frontiers | Frontiers in Communication

TYPE Original Research
PUBLISHED 04 October 2022
DOI 10.3389/fcomm.2022.957987

OPEN ACCESS

EDITED BY
Sandra Petroni,
University of Rome Tor Vergata, Italy

REVIEWED BY
Mihailo Antovic,
University of Niš, Serbia
Dusan Stamenkovic,
Södertörn University, Sweden

*CORRESPONDENCE
Valentijn Prové
valentijn.prove@kuleuven.be

SPECIALTY SECTION
This article was submitted to Multimodality of Communication, a section of the journal Frontiers in Communication

RECEIVED 31 May 2022
ACCEPTED 07 September 2022
PUBLISHED 04 October 2022

CITATION
Prové V (2022) Measuring embodied conceptualizations of pitch in singing performances: Insights from an OpenPose study.
Front. Commun. 7:957987.
doi: 10.3389/fcomm.2022.957987

Measuring embodied conceptualizations of pitch in singing performances: Insights from an OpenPose study

Valentijn Prové*

Department of Linguistics, KU Leuven, Leuven, Belgium

People conceptualize auditory pitch as vertical space: low and high pitch correspond to low and high space, respectively. The strength of this cross-modal correspondence, however, seems to vary across different cultural contexts and a debate on the different factors underlying this variation is currently taking place. According to one hypothesis, pitch mappings are semantically mediated. For instance, the use of conventional metaphors such as "falling" or "rising" melodies strengthens a pitch-height mapping to the detriment of other possible mappings (e.g., pitch as bright/dark color or small/big size). Hence, entrenched pitch terms shape specific conceptualizations. The deterministic role of language is called into question by the hypothesis that different pitch mappings share a less constraining conceptual basis. As such, conceptual primitives may be concretized *ad hoc* into specific domains so that more local variation is possible. This claim is supported, for instance, by the finding that musicians use language-congruent (conventional) and language-incongruent (*ad hoc*) mappings interchangeably. The present paper substantiates this observation by investigating the head movements of musically trained and untrained speakers of Dutch in a melody reproduction task, as embodied instantiations of a vertical conceptualization of pitch. The OpenPose algorithm was used to track the movement trajectories in detail. The results show that untrained participants systematically made language-congruent movements, while trained participants showed more diverse behaviors, including language-incongruent movements. The difference between the two groups could not be attributed to the level of accuracy in the singing performances. In sum, this study argues for a joint consideration of more entrenched (e.g., linguistic metaphors) and more context-dependent (e.g., musical training and task) factors in accounting for variability in pitch representations.

KEYWORDS

embodiment, singing, head movement, cross-modal correspondences, pitch

Introduction

As part of the present volume on kinesemiotics, this paper subscribes to a particular view on movement-based communication according to which bodies are always situated both in physical and cultural space (Maiorani, 2021). In physical space,

bodily movement follows dichotomic schemes that are essential and universal, such as upward/downward or forward/backward. However, the potential of these schemes to be used for meaning-making relies on contextual values (Maiorani, 2021, 27). Hence, meaningful movements are choices that are made relative to other acts in a physical repertoire and in a cultural context. In ballet performances, for instance, the cultural space is structured locally by the audience and the stage, which is divided into different portions by setting and lighting (Maiorani, 2021, 26). As such, dancers map a physical choice (e.g., making a forward movement in space) onto a cultural choice (e.g., addressing the audience as part of the space). The dual semiosis sketched here applies to different communicative processes and I believe it offers a refreshing perspective on the flexibility of meaning across socio-cultural contexts. In this paper, I will address the Research Topic of cross-cultural conceptualizations of auditory pitch. Humans make sense of pitch frequency by mapping it onto scalar, physical qualities such as high/low space, bright/dark color or small/big size. Since these mappings require different sensory modalities to be integrated, the term cross-modal correspondences is commonly used to refer to this phenomenon (Parise, 2015). The question as to how these arguably universal correspondences structure pitch conceptualizations in different cultural contexts has led to a vivid debate.

People naturally seem to associate auditory pitch with qualities pertaining to different sensory domains such as brightness, angularity, size, and height (cf. Spence, 2011; Walker, 2016 and Eitan, 2017 for reviews). Cross-modal correspondences form a relatively old yet very popular Research Topic because of their intuitive nature and their potential for application (Parise, 2015). Many studies have focused on how different pitch correspondences serve as a basis for communication. From a linguistic perspective, as a matter of fact, the multitude of metaphors lexicalizing pitch relationships is striking. To illustrate, tones are referred to as "thin" and "thick" in Turkish (Dolscheid et al., 2020) or "tight" and "loose" by the Kreung people Cambodja (Parkinson et al., 2012), to cite only two examples. Crucially, studies influenced by Cognitive Metaphor Theory (Zbikowski, 2002; Ashley, 2004; Shayan et al., 2011; Dolscheid et al., 2013, 2020; Casasanto, 2017; Cox, 2017; Fernandez-Prieto et al., 2017; Holler et al., 2022) propose that the way pitch qualities are coded in a language, shapes the way people conceive of pitch and vice versa. As such, pitch mappings become hard-wired during development because of the linguistic system and its conventional nature. A key finding challenging this claim is that Western participants are also consistent in applying unfamiliar metaphors for pitch (Eitan and Timmers, 2010). In a similar fashion, people can rely on higher-order schemes in making sense of visual pitch representations, for instance when the directionality on a vertical pitch axis is reversed (Antović et al., 2020). As such, people do not necessarily have a pre-existing percept of pitch as a spatial analog. This finding is supported by research on the SMARC effect, which suggests that the pitch-height mapping occurs at an early stage of processing as due to a generalized magnitude representation (Rusconi et al., 2006; Lidji et al., 2007; Prpic and Domijan, 2018). In an experiment carried out by Pitteri et al. (2021), reaction times for congruent pitch mappings were even faster if the pitch-height and the pitch-brightness mapping were combined. Therefore, the deterministic role of the linguistic system should be questioned.

