates), 292–321.

Sutherland, L. M., Middleton, P. F., Anthony, A., Hamdorf, J., Cregan, P., Scott, D., et al. (2006). Surgical simulation: a systematic review. *Ann. Surg.* 243, 291–300. doi: 10.1097/01.sla.0000200839.93965.26

Torkington, J., Smith, S. G. T., Rees, B. I., and Darzi, A. (2001). Skill transfer from virtual reality to a real laparoscopic task. *Surg. Endosc.* 15, 1076–1079. doi: 10.1007/s004640000233

Williamon, A. (2004). *Musical Excellence.* Oxford: Oxford University Press. doi: 10.1093/acprof:oso/9780198525356.001.0001

Williamon, A., Aufegger, L., Wasley, D., Looney, D., and Mandic, D. P. (2013). Complexity of physiological responses decreases in high stress musical performance. *J. R. Soc. Interface* 10, 1–6. doi: 10.1098/rsif.2013.0719

Xia, J. J., Shevchenko, L., Gateno, J., Teichgraeber, J. F., Taylor, T. D., Lasky, R. E., et al. (2011). Outcome study of computer-aided surgical simulation in the treatment of patients with craniomaxillofacial deformities. *J. Oral Maxillofac. Surg.* 69, 2014–2024. doi: 10.1016/j.joms.2011.02.018

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 09 September 2013; accepted: 09 January 2014; published online: 04 February 2014.

Citation: Williamon A, Aufegger L and Eiholzer H (2014) Simulating and stimulating performance: introducing distributed simulation to enhance musical learning and performance. Front. Psychol. ***5****:25. doi: 10.3389/fpsyg.2014.00025*

This article was submitted to Cognitive Science, a section of the journal Frontiers in Psychology.

frontiers in
PSYCHOLOGY

ORIGINAL RESEARCH ARTICLE
published: 10 December 2013
doi: 10.3389/fpsyg.2013.00921

Musical feedback during exercise machine workout enhances mood

Thomas H. Fritz[1,2,3] *****, **Johanna Halfpaap**[1], **Sophia Grahl**[1], **Ambika Kirkland**[1] **and Arno Villringer**[1]

[1] Max Planck Institute for Human Cognitive and Brain Science, Leipzig, Germany
[2] Institute for Psychoacoustics and Electronic Music, Gent, Belgium
[3] Department of Nuclear Medicine, University of Leipzig, Leipzig, Germany

Edited by:
Bruno Gingras, University of Vienna, Austria

Reviewed by:
Adrian Von Muhlenen, University of Warwick, UK
Jacques Launay, University of Oxford, UK

***Correspondence:**
Thomas H. Fritz, Max Planck Institute for Human Cognitive and Brain Sciences, Stephanstrasse 1a, 04103 Leipzig, Germany
e-mail: fritz@cbs.mpg.de

Music making has a number of beneficial effects for motor tasks compared to passive music listening. Given that recent research suggests that high energy musical activities elevate positive affect more strongly than low energy musical activities, we here investigated a recent method that combined music making with systematically increasing physiological arousal by exercise machine workout. We compared mood and anxiety after two exercise conditions on non-cyclical exercise machines, one with passive music listening and the other with musical feedback (where participants could make music with the exercise machines). The results showed that agency during exercise machine workout (an activity we previously labeled jymmin – a cross between jammin and gym) had an enhancing effect on mood compared to workout with passive music listening. Furthermore, the order in which the conditions were presented mediated the effect of musical agency for this subscale when participants first listened passively, the difference in mood between the two conditions was greater, suggesting that a stronger increase in hormone levels (e.g., endorphins) during the active condition may have caused the observed effect. Given an enhanced mood after training with musical feedback compared to passively listening to the same type of music during workout, the results suggest that exercise machine workout with musical feedback (jymmin) makes the act of exercise machine training more desirable.

Keywords: music therapy, exercise, agency, mood, emotional motor control, jymmin, esthetics

INTRODUCTION

A number of factors have been shown to beneficially modulate the performance of motor tasks in combination with music. Acoustic factors in music during a walking task have been shown to modulate the stride length of participants walking in synchrony with the music, thus entraining the speed of beat synchronized walking (Leman et al., 2013). Because the walking was performed in synchrony with the beat and all musical stimuli had the same duration and tempo, the differences in walking speed could only have been the result of music-induced differences in stride length, reflecting the vigor or physical strength of the movement. This is supported by other research showing that participants walked faster on music than on metronome ticks (Styns et al., 2007), and that the sound pressure level of the bass drum in dance music leads to more intense spontaneous hip movements and a higher degree of time entrainment (Van Dyck et al., 2013).

Furthermore, listening to music can have a regulating effect on movement performance in movement disorders such as Parkinson's disease, where it could be shown that rhythmical auditory cues promote movement speed up to 25% (Thaut et al., 1996; McIntosh et al., 1997) and stride length up to 12% (Thaut et al., 1996). Such a positive effect of musical parameters on the motor behavior of patients with Parkinson's disease has been reported both immediately during walking (Enzensberger et al., 1997), and as a long-term therapeutical effect (after weeks of training) (Thaut et al., 1996; De Bruin et al., 2010).

Interpersonal proximity in socially shared space will have an automatic effect on the motor behavior of participants and will, for example, increase unintentional coordinated tapping (Wu et al., 2013). Moreover, social interaction in comparison to performing alone or with a simulated partner seems to have a beneficial influence on musical motor performance. When children (of all ages) drum along with either a human partner, a drumming machine, or a drum sound coming from a speaker, they best synchronize in the social condition (Kirschner and Tomasello, 2009), suggesting that a shared representation of a musical goal may modify (in this scenario improve) motor behavior. A modulating influence as observed from social interaction to musical production can also be observed vice versa. Joint music making, including joint singing and dancing, more strongly promoted prosocial behavior such as spontaneous helping and spontaneous cooperative problem solving in 4-year olds (Kirschner and Tomasello, 2010).

