



SONIFICATION, PERCEPTUALIZING BIOLOGICAL INFORMATION

EDITED BY: Diego Minciocchi and David Rosenboom

PUBLISHED IN: Frontiers in Neuroscience and Frontiers in Neurology



frontiers

Frontiers eBook Copyright Statement

The copyright in the text of individual articles in this eBook is the property of their respective authors or their respective institutions or funders. The copyright in graphics and images within each article may be subject to copyright of other parties. In both cases this is subject to a license granted to Frontiers.

The compilation of articles constituting this eBook is the property of Frontiers.

Each article within this eBook, and the eBook itself, are published under the most recent version of the Creative Commons CC-BY licence.

The version current at the date of publication of this eBook is CC-BY 4.0. If the CC-BY licence is updated, the licence granted by Frontiers is automatically updated to the new version.

When exercising any right under the CC-BY licence, Frontiers must be attributed as the original publisher of the article or eBook, as applicable.

Authors have the responsibility of ensuring that any graphics or other materials which are the property of others may be included in the CC-BY licence, but this should be checked before relying on the CC-BY licence to reproduce those materials. Any copyright notices relating to those materials must be complied with.

Copyright and source acknowledgement notices may not be removed and must be displayed in any copy, derivative work or partial copy which includes the elements in question.

All copyright, and all rights therein, are protected by national and international copyright laws. The above represents a summary only. For further information please read Frontiers' Conditions for Website Use and Copyright Statement, and the applicable CC-BY licence.

ISSN 1664-8714

ISBN 978-2-88963-868-0

DOI 10.3389/978-2-88963-868-0

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

SONIFICATION, PERCEPTUALIZING BIOLOGICAL INFORMATION

Topic Editors:

Diego Minciacchi, University of Florence, Italy

David Rosenboom, California Institute of the Arts, United States



"2019-79314 Picture" by Cecilia Bello Minciacchi

Citation: Minciacchi, D., Rosenboom, D., eds. (2020). Sonification, Perceptualizing Biological Information. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-88963-868-0

Table of Contents

05	<i>Editorial: Sonification, Perceptualizing Biological Information</i> Diego Minciacchi, David Rosenboom, Riccardo Bravi and Erez James Cohen
07	<i>Movement Sonification: Effects on Motor Learning Beyond Rhythmic Adjustments</i> Alfred O. Effenberg, Ursula Fehse, Gerd Schmitz, Bjoern Krueger and Heinz Mechling
25	<i>More Feedback is Better than Less: Learning a Novel Upper Limb Joint Coordination Pattern With Augmented Auditory Feedback</i> Shinya Fujii, Tea Lulic and Joyce L. Chen
38	<i>Sonification of Arm Movements in Stroke Rehabilitation – A Novel Approach in Neurologic Music Therapy</i> Daniel S. Scholz, Sönke Rohde, Nikou Nikmaram, Hans-Peter Brückner, Michael Großbach, Jens D. Rollnik and Eckart O. Altenmüller
45	<i>Sensori-Motor Learning With Movement Sonification: Perspectives From Recent Interdisciplinary Studies</i> Frédéric Bevilacqua, Eric O. Boyer, Jules Françoise, Olivier Houix, Patrick Susini, Agnès Roby-Brami and Sylvain Hanneton
52	<i>Biomusic: An Auditory Interface for Detecting Physiological Indicators of Anxiety in Children</i> Stephanie Cheung, Elizabeth Han, Azadeh Kushki, Evdokia Anagnostou and Elaine Biddiss
62	<i>Interactive Sonification of Spontaneous Movement of Children—Cross-Modal Mapping and the Perception of Body Movement Qualities Through Sound</i> Emma Frid, Roberto Bresin, Paolo Alborn and Ludvig Elblaus
78	<i>3Mo: A Model for Music-Based Biofeedback</i> Pieter-Jan Maes, Jeska Buhmann and Marc Leman
91	<i>Commentary: Environmental Sound Training in Cochlear Implant Users</i> Nicholas Altieri
94	<i>On the Auditory-Proprioception Substitution Hypothesis: Movement Sonification in Two Deafferented Subjects Learning to Write New Characters</i> Jérémy Danna and Jean-Luc Velay
106	<i>Mapping Sonification for Perception and Action in Motor Skill Learning</i> John F. Dyer, Paul Stapleton and Matthew Rodger
110	<i>A Novel Sonification Approach to Support the Diagnosis of Alzheimer's Dementia</i> Letizia Gionfrida and Agnieszka Roginska
122	<i>Action Observation Plus Sonification. A Novel Therapeutic Protocol for Parkinson's Patient With Freezing of Gait</i> Susanna Mezzarobba, Michele Grassi, Lorella Pellegrini, Mauro Catalan, Bjorn Kruger, Giovanni Furlanis, Paolo Manganotti and Paolo Bernardis