A further non-linguistic factor that reveals the ambivalence of semantic strengthening effects is musical training. On the one hand, musicians have been shown to be more consistent than non-musicians in their gestural depictions of dynamic pitch contours in a communication task, demonstrating a more entrenched vertical conceptualization (Küssner et al., 2014). Indeed, musicians are more familiar with terms and techniques that embody a spatial pitch metaphor such as staff notation or specific instruments. To illustrate, Timmers and Li (2016) showed that pianistic expertise strengthens a lateral pitch-space mapping. It should be noted, however, that the experimental methods used in these studies, a communication task and a forced choice paradigm, respectively, prompt participants to react in an efficient way. This means that, if there is a convention available, it is likely to be used in these contexts. However, in a series of experiments involving more subjective pitch mappings in the domain of tonality (Maimon et al., 2021), musicians did not rely more on conventional metaphors than non-musicians did. Moreover, a qualitative study on the use of pitch metaphors in lyric singing classes (Prové and Feyaerts, 2022) showed that singing teachers blend conventional and unconventional conceptualizations in one multimodal expression. In one example, the teacher depicts how the student should sing a rising melody with a high note that is difficult to produce, by bending the knees, pointing downwards along her legs as the melody rises (non-conventional), and subsequently pointing upwards so as to indicate where the highest note should be (conventional). Hence, although musicians can be argued to have most contact with conventional pitch-height metaphors in Western cultures, they can also be characterized as more flexibly using both entrenched and *ad hoc* pitch mappings in order to adapt to different contexts. Such a reasoning makes it harder to assign a dominant role to semantic mediation.

In this paper, I am supporting this argument by hypothesizing that Western non-musicians (native speakers of Dutch, which is a pitch-height language) make more consistent language-congruent vertical head movements compared to musically trained participants while reproducing rising and falling melodies in a singing task. Whereas, it is likely that both groups of participants will react to musical tasks by making vertical head movements (Wöllner and Jensenius, 2017; Swarbrick et al., 2019; González Sánchez et al., 2020; Zelechowska et al., 2020), the directionality of the head movements may rely on different effects. On the one hand,

language-congruent movements (i.e., downward movement while singing falling pitch contours and vice versa) could be expected on the basis of the embodied simulation hypothesis (Barsalou, 1999; Casasanto and Gijssels, 2015; Cuccio and Fontana, 2017; Hostetter and Alibali, 2019) because they would be a physical manifestation of the more entrenched pitch-height mapping. On the other hand, the gesture-vocal coupling might disturb optimal body tension in singing performances (cf. Pearson and Pouw, 2022 for a recent discussion) and lead to reduced or even reversed (language-incongruent) movements For instance, stretching the body creates a tension in the vocal apparatus which is, paradoxically, suboptimal for singing high-pitched tones (cf. Turner and Kenny, 2011 for a review on the relation between posture and performance). As such, I hypothesize musicians to adapt their embodied behavior to the context of singing more than non-musicians do. Another possibility is that, independent from the participants' musical training, better singing performances correlate less with congruent head movements. This is a case in point for an *ad hoc* conceptualization of pitch: the musically trained (or better skilled) participants, albeit more familiar with the metaphorical pitch conventions, are expected to use the pitch-height mapping less in the specific singing task. The design of the empirical study is based on an unpublished paper (Baptist, 2014).

Based on the empirical study outlined above, I argue for a joint consideration of entrenched and *ad hoc* factors in the conceptualization of pitch. This reconciliatory position can also be motivated by recent developments in metaphor theory. In the last decade, metaphors have been increasingly studied as dynamic analogies that are made relevant in human interactions to different degrees (Müller and Tag, 2010; Kolter et al., 2012; Zlatev and Devylder, 2020). To illustrate, ballet teachers elaborate metaphors verbally and gesturally so that an analogy between two different sensory feelings can be interactionally negotiated (Müller and Ladewig, 2013). Although it should be made clear that analogies are a more complex phenomenon pertaining to the domain of communication, they may be indicative of the natural flexibility that cross-domain mappings offer. As Walker (2016) argues, mappings are transitive. For instance, if high is bright, and bright is thin, then high will be thin (Walker, 2016, 107). Crucially, some mappings have an ambiguous relation to magnitude: if high is "more" and thin is "less," high pitch is both "less" and "more" (Eitan and Timmers, 2010: 420). Hence, if pitch conceptualization involves the percept of an axis, it can be determined *ad hoc* which end is "more." Moreover, while increasing and decreasing qualities can be associated in synesthesia, different movement axes are compatible with these magnitude representations. The example *par excellence* of flexibility in activating cross-modal correspondences may concern orchestra conductors, who use a relatively limited collection of gestures to depict complex and interrelated properties of sound such as loudness, timbre and tempo (Globerson et al., 2021). As a result, one single property may be depicted on all three vertical, horizontal and sagittal movement axes (Meissl et al., 2022).