Physical workout as a motor activity is known to be healthy, and a therapeutic intervention with physical workout is known to be highly effective for a variety of illnesses and disorders such as depression and anxiety disorders (Ströhle, 2009), multiple sclerosis (Motl et al., 2009), and chronic pain (Vendrig and Lousberg, 1997). It is possible to systematically train distinct muscle groups with specifically developed exercise machines as used in many conventional fitness studios, but also with exercise machines specifically tailored to certain bodily deficits (e.g., after stroke). However, working out with exercise machines is often perceived as rather boring and is rejected by many (Ruby et al., 2011), especially because of the stereotype movements required to do the exercises. Sport performers often listen to music passively to make rather repetitive sports activity more pleasurable. This can have positive effects on sports performance (Terry et al., 2012), which is why athletes often use music in fitness studios and during the preparation of sport competitions (Lim et al., 2009). The musical parameters tempo and rhythm, for example, have been shown to have a motivating and ergogenic influence on performance in sports (Karageorghis et al., 1999; Simpson and Karageorghis, 2006). Furthermore, it has been previously reported that people who listen passively to music while working out may be able use their muscles more effectively (Szmedra and Bacharach, 1998). Supporting the idea that listening to music benefits physical activity, evidence has suggested that multi-year amateur dancing helps preserve cognitive, motor, and perceptual abilities, and thus everyday life competence of elderly individuals (Kattenstroth et al., 2010).

Making music as a typically interactive social activity that involves agency and synchronization has the potential to integrate these factors that are beneficial during motor tasks (as described above). Making music, in comparison to only passively listening to music, thus probably mediates another set of beneficial effects for the performer. It has been reported to have a relaxation effect on performers similar to well-established recreation practices such as progressive muscle relaxation (Kibler and Rider, 1983). It probably has an activating effect on the immune system, as evidenced by measuring Salivary Immunoglobulin A (SlgA) during instrument playing or singing, in comparison to passive music listening (Kuhn, 2002; Kreutz et al., 2004), and an increase of cortisol levels during choir singing. Furthermore, it could be shown that music can increase perceived movement motivation and experienced mood (van der Vlist et al., 2011).

Notably, it has been shown that high energy musical activities seem to more strongly induce a release of endorphins, as measured using pain threshold. Singing, dancing, and drumming all increased the post-activity pain tolerance, and also resulted in elevated positive affect (Dunbar et al., 2012). A recent study explored a novel method to make music while systematically increasing physiological arousal by exercise machine workout (Fritz et al., 2013). This study showed evidence for a massive decrease in perceived exertion when combining workout with musical agency (jymmin) and also an increased motor effectivity when comparing this condition to a control condition where music was listened to passively during workout (similar to conventional training in fitness studios). Note that here we adapt a definition of agency as a performance of bodily movement guided by an agent, and governed by a goal or intention. We employ the term musical agency, because the goal/intention of the musical agency condition is a modulation of musical sounds. In the current study we expected to find effects of musical agency during workout (jymmin) on perceived mood and anxiety compared to passively listening to music during workout. To this end, the same set of machines was used as in the previous study, which allowed for a musical feedback during guided movements on fitness machines.

MATERIALS AND METHODS

PARTICIPANTS

Fifty-two participants (27 male) took part in the experiment. The age range was 20–49 years for males (mean = 27.07) and 20–43 years for females (mean = 27.33). Participants used the fitness machines in groups of three. None of the participants were professional athletes, body builders, or musicians.

EXPERIMENTAL DESIGN

The experiment comprised two conditions: In one condition the participants operated fitness machines while passively listening to music ("passive listening" condition); in a second condition they operated the fitness machines while listening to the musical feedback of their movements ("musical agency" condition). The physical workout was conducted with three different fitness machines, a tower (for a depiction see Fritz et al., 2013), a stomach trainer, and a stepper. All three machines are standard fitness machines that are commercially available from several companies, all allow for guided movements, and feature joints where sensors can easily be attached. The tower features a metal bar attached to weights via a cable. Performers pull down on the metal bar in a well-controlled movement to train the biceps. In the stomach trainer the legs can be fixed in an angled position, and sit-ups are performed, where the movement can be supported by moving the back support of the machine forward while pushing/pulling a semicircular handle with the arms. This way both stomach and arms muscles can be trained, so that the very exhausting sit-up movement can be performed over a longer period of time. The stepper consists of two platforms, one for each foot, so that the performer stands on the machine. Performers push down on the platform with first the left and then the right foot shifting their body weight so that the platforms move accordingly. A resistance is introduced so that the workout becomes more exhausting.