135 *Interactive Sonification Exploring Emergent Behavior Applying Models for Biological Information and Listening*

Insook Choi

164 *The Use of Footstep Sounds as Rhythmic Auditory Stimulation for Gait Rehabilitation in Parkinson's Disease: A Randomized Controlled Trial*

Mauro Murgia, Roberta Pili, Federica Corona, Fabrizio Sors,
Tiziano A. Agostini, Paolo Bernardis, Carlo Casula, Giovanni Cossu,
Marco Guicciardi and Massimiliano Pau



Editorial: Sonification, Perceptualizing Biological Information

Diego Minciacchi^{1*}, David Rosenboom², Riccardo Bravi¹ and Erez James Cohen¹

¹ Physiological Sciences Section, Department of Experimental and Clinical Medicine, University of Florence, Florence, Italy,

² The Herb Alpert School of Music, California Institute of the Arts, Valencia, CA, United States

Keywords: audio display, music, movement, sound, sonification

Editorial on the Research Topic

Sonification, Perceptualizing Biological Information

Sonification is a subtype of auditory displays that use sound structures devoid of linguistic elements to represent information. Kramer et al. (1999) have neatly defined sonification as “the transformation of data relations into perceived relations in an acoustic signal for the purposes of facilitating communication or interpretation.” The sense of hearing has the potential to convey in a simple way information that is complementary or alternative to visualization. Applications of sonification are by no means recent, in fact from early formulation of auditing, such as comparing the sound of commodities or “hearing of accounts” in Mesopotamia, as early as 3500 BCE (Worrall, 2009), to the three inventions from the nineteenth century, the Bell telephone, the Edison phonograph, and the Marconi radiotelegraphy, sound and audio were used and transformed to convey information. Many of these developments included translations of information into sound and were so powerful and rich of outcomes to change ultimately our relation to hearing in general. Commonly, implementations of sound constructions for information display are described for alarms, alerts, and warnings, status, process, and monitoring messages, data exploration, and finally for entertainment, sports, exercise, and art (Walker and Nees, 2011).

Though simplistic forms of sonification were always employed to represent phenomena from the physical world, the rapid developments at the end of last century in psychoacoustics, data manipulation, sound synthesis, and sonification techniques, caused an outburst of research in this field (Kramer, 1994; Hermann et al., 2011). With technological advancements the amount of data being processed is immense. In current research, regardless of the field, resorting to datamining is often inevitable. For this exact reason, representation of data has an important role. For certain applications, the most immediate representation of data is confined to the visual. However, there are still limitations when it comes to the visual representation of data by itself. For example, within medical diagnostics, even the most refined machinery is limited by our own perception of reality. Also, a visual representation is often a static one, and therefore a partial representation. Audio on the other hand, a time based experience by definition, is a dynamic representation, though also limited by our perception, it does add another dimension to our exploration of data. The addition of another dimension for evaluating data therefore, appears not only to be necessary but also natural. Any interaction of this world is done with our senses, and as such the more participating senses are during the exploration, the clearer the image and the subtleties of the world. Conversely, confinement of the exploration to a single dimensionality, is bound to limit our understanding.

With these premises, the implementation of sonification into various fields could provide major advances for the interpretation of data. Specifically, the approach was recently integrated within the field of neuroscience to facilitate the understanding of biological mechanisms and structures.