Materials and methods

In order to test the hypothesis outlined above, I designed an experiment that required participants to reproduce melodies while being filmed. I obtained written consent from 38 native speakers of Dutch (age range 18–25) to record their data for this study. The choice of participants was forced in that (a) they were not allowed to be trained in singing and (b) I created a group with no musical training at all (i.e., unable to read staff notation, $n = 19$, "untrained") and a musically trained group (i.e., able to read staff notation, $= 19$, "trained") to create a between-subject variable for *training*. The latter group had been participating in music trainings for 7.35 years on average (SD $= 4.18$). All participants were students at the author's institution and they received a cinema ticket as a reimbursement for their participation. One participant (group "untrained") was excluded from the analysis because of data loss during the recording.

As for the experimental procedure, I copied the task and the musical stimuli from an unpublished pilot study (Baptist, 2014). I asked the participants to stand on a marked spot on the ground while being filmed both from a frontal perspective and in profile using two camcorders (Sony HDR-CX160, 720x576, 25 fps). They would hear a melody being played twice in a row from a speaker behind them, which they would reproduce using the vocalization "la." It should be noted that the "a" vocal is produced relatively low in the mouth cavity, which might have interfered with the head movements. I will address this potential confound again in the discussion. I rehearsed the procedure two times to allow the participants to get acquainted with the task. During this try-out, they were allowed to ask questions. Subsequently, I played seven more melodies in a random order, which constituted the actual experiment. After the experiment, the participants received an explanation about the objectives of this study and they could withdraw their data if they wished (which did not happen).

For this procedure, I used three types of melodies with two or three different difficulty levels per type because it was hard to predict the singing skills of the participants. All the stimuli and their notations can be found in Supplementary materials. The falling melodies ("f1" and "f2") were designed in such a way that there was a static phase (three times the same tone) and a linear dynamic phase (a scale to the octave below). The rising melodies ("r1," "r2" and "r3") were constructed in exactly the same way, using the octave above in reference to the start tone. Finally, the interval melodies ("i1" and "i2") were rising too, but I used interval of a fifth and the range was one octave and a fifth. Interval melodies were considered to be more difficult variants of the rising melodies because the intervals between the notes are more difficult to reproduce. To construct the variable "*melody type*" in the analysis, one melody per type was selected

("r2", "i2" and "f2"), according to the hypothesis that lower notes in the falling melodies and higher notes in the rising melodies would elicit more prominent head movement. Melody "r1" was excluded because some participants reported to feel uncomfortable to sing it so "r2" was selected as the highest version. Melody "i2" was selected as its counterpart with larger intervals because the end note is the same. Finally, melody "f2" was selected as the lowest version of the falling melody type. I composed the musical stimuli using the Musescore[1] software and I exported the sound with a built-in piano sample.

In order to asses vertical head movements, I used the videos that showed the participants in profile and I automatically estimated the two-dimensional position of their nose in pixels (px) using OpenPose (Cao et al., 2017), which is an open-source algorithm for video-based body part tracking. I adjusted incorrect data points by applying a smoothing filter[2]. Concerning auditory pitch, I used Praat (Boersma and Weeninck, 2022) to extract the pitch contours of both the stimuli and the participants' sound production in Herz (Hz), based on the audio that was recorded by the frontal camera. I resampled the pitch data according to the frame rate of the videos (25 fps) and the same smoothing filter was applied to exclude incorrect data points. As a next step, I used the ELAN-software (Wittenburg et al., 2006) to segment the videos into action units where the participants were singing the melodies. I used a visualization of the audio waveform to manually determine the beginning and the end point of the actions. As for the following steps, I always used the time stamps of these actions to process the data and construct the variables in R Markdown (Xie et al., 2018)[3]. For each action unit, I z-scaled both the movement and pitch data (i.e., in terms of the distance of each data point from the mean in standard deviation units) and I centered the data relative to the first data point at the beginning of the participants' sound production, that is the first data point is subtracted from itself and every subsequent observation. This means that the starting point of every movement trajectory or pitch contour was a zero. Moreover, the standardized scores made the trajectories comparable across the participants, eliminating individual differences between body sizes and positions.

On the basis of the data described above, I calculated two measures. First, the slope of vertical movement measures how strongly the movement is directed on the vertical axis (hereafter: *vertical directionality*). This variable was constructed by calculating the correlation coefficient of the vertical movement data in time. A value of zero entails no directionality, a −1 entails a perfect downward relationship and a 1 entails a perfect upward relationship. Second, *singing accuracy* was calculated using Dynamic Time Warping (DTW, Pouw and Dixon, 2020), which rendered a quantification of the similarity between the standardized pitch contours of the stimulus and the participants' sound production. A zero entails perfect accuracy, because the closer the value is to zero, the less one time series has to be stretched and truncated to match the other, that is the more aligned the pitch contours are. I emphasize that singing accuracy was measured in terms of pitch only.

In what follows, I present a regression analysis with the variables *singing accuracy* (response variable or covariate), *vertical directionality* (response variable), *training* (independent variable, between-subject) *and melody type* (independent variable, within-subject) outlined in this section. I built nested linear regression models using the "lme4" package in R (Bates et al., 2015). I always started with a null model containing a random intercept for the individual participants and I gradually added the *training* and *melody type* factors and their two-way interaction. In order to determine the best model and its significance, I used the Likelihood Ratio test as prescribed in Winter (2013).