In the "musical agency" condition the movement of the fitness machines was mapped to a musical composition software (Ableton Live 8) so that the deflection of the fitness machines corresponded to musical parameters of an acoustic feedback signal. In the music composition software we prepared a series of musical loops (either wav-audio files or midi sequences), which were set to repeat, and were set to temporally synchronize at a constant tempo of 130 bpm. The style of the music composition used in the experiment was rather simplistic electronic (dance) music. In the composition software, several of the loops could be attributed to one of the "fitness instruments." For each of these loops, a track with an effect section was created. The movements of each fitness machine were mapped to modulate different parameters for each loop, which were specified in the effects section of each track. Thus, different loops (audio- and midi-) could be influenced by several audio effects by each fitness machine simultaneously. The effects and loops were chosen so that even relatively small movements in the centimeter range created a perceivable musical effect for the performer, culminating in one interesting musical dimension associated with each fitness instrument. Composition effects used in the experiment were bandpass filter (Ableton VST plugin autofilter; the cutoff of the bandpass filter has a strong effect on the perceived timbre of the audio signal; this was used on all three fitness machines), and pitch shift in association with the Ableton

VST plugin scale stance filter (allowed for the generation of simple melodies within a scale; this VST plugin was used on the tower). For example, on the tower the cutoff frequencies of two bandpass filters (located in the effects section of two different tracks) were set to modulate a driving techno beat and a bassline (each located in one of the two respective tracks). The cutoff frequencies of the bandpass filters were mapped to the movements of the tower. This meant that in the absence of exerted power, the cutoff was very low (so that no sound except some very deep bass frequencies was audible) and increased in relation to the distance the weights were pulled (so that simultaneously the bassline and the drum loops would blend in their higher frequency spectrum; an effect often used in dance music). While on the bassline this effect is strongly perceived on the low and medium spectral range, on the beat it is also strongly perceived in the high frequency range where cymbals are effected. In the high frequency range additionally the pitch of a midi loop (triggering a synthesizer) was effected, so that a simple melody on a software-synthesizer could be created by moving the weights in the top range of displacement.

We call this musical feedback technology "jymmin," a cross between jammin and gym. The musical feedback was designed so that the three fitness machines created sounds that could be interactively combined into a holistic musical piece at a constant tempo of 130 bpm. The musical interaction was constrained for its degrees of freedom, by predefining the sounds to be modulated, the parameters to be modulated, and the pulse of the music. Although the musical piece had a clear metric pulse as defined by beat loops manipulated both on the stepper and the tower, the movements of the participants were not strictly coupled to the meter, but could, for example, include slow bandpass filter movements over several (musical) measures.

Note that the musical soundscape in both conditions was comparable, because the music to which participants worked out in the "passive listening" condition was created by musical interactions similar to the "musical agency" condition by nine different groups of participants (ensuring an ecological validity of the passive listening musical baseline signal), which were recorded previous to the experiment. We controlled for the sequence in which the conditions were performed, so that half the participants first performed "passive listening," and the other half first performed "musical agency."

EXPERIMENTAL PROCEDURE

The participants from all groups met for the first time during the experiment and were asked to choose their preferred fitness machines. The task for all conditions was identical: "Use the fitness machines in a way in which you are physically comfortable." All participants filled out general information questionnaires to assess sex and age, the short form of the Multidimensional Mood Questionnaire (MDMQ; Mehrdimensionaler Befindlichkeitsfragebogen; Steyer et al., 1997), and the State-Trait-Anxiety-Inventory (STAI; State-Trait-Angstinventar; Laux et al., 1981) to assess a baseline of their mental state before the experiment. Each condition was performed for 10 min; a speaker system was used so that the sound was heard by all participants. Although 10 min of fitness training may seem a rather short duration (often fitness training is undertaken for 15–20 min), we chose this shorter time frame because (1) participants perform two conditions, so the time adds up to 2 × 10 = 20 min and we did not want to overexert participants, and (2) we wanted to avoid the risk that a too lengthy musical piece might with time become boring for the participants (an extra long disco mix track usually lasts a maximum of around 10 min). Each condition was followed by a 10 min break to relax and again fill out the STAI questionnaire; additionally after each condition the Multidimensional Mood Questionnaire (MDMQ; Mehrdimensionaler Befindlichkeitsfragebogen; Steyer et al., 1997) was filled out. The MDMQ is a mental state assessment instrument that measures the current mental state with three bipolar designed subscales (good/bad mood, alertness/fatigue, calmness/restlessness) with a total of 12 items. Each subscale consists of four adjectives, of which two belong to the negative (e.g., "bad," "uncomfortable") and two to the positive (e.g., "well," "satisfied") pole. These items are rated on a five-point Likert-scale with one ("not at all") to five ("very"). The MDMQ is designed for adolescents as well as adults and can be used to evaluate the progress of mood influencing therapies and interventions. The short form of the test, which is used here can be filled out in less than 6 min and the possible scores are between 12 (bad mental state) and 40 (good mental condition). Despite the brevity of the MDMQ, it has been shown to be highly reliable (Heinrichs and Nater, 2002), and the internal consistency (Cronbach's alpha) of the short form scales are between $\alpha = 0.73$ and $\alpha = 0.89$. The STAI can measure both anxiety as a trait and anxiety as a state. For the current study we were interested in measuring state anxiety. The questionnaire consists of 20 items describing the current mental state of the participants (for example "I feel nervous." or "I feel relaxed."). The items are rated on a four-point-Likert scale from one ("not at all") to four ("very"). The questionnaire is well approved in research and clinical practice and shows high internal consistency (Cronbach's alpha of $\alpha = 0.90$).

DATA ANALYSIS

The behavioral data were analyzed using SPSS 18 (IBM). Participants who had missing responses because a questionnaire was filled out incompletely were excluded from the analysis. In total, data from the MDMQ were analyzed for 45 participants and data from the STAI were analyzed for 51 participants.

In order to obtain mean scores for each subscale on the MDMQ, responses to the four questions corresponding to each subscale were averaged. These mean scores range from 1 to 5, reflecting the one to five Likert scale used to rate each item. Items belonging to the negative pole (e.g., "bad" or "uncomfortable") were reverse scored, (e.g., higher scores on items relating to bad mood result in a lower score). Therefore, a higher mean score on the "good vs. bad mood" subscale corresponds to a better mood, a higher mean score on the "calmness vs. agitation" subscale indicates a greater degree of calmness, and a higher mean score on the "alertness vs. tiredness" subscale indicates a higher level of alertness.