OPEN ACCESS

Edited and reviewed by:

Mikhail Lebedev,
Duke University, United States

*Correspondence:

Diego Minciacchi
diego.minciacchi@unifi.it

Specialty section:

This article was submitted to
Neuroprosthetics,
a section of the journal
Frontiers in Neuroscience

Received: 29 February 2020

Accepted: 04 May 2020

Published: 04 June 2020

Citation:

Minciacchi D, Rosenboom D, Bravi R
and Cohen EJ (2020) Editorial:
Sonification, Perceptualizing Biological
Information. *Front. Neurosci.* 14:550.
doi: 10.3389/fnins.2020.00550

Applications are manifold including behavioral monitoring, complex data extracting and analysis, algorithm development and interface implementations. Also, central application areas such as data navigation, status and process monitoring, motor tutoring for sports and exercise, and assistive technology for people with disabilities, became sites of strategic interest for sonification. Finally, sonification is increasingly being used as an aesthetic concept and method in the artistic and entertainment domain.

The articles presented in this Research Topic can be divided into three general sections. The first section includes articles concentrated on the implementations of sonification in movement research. Specifically, studies were oriented toward the examination of the effect of augmented sensorial feedback, by means of sonification, on motor learning. Movement sonification is applied on motor learning in sports (Effenberg et al.) and to test augmented sonified feedback on the motor learning of a novel joint coordination pattern (Fujii et al.). This line was further expanded by examining the possible benefits of sonification methods for sensori-motor learning with movement sonification (Bevilacqua et al.) and a perception-action approach to sonification used as feedback for skill learning (Dyer et al.), which may lead researchers toward encouraging applications in rehabilitation, sport training or product design. The section is concluding with interactive and real time implementations of sonification methods to the benefit of psychological and physical optimization of sports and motor rehabilitation tasks through the main functions Motivate, Monitor, and Modify of the 3Mo model (Maes et al.) and to analyze children's spontaneous movement in terms of energy, smoothness and directness (Frid et al.).

The second section focuses mainly on sonification as an open-ended design task to construct general sound information processes that translate data into sound maintaining reproducibility when sources exhibit non-linear properties

of self-organization and emergent behavior (Choi) and a specific supportive auditory tool to aid in diagnosing patients with different levels of Alzheimer's that introduces an audible parameter mapped upon different brain's lobes (Gionfrida and Roginska), and an auditory interface for intuitive detection and management of anxiety from physiological signals (Cheung et al.).

The third section overviews implementations of sonification methods as means of therapeutics. For rehabilitation purposes in socially relevant movement disorders such as Parkinson's disease, where sonification has been used to help relearn gait movements and to reduce freezing episodes (Mezzarobba et al.; Murgia et al.), and stroke patients, where is presented an innovative musical sonification therapy, designed to retrain patients' gross motor functions of the upper extremity (Scholz et al.). Furthermore, a real-time auditory feedback based on movement sonification approach is used to compensate proprioceptively deafferented subjects (Danna and Velay), and a modification of the environmental sound training procedure is utilized to enhance neural plasticity, and reconstruct auditory representations that have become degraded after chronic use of cochlear implants (Altieri et al.).

Though the benefits of sonification are clear and become well-evidenced in this topic and most recent literature (see, Schaffert et al., 2019), widespread implementation is still scarce. We believe that a collective endeavor is necessary to put theory into practice and, perhaps, to start to exploit the large potential of sonifications.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

REFERENCES

- Hermann, T., Hunt, A., and Neuhoﬀ, J. (eds.). (2011). *The Sonification Handbook*. Berlin: Logos Verlag Berlin GmbH.
- Kramer, G. (ed.). (1994). *Auditory Display: Sonification, Audification, and Auditory Interfaces*. Boca Raton, FL: CRC Press.
- Kramer, G., Walker, B., Bonebright, T., Cook, P., Flowers, J. H., Miner, N., et al. (1999). *Sonification Report: Status of the Field and Research Agenda*. Faculty Publications, Department of Psychology.
- Schaffert, N., Janzen, T. B., Mattes, K., and Thaut, M. H. (2019). A review on the relationship between sound and movement in sports and rehabilitation. *Front. Psychol.* 10:244. doi: 10.3389/fpsyg.2019.00244
- Walker, B., and Nees, M. (2011). "Theory of sonification," in *The Sonification Handbook*, eds T. Hermann, A. Hunt, and J. Neuhoﬀ (Berlin: Logos Publishing House), 9–39.