Results

Before carrying out the statistical analyses, I visually explored the contours of the vertical head movement trajectories by plotting their temporal unfolding. From a visual inspection of the trajectories per melody (as illustrated in Figure 1), it is clear that the participants behaved in different individual ways. Some movements feature clear upward or downward movement, whereas others have more variable contours. Therefore, it should be noted again that directionality refers to a general trend in the movement trajectory.

As for *singing accuracy* (M = 0.33, SD = 0.21), the scores ranged from almost perfectly matching (0.06) to highly dissimilar (0.79) pitch contours. The group of trained participants did not perform better, increasing the accuracy by only 0.05 units (SE = 0.03, t-score = −1.64) compared to the untrained group that could not [$\chi^2_{(1)} = 2.66$, p = ns]. Adding *melody type* significantly improved the null model [$\chi^2_{(2)} = 96.64$, $p < 0.001$], with interval melodies decreasing the quality by 0.15 units (SE = 0.02, t-score = 6.68) and rising melodies decreasing the quality by 0.29 units (SE = 0.02, t-score = 12.99). There was no significant interaction effect between *training* and *melody type*, which implies that interval and rising melodies were reproduced less accurately compared to falling melodies in both groups.

The best regression model for *vertical directionality* involved both the *training* and the *melody type* factors and their two-way interaction effect [$\chi^2_{(2)} = 9.18$, $p < 0.05$]. Falling melodies systematically correlated with downward head movement in

1 This open-source software is freely downloadable from https://musescore.org/en.

2 I used a Kolmogorov–Zurbenko filter (window size = 5, number of iterations = 3).

3 The scripts and datasets created for this study can be accessed in the Zenodo repository (https://doi.org/10.5281/zenodo.7082438) under a Creative Commons 4.0 International license.

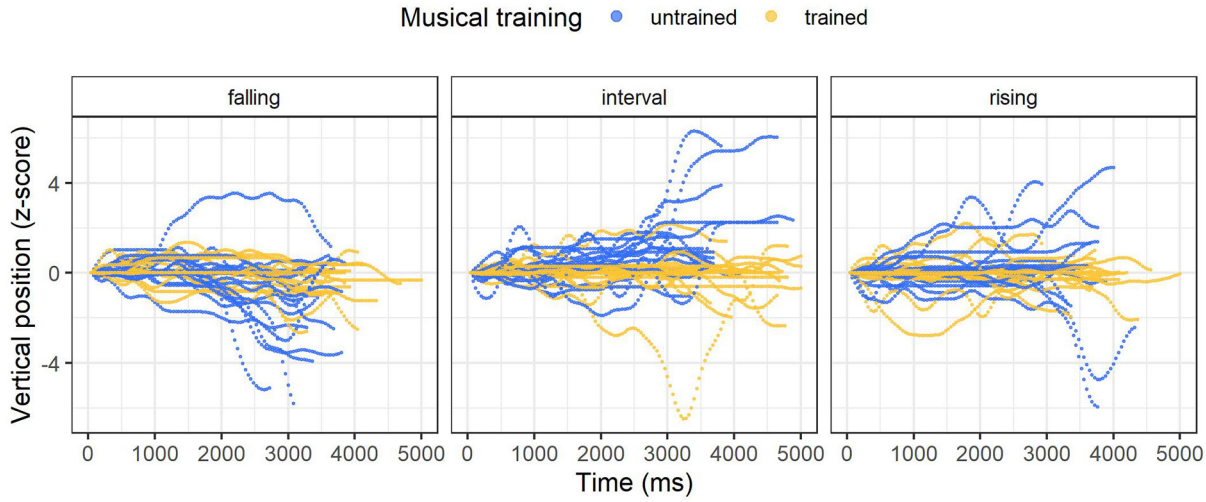

FIGURE 1
Plot of the vertical head movement trajectories per melody type. Participants who can and cannot read staff notation are marked yellow and blue, respectively.

both the untrained [model fit = −0.46, 95 % CI (−0.71, −0.20)] and the trained [model fit = −0.34, 95 % CI (−0.59, −0.09)] groups. By contrast, interval and rising melodies yielded different effects when comparing the two groups. In the untrained group, both the interval melodies [model fit = 0.55, 95 % CI (0.28, 0.82)] and linear rising melodies [model fit = 0.29, 95 % CI (0.04, 0.55)] correlated with upward head movements. Conversely, the trained group decreased the head movement slopes for both the interval melodies (by 0.69 units, SE = 0.22, t-score = −3.07) and the linear rising melodies (by 0.44 units, SE = 0.22, t-score = −1.99). As a consequence, the head movement slopes associated with these melody types were fitted close to zero [resp. −0.02, 95 % CI (−0.27, 0.23) and −0.03 95 % CI (−0.28, 0.22)]. In sum, this result implies that the musically trained participants made systematic downward movements when reproducing falling melodies, but no systematic upward or downward movements in the case of interval or linear rising melodies.