A multivariate analysis of variance (MANOVA) with repeated measures was performed to test the effect of condition (musical agency vs. passive listening) and condition order (passive listening first vs. musical agency first) on the linear combination of the three subscales of the mood and activation questionnaire which was rated after each condition. Condition was a within-subjects variable and condition order varied between subjects. Follow-up univariate analysis of variance (ANOVA) was performed for each of the three subscales to determine whether differences in mood or differences in activation were driving the effect.

A repeated-measures ANOVA was also carried out to test the effect of condition (baseline, musical agency, and passive listening) and condition order (passive listening first vs. musical agency first) on anxiety, as measured by the STAI.

RESULTS

A MANOVA revealed that the main effect of condition on the three mood subscales was statistically significant, Pillai's Trace = 0.21, $F(3, 41) = 3.56$, $p < 0.05$. The main effect of condition order was not significant, Pillai's Trace = 0.91, $F(3, 41) = 1.36$, $p = 0.27$; the interaction between condition and condition order, however, was significant, Pillai's Trace = 0.23, $F\ (3, 41) = 4.10$, $p < 0.05$.

Follow-up univariate ANOVAs showed a significant main effect of condition on the "good vs. bad mood" subscale, $F(1, 43) = 10.67$, $p < 0.05$. As shown in **Figure 1**, the mean score on this subscale was higher following the "musical agency" condition ($M = 3.63$, SD = 0.99) than following the "passive listening" condition ($M = 3.82$, SD = 1.03). However, condition did not have a significant main effect on either the "calmness vs. agitation" subscale, $F(1, 43) = 1.83$, $p = 0.18$ or the "alertness vs. tiredness" subscale, $F(1, 43) = 0.46$, $p = 0.50$.

The univariate main effect of condition order was significant for the "calmness vs. agitation" subscale, $F(1, 43) = 4.12$, $p < 0.05$. As **Figure 2** shows, participants scored higher on this subscale when they performed the "musical agency" condition first ($M = 3.81$, SD = 0.94) than when they performed "passive listening" first ($M = 3.51$, SD = 0.95). There was no significant effect of condition order for the "good vs. bad mood" subscale, $F(1, 43) = 2.34$, $p = 0.13$ or for the "alertness vs. tiredness" subscale, $F(1, 43) = 2.50$, $p = 0.12$.

The univariate interaction between condition and condition order was significant for the "good vs. bad mood" subscale, $F(1, 43) = 8.24$, $p < 0.05$. As seen in **Figure 3**, the difference between the "passive listening" condition and the "musical agency" condition was greater when the "passive listening" condition came first. There was no significant interaction for the "calmness vs. agitation" subscale, $F(1, 43) = 1.10$, $p = 0.30$, or for the "alertness vs. tiredness" subscale, $F(1, 43) = 1.03$, $p = 0.32$.

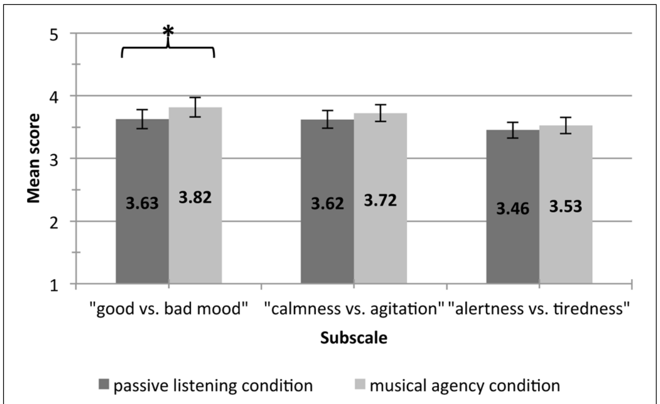


FIGURE 1 | The figure depicts mean ratings for each Multidimensional Mood Questionnaire subscale by condition; error bars show the standard error and *indicates a significant univariate effect with a *p*-value of <0.05.

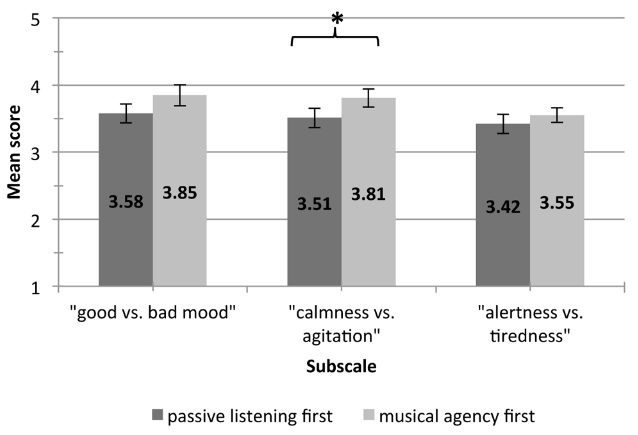


FIGURE 2 | The figure depicts mean ratings for each Multidimensional Mood Questionnaire subscale by condition order; error bars show the standard error and *indicates a significant univariate effect with a *p*-value of <0.05.