The plot in Figure 2 illustrates the fitted values for each factor level in *melody type* in both groups. The red bars represent the 95% confidence intervals and the dots represent the observed data points. The color of the dots is determined by the *singing accuracy* variable (green = perfect match, red = highly dissimilar) and the gray lines connect the observations from the individual participants. The violin boxes indicate the density of the observed data. Importantly, from a closer visual inspection of the violin plots, it is clear that the data points for interval and the rising melodies in the trained group are distributed along the entire axis without being skewed at a particular point. In relation to the fitted values that were close to zero (resp. −0.02 and −0.03, cf. previous paragraph), this entails that the null-effect is due to an approximately equal amount of melody-congruent and melody-incongruent movements. That is, the musically trained group should be characterized as showing a range of different behaviors, including more incongruent movements compared to the untrained group. Moreover, the reactions to the different types of melodies within the same participants, as illustrated by the gray lines that connect the observations in the plot, are less consistent in the trained group.

To conclude, adding *singing accuracy* as a covariate in the regression model predicting *vertical directionality* did not improve the model significantly [$\chi^2_{(3)} = 2.34$, p = ns]. Given that *singing accuracy* was not influenced by the training level (cf. *supra*), this result entails that it is the degree of musical training as a relatively broad factor encompassing different types of learning and skills and not the singing skill in itself that reduces congruent head movement when singing melodies.

Discussion

This paper has investigated vertical head movements during a scale reproduction task in musically trained and untrained speakers of Dutch. In doing so, it has explored a new setting to investigate contextual factors that may influence the *ad hoc* conceptualization of pitch as verticality. In line with Antović et al. (2020), it adds to the literature questioning the predominance of the semantic strengthening effect occurring, for instance, in Western societies that use linguistic expressions pertaining to the domain of vertical space such as “falling” and “rising” melodies.

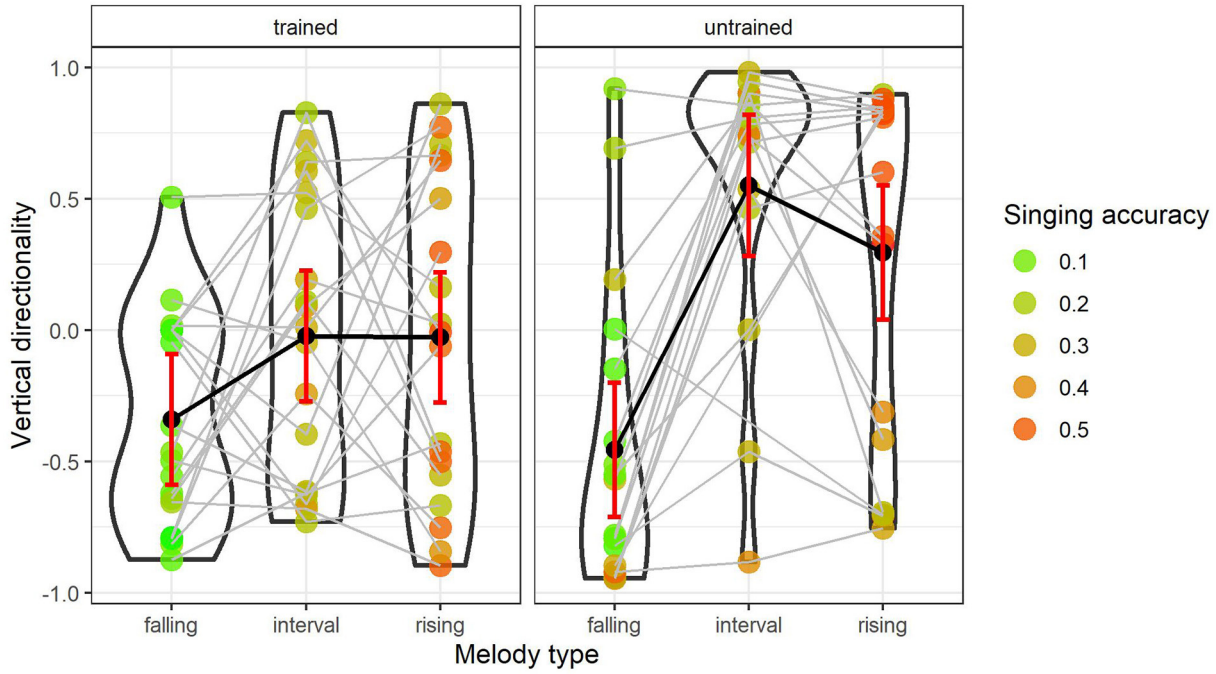

FIGURE 2
Estimations for vertical directionality. The red bars indicate 95% confidence intervals and the dots represent the individual data points (head movement slope per melody) per condition in each of the two groups. The colors of the dots represent the singing accuracy: green is more accurate and red is less accurate.

The semantic mediation hypothesis entails that our linguistic system shapes the way we structure the concept of pitch (Casasanto, 2017). As such, investigating pitch metaphors provides a window on mental pitch representations. To illustrate, the Farsi language offers a low codability for pitch (Holler et al., 2022). Whereas, a pitch-thickness metaphor is commonly preferred, other expressions involving verticality can also be used. Therefore, Farsi speakers have less consistent conceptualizations of pitch. Conversely, the Dutch language almost exclusively features a conventional metaphor for pitch-height so that the vertical conceptualization is more stable.

This claim can be nuanced by investigating the effect of musical training in a music-relevant setting. Musicians have been shown to be both more consistent in their use of conventional metaphors (Küssner et al., 2014) and in their flexible use of unfamiliar mappings (Eitan and Timmers, 2010; Maimon et al., 2021). Observing language-incongruent preferences in musicians demonstrates that a choice is made between different possible mappings that may or may not be supported by language. This trade-off effect is supported by evidence that pitch mappings are transitive (Walker, 2016) and that they rely on higher-order schemes such as generalized magnitude representations (Eitan and Timmers, 2010; Pitteri et al., 2021) or amodal conceptual primitives (Antović et al., 2020).

In order to lend support to the latter hypothesis, I conducted a behavioral experiment in which speakers of Dutch had to reproduce falling and rising scales and rising tone sequences with larger intervals using the vocalization "la." I hypothesized that a musically trained group (n = 19) would make less congruent head movements compared to a musically untrained group (n = 19). Musicians, although very familiar with Western musical conventions such as staff notation, are expected to be better at "escaping" the language-congruent vertical mapping that might have detrimental effects on the singing quality if gestures start to regulate suboptimal tension in the vocal apparatus (cf. Turner and Kenny, 2011). The directionality of the participants' head movements during their singing performances was computed using the vertical movement trajectories of the nose as tracked by the OpenPose (Cao et al., 2017) algorithm. The accuracy of their performance was assessed using Dynamic Time Warping (Pouw and Dixon, 2020), which yielded a measure of how similar the pitch contours of the stimuli and the participant's melody reproductions were.

As for the results, the musically untrained group of participants systematically made downward head movements when singing falling melodies and upward movements when singing rising melodies (both for the interval type and the linear scale type). Musically trained participants behaved in

a less equivocal way. While reproducing falling scales, they made systematic downward head movements as well, although this effect was less strong compared to the untrained group. Crucially, their movements aligning with both types of rising melodies were more diverse, including more incongruent movements. Hence, I find partial support for the claim that trained musicians adhere less to the more entrenched vertical pitch mapping of the Dutch language. This inconsistent behavior can be attributed to interconnected conceptualizations of auditory qualities that may interfere (cf. Parise, 2015 for a discussion). In contrast to static pitch in isolated tones, for instance, pitch sequences may trigger perceptions of loudness and dynamic progression (Eitan, 2013). Magnitude representations that underly pitch mapping are relevant for these dimensions as well. Moreover, the vocalization "la," which involves a vowel that is produced relatively low in the mouth cavity, might have elicited downward movement as well.

Moreover, I excluded the possibility that the effect of the musically trained and untrained groups of participants on pitch-congruent head movements was merely due to more accurate singing performances. I found two arguments to support this claim. On the one hand, the musically trained group did not perform better than the untrained group. Rather, falling melodies were reproduced more precisely compared to the interval and rising type in both groups. On the other hand, adding singing accuracy to the model predicting the directionality of head movements did not result in a significant effect. Hence, musical training was independent from singing skill and it was a much better predictor. On the whole, the results of the present empirical study offer support for the claim that cross-modal correspondences for auditory pitch share underlying associative processes allowing for contextual variation that cannot be explained by linguistic structure alone (Walker, 2016; Antović et al., 2020; Maimon et al., 2021).

The most important limitation of this study is that it remains unclear how musically trained and trained participants exactly differ from each other. Different skills may have proven useful in executing the experimental task: more experience with scales, more experience with live performance, a better development of imagery, a better familiarity with the discourse on suboptimal musical behavior, a transfer of their skill in playing musical instruments to singing etc. In this vein, the hypothesis that musician use more flexible and diverse mappings in their online conceptualization of pitch still needs more rigorous investigation. Future studies could also compare musicians and non-musicians from different linguistic groups to offer more conclusive results in the semantic-mediation debate.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://zenodo.org/record/7082438.

Ethics statement

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. The patients/participants provided their written informed consent to participate in this study. Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

Author contributions

The author confirms being the sole contributor of this work and has approved it for publication.

Funding

This research was supported by the Research Foundation - Flanders (FWO), grant number 1119321N.

Acknowledgments

I would like to thank Prof. Dr. Kurt Feyaerts, Katharina Meissl, and Clarissa de Vries wholeheartedly for their valuable comments on earlier versions of this paper and their enthusiasm about this research.

Conflict of interest

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.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fcomm.2022.957987/full#supplementary-material

References

Antović, M., Mitić, J., and Benecasa, N. (2020). Conceptual rather than perceptual: cross-modal binding of pitch sequencing is based on an underlying schematic structure. *Psychol. Music* 48, 84–104. doi: 10.1177/0305735618785242

Ashley, R. (2004). "Musical pitch space across modalities: Spatial and other mappings through language and culture," in *Proceedings of the 8th International Conference on Music Perception and Cognition*, eds S. D. Lipscomb, R. Ashley, R. O. Gjerdingen, and P. Webster (Adelaide: Causal Productions), 64–71.

Baptist, M. (2014). *De invloed van de Westerse conventionele verticaliteitsmetafoor voor toonhoogte op lichamelijk gedrag tijdens het zingen* (master's thesis). Leuven, Belgium: KU Leuven.

Barsalou, L. W. (1999). Perceptual symbol systems. *Behav. Brain Sci.* 22, 577–660. doi: 10.1017/S0140525X99002149

Bates D., Mächler M., Bolker B., and Walker S. (2015). Fitting linear mixed-effects models using lme4. *J. Statist. Sofw*. 67, 1–48. doi: 10.18637/jss.v067.i01

Boersma, P., and Weeninck, D. (2022). *Praat: Doing Phonetics by Computer*. Available online at: http://www.praat.org/ (accessed May 31 2022).