Results for the anxiety questionnaire ANOVA showed a significant effect of condition, $F(2, 98) = 8.32$, $p < 0.05$. The mean STAI scores for each condition can be seen in **Figure 4**. Tests of within-subject contrasts revealed that both "passive listening" and "musical agency" lowered anxiety relative to the baseline. Anxiety scores after "passive listening" ($M = 33.76$, SD = 6.88) were significantly lower than the baseline ($M = 35.8$, SD = 6.28), $F(1, 49) = 7.54$, $p < 0.05$. Anxiety scores following the "musical agency" condition ($M = 33.02$, SD = 7.33) were also lower than the baseline, $F(1, 49) = 13.02$, $p < 0.05$. However, anxiety scores for the "passive listening" condition were not significantly different from anxiety scores for the "musical agency" condition, $F(1, 49) = 1.64$, $p = 0.21$. The main effect of condition order on anxiety was not significant, $F(1, 49) = 0.01$, $p = 0.93$, nor was the interaction between condition and condition order, $F(2, 98) = 1.04$, $p = 0.357$.

DISCUSSION

The data indicate that musical agency during exercise machine workout can more strongly increase the subjectively perceived mood. Findings from a recent study suggest that the physical exertion during musical performance is proportionate to endorphin release (Dunbar et al., 2012). Authors also reported corresponding changes in positive affect. Note that in the current study we could compare the musical agency condition with a control condition that was comparably high with respect to physiological activity. In fact in a different experiment, which is not reported here where we measured spirometry, we found that physiological activity during a stereotypical workout control condition as used in the current experiment is even higher than during the musical agency condition (Fritz et al., 2013). Thus it appears that the amount

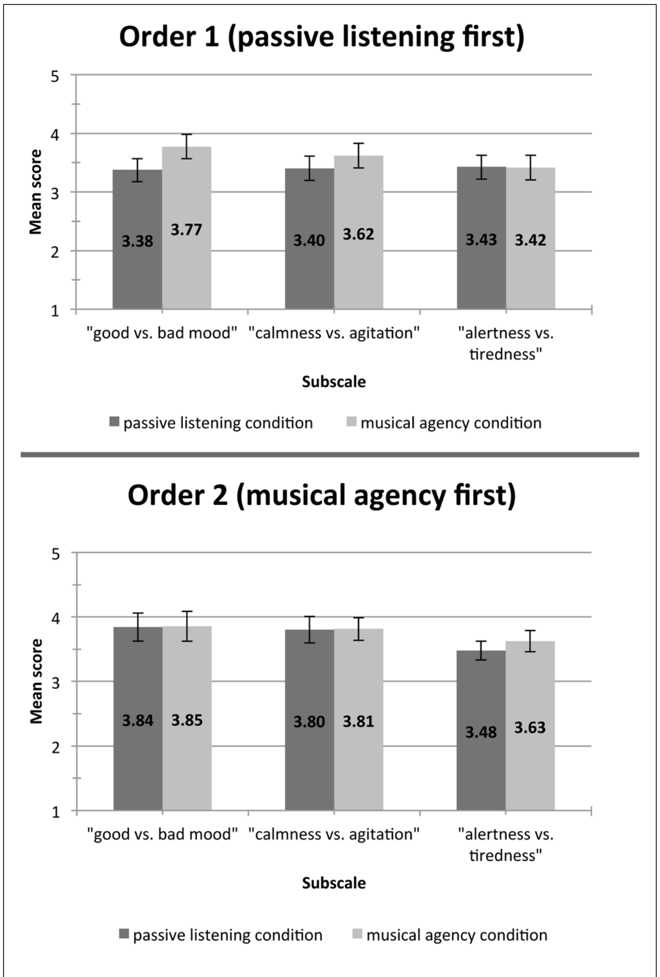


FIGURE 3 | The figure depicts mean ratings for the "good vs. bad mood" subscale by condition and condition order; error bars show the standard error.

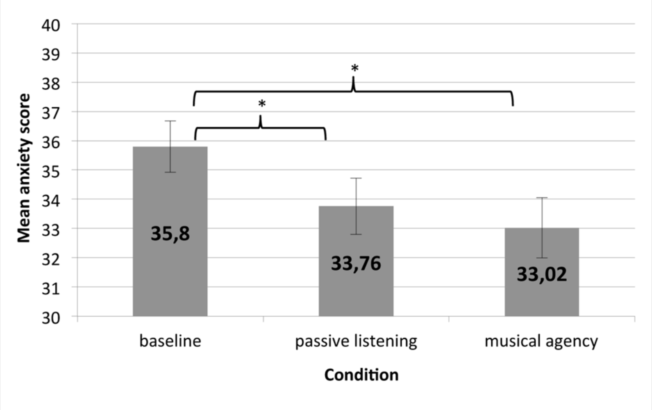


FIGURE 4 | The figure depicts mean anxiety scores on the State Trait Anxiety Inventory by condition; error bars show the standard error and *indicates a significant effect with a *p*-value of <0.05.

of physical exertion (and corresponding endorphin release) does not alone modulate mood in the participants. However, it cannot be excluded that mood is influenced by physiological differences between the conditions related to the different (less stereotype) movement patterns during the active condition (Fritz et al., 2013).

The mechanism by which mood may be enhanced by the musical agency condition is not yet completely apparent, but it is probably related to hormonal processes, which are known to have the capacity to influence mood (Zubieta et al., 2003). The time period between conditions (10 min) is not sufficient for hormones like cortisol and endorphins to return to a baseline level and the observed lingering effect of the musical agency condition (effect of order; **Figure 3**) is likely to be due to such a longer term hormonal effect. The agency condition more strongly creates emotional experience in the performer. In dimensional emotion models physiological arousal is almost always a basic dimension (Fritz, 2009). If such a physiological arousal (systematically created by workout) in the performer is combined with expressional behavior as during the musical agency condition, it may facilitate emotional experience in the performer. This would be further facilitated by the fact that to a great degree musical expression in Western music comprises emotional expression (Fritz et al., 2009). The physiological exercise carried out by the participants did not result in significant differences in the dimensions "calmness vs. agitation" and "alertness vs. tiredness" (two of the subscales of the Multidimensional Mood Questionnaire), but in the "good vs. bad mood" dimension it did. This further underlines the capacity that active music making has an intensifying (ameliorating) emotional effect on the performer.