Cao, Z., Simon, T., Wei, S.-E., and Sheikh, Y. (2017). "Realtime multi-person 2D pose estimation using part affinity fields," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (Honolulu, HI: IEEE). doi: 10.1109/CVPR.2017.143

Casasanto, D. (2017). "Relationships between language and cognition," in *The Cambridge Handbook of Cognitive Linguistics*, ed B. Dancygier (Cambridge: Cambridge University Press), 19–37. doi: 10.1017/9781316339732.003

Casasanto, D., and Gijssels, T. (2015). What makes a metaphor an embodied metaphor? *Ling. Vanguard* 1, 327–337. doi: 10.1515/lingvan-2014-1015

Cox, A. (2017). *Music and Embodied Cognition: Listening, Moving, Feeling, and Thinking*. Bloomington, IN: Indiana University Press. doi: 10.2307/j.ctt200610s

Cuccio, V., and Fontana, S. (2017). "Embodied Simulation and metaphorical gestures," in *Embodied Simulation and Metaphorical Gestures*, eds F. Ervas, E. Gola, and M. G. Rossi (Berlin; Boston, MA: De Gruyter Mouton), 77–92. doi: 10.1515/9783110549928-005

Dolscheid, S., Çelik, S., Erkan, H., Küntay, A., and Majid, A. (2020). Space-pitch associations differ in their susceptibility to language. *Cognition* 196, 104073. doi: 10.1016/j.cognition.2019.104073

Dolscheid, S., Shayan, S., Majid, A., and Casasanto, D. (2013). The thickness of musical pitch: psychophysical evidence for linguistic relativity. *Psychol. Sci.* 24, 613–621. doi: 10.1177/0956797612457374

Eitan, Z. (2013). "How pitch and loudness shape musical space and motion," in *The Psychology of Music in Multimedia*, eds. S. L. Tan, A. J. Cohen, S. D. Lipscomb and R. A. Kendall (New York, NY: Oxford University Press), 165–191. doi: 10.1093/acprof:oso/9780199608157.003.0008

Eitan, Z. (2017). "Musical connections: cross-modal correspondences." in *The Routledge Companion to Music Cognition*, eds R. Timmers and R. Ashley (Abingdon: Routledge). doi: 10.4324/9781315194738-18

Eitan, Z., and Timmers, R. (2010). Beethoven's last piano sonata and those who follow crocodiles: cross-domain mappings of auditory pitch in a musical context. *Cognition* 114, 405–422. doi: 10.1016/j.cognition.2009.10.013

Fernandez-Prieto, I., Spence, C., Pons, F., and Navarra, J. (2017). Does language influence the vertical representation of auditory pitch and loudness? *iPerception* 8, 2041669517716183. doi: 10.1177/2041669517716183

Globerson, E., Flash, T., and Eitan, Z. (2021). "Space, time and expression in orchestral conducting," in *Space-Time Geometries for Motion and Perception in the Brain and the Arts*, eds T. Flash and A. Berthoz (New York, NY: Springer International Publishing), 199–212. doi: 10.1007/978-3-030-57227-3_10

González Sánchez, V., Zelechowska, A., and Jensenius, A. R. (2020). Analysis of the movement-inducing effects of music through the fractality of head sway during standstill. *J. Mot. Behav.* 52, 734–749. doi: 10.1080/00222895.2019.1689909

Holler, J., Drijvers, L., Rafiee, A., and Majid, A. (2022). Embodied space-pitch associations are shaped by language. *Cogn. Sci.* 46, e13083. doi: 10.1111/cogs.13083

Hostetter, A. B., and Alibali, M. W. (2019). Gesture as simulated action: revisiting the framework. *Psychon. Bull. Rev.* 26, 721–752. doi: 10.3758/s13423-018-1548-0

Kolter, A., Ladewig, S. H., Summa, M., Cornelia, M., Koch, S. C., and Fuchs, T. (2012). "Body memory and the emergence of metaphor in movement and speech: an interdisciplinary case study," in *Body Memory, Metaphor and Movement*, eds S. C. Koch, T. Fuchs, and M. Summa (Amsterdam: John Benjamins Publishing), 201–226 doi: 10.1075/aicr.84.16kol

Küssner, M. B., Tidhar, D., Prior, H. M., and Leech-Wilkinson, D. (2014). Musicians are more consistent: gestural cross-modal mappings of pitch, loudness and tempo in real-time. *Front. Psychol.* 5, 789. doi: 10.3389/fpsyg.2014.00789

Lidji, P., Kolinsky, R., Lochy, A., and Morais, J. (2007). Spatial associations for musical stimuli: a piano in the head? *J. Exp. Psychol. Hum. Perception Performance* 33, 1189–1207. doi: 10.1037/0096-1523.33.5.1189

Maimon, N. B., Lamy, D., and Eitan, Z. (2021). Space oddity: musical syntax is mapped onto visual space. *Sci. Rep.* 11, 22343. doi: 10.1038/s41598-021-01393-1

Maiorani, A. (2021). *Kinesemiotics*. New York, NY: Routledge. doi: 10.4324/9780429297946

Meissl, K., Sambre, P., and Feyaerts, K. (2022). Mapping musical dynamics in space. A corpus-based analysis of conductors' movements in orchestra rehearsals. *Front. Commun.*

Müller, C., and Ladewig, S. H. (2013). "Metaphors for sensorimotor experiences. Gestures as embodied an dynamic conceptualizations of balance in dance lessons," in *Language and the Creative Mind*, eds B. Dancygier, J. Hinnell, and M. Borkrent (Stanford, CA: CLSI), 295–324.

Müller, C., and Tag, S. (2010). The dynamics of metaphor: foregrounding and activating metaphoricity in conversational interaction. *Cogn. Semiotics* 6, 85–120. doi: 10.1515/cogsem.2010.6.spring2010.85

Parise, C. V. (2015). Crossmodal correspondences: standing issues and experimental guidelines. *Multisensory Res.* 29, 7–28. doi: 10.1163/22134808-00002502

Parkinson, C., Kohler, P. J., Sievers, B., and Wheatley, T. (2012). Associations between auditory pitch and visual elevation do not depend on language: Evidence from a remote population. *Perception*. 41, 854–861. doi: 10.1068/p7225

Pearson, L., and Pouw, W. (2022). Gesture-vocal coupling in Karnatak music performance: a neuro-bodily distributed aesthetic entanglement. *Ann. N. Y. Acad. Sci*. 2022, 1–18. doi: 10.1111/nyas.14806

Pitteri, M., Marchetti, M., Grassi, M., and Priftis, K. (2021). Pitch height and brightness both contribute to elicit the SMARC effect*:* a replication study with expert musicians. *Psychol. Res.* 85, 2213–2222. doi: 10.1007/s00426-020-01395-0

Pouw, W., and Dixon, J. A. (2020). Gesture networks: Introducing dynamic time warping and network analysis for the kinematic study of gesture ensembles. *Discour. Process*. 57, 301–319. doi: 10.1080/0163853X.2019.1678967

Prové, V., and Feyaerts, K. (2022). Pitch metaphors and the body in singing classes. *CogniTextes*, 22. doi: 10.4000/cognitextes.2037

Prpic, V., and Domijan, D. (2018). Linear representation of pitch height in the SMARC effect. *Psychol. Topics* 27, 437–452. doi: 10.31820/pt.27.3.5

Rusconi, E., Kwan, B., Giordano, B. L., Umiltà, C., and Butterworth, B. (2006). Spatial representation of pitch height: the SMARC effect. *Cognition* 99, 113–129. doi: 10.1016/j.cognition.2005.01.004

Shayan, S., Ozturk, O., and Sicoli, M. A. (2011). The thickness of pitch: crossmodal metaphors in Farsi, Turkish, and Zapotec. *Senses Soc.* 6, 96–105. doi: 10.2752/174589311X12893982233911

Spence, C. (2011). Crossmodal correspondences: a tutorial review. *Atten. Percept. Psychophys.* 73, 971–995. doi: 10.3758/s13414-010-0073-7

Swarbrick, D., Bosnyak, D., Livingstone, S. R., Bansal, J., Marsh-Rollo, S., Woolhouse, M. H., et al. (2019). How live music moves us: head movement differences in audiences to live versus recorded music. *Front. Psychol.* 9, 2682. doi: 10.3389/fpsyg.2018.02682

Timmers, R., and Li, S. (2016). Representation of pitch in horizontal space and its dependence on musical and instrumental experience. *Psychomusicol. Music Mind Brain* 26, 139–148. doi: 10.1037/pmu0000146

Turner, G., and Kenny, D. T. (2011). Restraint of body movement potentially reduces peak SPL in western contemporary popular singing. *Musicae Sci.* 16, 357–371. doi: 10.1177/1029864911423164

Walker, P. (2016). Cross-sensory correspondences: a theoretical framework and their relevance to music. *Psychomusicol. Music Mind Brain* 26, 103–116. doi: 10.1037/pmu0000130

Winter, B. (2013). Linear models and linear mixed effects models in R with linguistic applications. *arXiv [Preprint]*. arXiv:1308.5499. Available online at: http://arxiv.org/pdf/1308.5499.pdf (accessed September 14, 2022).

Wittenburg, P., Brugman, H., Russel, A., Klassmann, A., and Sloetjes, H. (2006). "ELAN : a professional framework for multimodality research," in *Proceedings of LREC 2006, Fifth International Conference on Language Resources and*

Evaluation, 1556–1559. Available online at: https://pure.mpg.de/pubman/faces/ViewItemOverviewPage.jsp?itemId=item_60436 (accessed December 13, 2021).

Wöllner, C., and Jensenius, A. R. eds. (2017). "Exploring music-related micromotion," in *Body, Sound and Space in Music and Beyond: Multimodal Explorations* (London: Routledge), 29–48. doi: 10.4324/9781315569628-3

Xie, Y., Allaire, J. J., and Grolemund, G. (2018). *R Markdown: The Definitive Guide*. Boca Raton, FL: Chapman and Hall/CRC. doi: 10.1201/9781138359444

Zbikowski, L. M. (2002). *Conceptualizing Music: Cognitive Structure, Theory, and Analysis*. Oxford: Oxford University Press. doi: 10.1093/acprof:oso/9780195140231.001.0001

Zelechowska, A., Gonzalez-Sanchez, V. E., Laeng, B., and Jensenius, A. R. (2020). Headphones or Speakers? An exploratory study of their effects on spontaneous body movement to rhythmic music. *Front. Psychol.* 11, 698. doi: 10.3389/fpsyg.2020.00698

Zlatev, J., and Devylder, S. (2020). "Cutting and breaking metaphors of the self and the Motivation and Sedimentation Model," in *Figurative Meaning Construction in Thought and Language*, ed A. Baicchi (Amsterdam: John Benjamins Publishing Company), 225–282.