How may such emotional experience be facilitated during the active condition? Obviously, the musical agency condition is cognitively more demanding than the passive listening condition. It seems to be this additional cognitive challenge in terms of personal expression that brings about the observed mood changes. This observed benefit from higher expressional cognitive demand during the active condition, where participants have control over the music and interact demonstrates that human beings gain from expressing inner thought processes and being creative. This would also indicate why humans are devoted to being social animals. Furthermore it may entail that jymmin may be a convenient method for influencing an asocial mindset.

The finding that anxiety does not differ between the active and passive condition may relate to previous findings showing that making music while modulating positive affect did not influence negative affect (Dunbar et al., 2012) as assessed with the PANAS (Watson et al., 1988). On the other hand, it may be that musical agency does have a capacity to more strongly influence anxiety, but that given the relatively low anxiety of participants to begin with, there is a ceiling effect. It would thus be interesting to repeat the experiment with a high anxiety cohort.

While the exercise movements were guided as a consequence of the design of the fitness machines, they were also less stereotypical than usual during exercise machine workout. This raises the question if it can be regarded a more or less healthy movement than those usually performed on exercise machines. Such are stereotypical movements, which are repeated for a defined number of times (often 10–12) and which are evenly performed from beginning to end, and consist of almost exclusively isotonic movement. In contrast, the current musical feedback technology (jymmin) rather encouraged less stereotypical movement patterns with a greater proportion of isometric movement, for example when participants held the weights at a certain position and someone else tried to play a "solo."

The idea that movements on exercise machines are most effective when evenly performed from beginning to end and for a defined number of times is largely a historical product, and not proven by physiological data (Hartmann et al., 2010). Given the greater proportion of isometric movements and the smaller stereotypicity, the current movement pattern may be regarded to more closely resemble climbing than exercise machine movement. Moreover, climbing has been argued to be quite healthy (Heitkamp et al., 2005), among other reasons because it is non-stereotypical and accordingly puts less strain on joints than stereotypical movements such as jogging and exercise machine workout.

On a critical note, the musical excerpts listened to in the "passive listening" condition were (similar to those created the in "musical agency" condition) relatively basic musical compositions (electronic music) arranged by non-musicians. Thus it may well be that participants deem other commercially available songs more pleasant to listen to. This is also relevant to the current findings, because it cannot be excluded that the agency in the "musical agency" condition only rendered the listening to the soundscapes less unpleasant than passively listening to them in the "passive listening" condition (and that workout with a neutral soundscape or one's own favorite music may have more strongly increased the mood than during the musical agency condition). Note however that after the experiment, in a post-experimental interview, participants were asked how they generally felt during the Experiment, and how they liked the music. It was obvious that the large majority of the participants enjoyed the music (also the passive listening part). However, this parameter was unfortunately not quantitatively assessed, so that the above argument has to be regarded a limitation of the current study, and should be considered in future studies. Furthermore, for future experiments it would be helpful to qualitatively assess the experience of the participants in the "musical agency" condition in order to deduce ideas about psychological factors underlying the observed mood effect.

INDIVIDUAL FACTORS THAT MAY MODULATE THE EFFECT OF JYMMIN ON MOOD

At this point the mechanism by which jymmin has an effect on mood must remain hypothetical. However, there is evidence that during music making, in comparison to passive music listening, a number of body-physiology related parameters are different due to emotion-related autonomic nerve activity (Nakahara et al., 2011), and a greater degree of flow state experience (Wrigley and Emmerson, 2011). It will have to be specified what it is about musical agency that creates the observed psychological effect on mood. We propose that a strong experience of individual expression through music may engage a social communication program that can strongly modulate perception, especially when related to proprioception. Note that the nature and function of such communicative musicality (Malloch and Trevarthen, 2008) has been a focus of recent music research. A manipulation of physiological arousal by music-making-associated workout may enhance the intensity of this experience. The communicative aspect of the musical agency condition is further underscored by the fact that alternating soloing was encouraged by the design of the musical interaction (see above).

Note that it is to be expected that the modulating influence of music making on mood is also influenced by several other factors in the individual that would need to be addressed in future studies: (1) There may be interpersonal factors, such that some performers find each other more (or less) attractive, or friendly. (2) Participants may have in their dynamic interaction a better (or worse) "chemistry" together, which means they may respond more or less sensitively to each other's actions. (3) The influence on their mood may be modulated by their current mental state for example their daily mood or hormonal disposition (Hunter et al., 2011). (4) The effect on their mood may relate to their openness to experience such a form of rather ecstatic togetherness with others. (5) It may vary depending on how much individuals generally derive enjoyment from musical expression and interaction. (6) It may depend on the musical style of the musical feedback participants produce (if for example they considered it part of their repertoire of favorite music) (Dyrlund and Wininger, 2007; Istók et al., 2013). (7) Finally, the combination of music making and high bodily arousal as conducted in the jymmin music feedback technology also gives rise to the possibility that some individuals derive more positive effect on their mood through physiological experience or disposition. For example, that they can experience greater or lesser reward through a disposition of their dopaminergic system. This would, for example, also relate to patients who suffer from a disorder of the dopaminergic system, as in Parkinson's disease, or in depression.

In conclusion, we present a method that combines exercise machine workout and music making, and by this measure makes exercise machine workout more desirable through an influence on experienced mood. A key to its impact on the individual (and thus its potential as a recreational activity and therapeutical approach) may be that movements motivated by social interaction for a common esthetic goal in a different way engage motor control than the stereotypical conventional exercise machine movements. It is discussed which factors in the individual may modulate the effect of this so-called "jymmin." Furthermore, we outline putative therapeutical benefits of this method.

REFERENCES

De Bruin, N., Doan, J. B., Turnbull, G., Suchowersky, O., Bonfield, S., Hu, B., et al. (2010). Walking with music is a safe and viable tool for gait training in Parkinson's disease: The effect of a 13-week feasibility study on single and dual task walking. *Parkinson's Dis.* 2010, 483530. doi: 10.4061/2010/483530

Dunbar, R., Kaskatis, K., MacDonald I., and Barra, V. (2012). Performance of music elevates pain threshold and positive affect: implications for the evolutionary function of music. *Evol. Psychol.* 10, 688–702.

Dyrlund, A. K., and Wininger, S. R. (2007). The effects of music preference and exercise intensity on psychological variables. *J. Music Ther.* 45, 114–134.

Enzensberger, W., Oberländer, U., and Stecker K. (1997). Metronome therapy in patients with Parkinson disease. *Der Nervenarzt* 68, 972–977. doi: 10.1007/s001150050225

Fritz, T. (2009). *Emotion Investigated With Music of Variable Valence: neurophysiology and Cultural Influence.* Leipzig: Max-Planck-Institute for Human Cognitive and Brain Sciences.

Fritz, T., Jentschke, S., Gosselin, N., Sammler, D., Peretz, I., Turner, R., et al. (2009). Universal recognition of three basic emotions in music. *Curr. Biol.* 19, 573–576. doi: 10.1016/j.cub.2009.02.058

Fritz, T. H., Hardikar, S., Demoucron, M., Niessen, M., Demey, M., Giot, O., et al. (2013). Musical agency reduces perceived exertion during strenuous physical performance. *Proc. Natl. Acad. Sci. U.S.A.* 110, 17784–17789. doi: 10.1073/pnas.1217252110

Hartmann, U., Platen, P., Niessen, M., Mank, D., Marzin, T., Bartmus, U., et al. (2010). *Krafttraining im Nachwuchsleistungssport [Workout training in junior high-performance sport]*. Bonn: Bundesinstitut für Sportwissenschaft.

Heinrichs, M., and Nater, U. (2002). Der Mehrdimensionale Befindlichkeitsfragebogen (MDBF). [The Multidimensional Mood Questionnaire (MMQ)]. *Z. Klin. Psychol. Psychother. Psychopathol.* 31, 66–67. doi: 10.1026//1616-3443.31.1.66

Heitkamp, H., Wörner, C., and Horstmann, T. (2005). Klettertraining bei Jugendlichen: Erfolge für die wirbelsäulenstabilisierende Muskulatur. [Climbing training in teenagers: A success for the muscular system supporting the spinal column] *Sportverletz. Sportschaden* 19, 28–32. doi: 10.1055/s-2005-857953

Hunter, P. G., Schellenberg, E. G., and Griffith, A. T. (2011). Misery loves company: mood-congruent emotional responding to music. *Emotion* 11, 1068. doi: 10.1037/a0023749

Istók, E., Brattico, E., Jacobsen, T., Ritter, A., and Tervaniemi, M. (2013). 'I love rock 'n' roll' – music genre preference modulates brain responses to music. *Biol. Psychol.* 92, 142–151. doi: 10.1016/j.biopsycho.2012.11.005

Karageorghis, C. I., Terry, P. C., and Lane, A. M. (1999). Development and initial validation of an instrument to assess the motivational qualities of music in exercise and sport: the Brunel music rating inventory. *J. Sports Sci.* 17, 713–724. doi: 10.1080/026404199365579

Kattenstroth, J.-C., Kolankowska, I., Kalisch T ., and Dinse, H. R., (2010). Superior sensory, motor, and cognitive performance in elderly individuals with multi-year dancing activities. *Front. Aging Neurosci.* 2:31. doi: 10.3389/fnagi.2010.00031

Kibler, V. E., and Rider, M. S. (1983). Effects of progressive muscle relaxation and music on stress as measured by finger temperature response. *J. Clin. Psychol.* 39, 213–215. doi: 10.1002/1097-4679(198303)39:2<213::AID-JCLP2270390211>3.0.CO;2-2

Kirschner, S., and Tomasello, M. (2009). Joint drumming: social context facilitates synchronization in preschool children. *J. Exp. Child Psychol.* 102, 299–314. doi: 10.1016/j.jecp.2008.07.005

Kirschner, S., and Tomasello, M. (2010). Joint music making promotes prosocial behavior in 4-year-old children. *Evol. Hum. Behav.* 31, 354–364. doi: 10.1016/j.evolhumbehav.2010.04.004

Kreutz, G., Bongard, S., Rohrmann, S., Hodapp, V., and Grebe, D. (2004). Effects of choir singing or listening on secretory immunoglobulin A, cortisol, and emotional state. *J. Behav. Med.* 27, 623–635. doi: 10.1007/s10865-004-0006-9

Kuhn, D. (2002). The effects of active and passive participation in musical activity on the immune system as measured by salivary immunoglobulin A (SIgA). *J. Music Ther.* 39, 30.

Laux, L., Glanzmann, P., Schaffner, P., and Spielberger, C. D. (1981). *Das State-Trait-Angstinventar [The State-Trait Anxiety Inventory]*. Beltz: Weinheim.

Leman, M., Moelants, D,. Varewyck, M., Styns, F., van Noorden, L., Martens, J. P., et al. (2013). Activating and relaxing music entrains the speed of beat synchronized walking. *PLoS ONE* 8:e67932. doi: 10.1371/journal.pone.0067932

Lim, H. B. T., Atkinson, G., Karageorghis, C. I., and Eubank M. R., (2009). Effects of differentiated music on cycling time trial. *Int. J. Sports Med.* 30, 435–442. doi: 10.1055/s-0028-1112140

Malloch, S., and Trevarthen, C. (2008). "Musicality: Communicating the vitality and interests of life," in *Communicative Musicality*, eds S. Malloch and C. Trevarthen. (Oxford: Oxford University Press), 1–11.

McIntosh, G. C., Brown, S. H., and Jalali, B. (1997). Rhythmic auditory-motor facilitation of gait patterns in patients with Parkinson's disease. *J. Neurol. Neurosurg. Psychiatry* 62, 22–26. doi: 10.1136/jnnp.62.1.22

Motl, R. W., McAuley, E., Snook, E. M., and Gliottoni, R. C. (2009). Physical activity and quality of life in multiple sclerosis: intermediary roles of disability, fatigue, mood, pain, self-efficacy and social support. *Psychol. Health Med.* 14, 111–124. doi: 10.1080/13548500802241902

Nakahara, H., Furuya, S., Masuko, T., Francis, P. R., and Kinoshita, H. (2011). Performing music can induce greater modulation of emotion-related psychophysiological responses than listening to music. *Int. J. Psychophysiol.* 81, 152–158. doi: 10.1016/j.ijpsycho.2011.06.003

Ruby, M. B., Dunn, E. W., Perrino, A., Gillis, R., and Viel, S. (2011). The invisible benefits of exercise. *Health Psychol.* 30, 67–74. doi: 10.1037/a0021859

Simpson, S. D., and Karageorghis, C. I. (2006). The effects of synchronous music on 400-m sprint performance. *J. Sports Sci.* 24, 1095–1102. doi: 10.1080/02640410500432789

Steyer, R., Schwenkmezger, P., Notz, P., and Eid, M. (1997). *Der Mehrdimensionale Befindlichkeitsfragebogen* [MDBF; The Multidimensional Mood State Questionnaire, MDMQ]. Hogrefe: Handanweisung Göttingen.

Ströhle, A. (2009). Physical activity, exercise, depression and anxiety disorders. *J. Neural. Transm.* 116, 777–784. doi: 10.1007/s00702-008-0092-x

Styns, F., van Noorden, L., Moelants, D., and Leman, M. (2007). Walking on music. *Hum. Mov. Sci.* 26, 769–785. doi: 10.1016/j.humov.2007.07.007

Szmedra, L., and Bacharach, D. (1998). Effect of music on perceived exertion, plasma lactate, norepinephrine and cardiovascular hemodynamics during treadmill running. *Int. J. Sports Med.* 19, 32–37. doi: 10.1055/s-2007-971876

Terry, P. C., Karageorghis, C. I., and Saha, A. M., and D'Auria, S. (2012). Effects of synchronous music on treadmill running among elite triathletes. *J. Sci. Med. Sport* 15, 52–57. doi: 10.1016/j.jsams.2011.06.003

Thaut, M., McIntosh, G., Rice R. R., Miller, R. A., Rathbun, J., and Brault, J. M. (1996). Rhythmic auditory stimulation in gait training for Parkinson's disease patients. *Mov. Disord.* 11, 193–200. doi: 10.1002/mds.870110213

van der Vlist, B., Bartneck, C., and Mäueler, S. (2011). moBeat: Using interactive music to guide and motivate users during aerobic exercising. *Appl. Psychophysiol. Biofeedback* 36, 135–145. doi: 10.1007/s10484-011-9149-y

Van Dyck, E., Moelants, D., Demey, M., Deweppe, A., Coussement, P., Leman, M., et al. (2013). The impact of the bass drum on human dance movement. *Music Percep.* 30, 349–359. doi: 10.1525/mp.2013.30.4.349

Vendrig, A. A., and Lousberg, R. (1997). Within-person relationships among pain intensity, mood and physical activity in chronic pain: a naturalistic approach. *Pain* 73, 71–76. doi: 10.1016/S0304-3959(97)00075-4

Watson, D., Clark, L. A., and Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: the PANAS scales. *J. Pers. Soc. Psychol.* 54, 1063–1070. doi: 10.1037/0022-3514.54.6.1063

Wrigley, W. J., and Emmerson, S. B. (2011). The experience of the flow state in live music performance. *Psychol. Music* 41, 292–305. doi: 10.1177/03057356114 25903

Wu, D. W., Chapman, C. S., Walker, E., Bischof, W. F., and Kingstone, A. (2013). Isolating the Perceptual From the Social: Tapping in Shared Space Results in Improved Synchrony. *J. Exp. Psychol. Hum. Percept. Perform.* 39, 1218–1223. doi: 10.1037/a0033233

Zubieta, J.-K., Ketter, T. A., Bueller, J. A., Xu, Y., Kilbourn, M. R., Young, E. A., et al. (2003). Regulation of human affective responses by anterior cingulate and limbic {micro}-opioid neurotransmission. *Arch. Gen. Psychiatry*, 60, 1145. doi: 10.1001/archpsyc.60.11.1145

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 24 May 2013; accepted: 20 November 2013; published online: 10 December 2013.

*Citation: Fritz TH, Halfpaap J, Grahl S, Kirkland A and Villringer A (2013) Musical feedback during exercise machine workout enhances mood. Front. Psychol. **4**:921. doi: 10.3389/fpsyg.2013.00921*

This article was submitted to Cognitive Science, a section of the journal Frontiers in Psychology.