- Worrall, D. (2009). *Sonification and information: concepts, instruments and techniques* (Doctoral thesis). University of Canberra (AU).

Conflict of Interest: 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.

Copyright © 2020 Minciacchi, Rosenboom, Bravi and Cohen. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Movement Sonification: Effects on Motor Learning beyond Rhythmic Adjustments

Alfred O. Effenberg^{1*}, Ursula Fehse¹, Gerd Schmitz¹, Bjoern Krueger² and Heinz Mechling³

¹ Faculty of Humanities, Institute of Sports Science, Leibniz Universität Hannover, Hanover, Germany, ² Computer Science, Faculty of Mathematics and Natural Sciences, Institute of Computer Science II, University of Bonn, Bonn, Germany, ³ Institute of Sport Gerontology, German Sport University Cologne, Cologne, Germany

OPEN ACCESS

Edited by:

Diego Minicacchi,
University of Florence, Italy

Reviewed by:

Tiziano A. Agostini,
University of Trieste, Italy
Martin Lotze,
University of Greifswald, Germany

*Correspondence:

Alfred O. Effenberg
alfred.effenberg@sportwiss.
uni-hannover.de

Specialty section:

This article was submitted to
Neuroprosthetics,
a section of the journal
Frontiers in Neuroscience

Received: 28 February 2016

Accepted: 02 May 2016

Published: 27 May 2016

Citation:

Effenberg AO, Fehse U, Schmitz G,
Krueger B and Mechling H (2016)
Movement Sonification: Effects on
Motor Learning beyond Rhythmic
Adjustments. *Front. Neurosci.* 10:219.
doi: 10.3389/fnins.2016.00219

Motor learning is based on motor perception and emergent perceptual-motor representations. A lot of behavioral research is related to single perceptual modalities but during last two decades the contribution of multimodal perception on motor behavior was discovered more and more. A growing number of studies indicates an enhanced impact of multimodal stimuli on motor perception, motor control and motor learning in terms of better precision and higher reliability of the related actions. Behavioral research is supported by neurophysiological data, revealing that multisensory integration supports motor control and learning. But the overwhelming part of both research lines is dedicated to basic research. Besides research in the domains of music, dance and motor rehabilitation, there is almost no evidence for enhanced effectiveness of multisensory information on learning of gross motor skills. To reduce this gap, movement sonification is used here in applied research on motor learning in sports. Based on the current knowledge on the multimodal organization of the perceptual system, we generate additional real-time movement information being suitable for integration with perceptual feedback streams of visual and proprioceptive modality. With ongoing training, synchronously processed auditory information should be initially integrated into the emerging internal models, enhancing the efficacy of motor learning. This is achieved by a direct mapping of kinematic and dynamic motion parameters to electronic sounds, resulting in continuous auditory and convergent audiovisual or audio-proprioceptive stimulus arrays. In sharp contrast to other approaches using acoustic information as error-feedback in motor learning settings, we try to generate additional movement information suitable for acceleration and enhancement of adequate sensorimotor representations and processible below the level of consciousness. In the experimental setting, participants were asked to learn a closed motor skill (technique acquisition of indoor rowing). One group was treated with visual information and two groups with audiovisual information (sonification vs. natural sounds). For all three groups learning became evident and remained stable. Participants treated with additional movement sonification showed better performance compared to both other groups. Results indicate that movement sonification enhances motor learning of a complex gross motor skill—even exceeding usually expected acoustic rhythmic effects on motor learning.

Keywords: audiovisual information, motor learning, motor perception, motor rehabilitation, movement sonification, multisensory integration

INTRODUCTION

When looking back to our sport classes, recalling how breaststroke swimming, the overhand technique in volleyball or even rowing was taught, we remember our teachers explaining and demonstrating the techniques. Technique acquisition in sports is usually shaped by visual demonstrations and verbal information as getting evident in popular sportscientific textbooks (Newell and Corcos, 1993; Schmidt and Lee, 2005). Also in perceptually directed research in sport science, processes of motor perception, motor control and motor learning have been studied primarily related to single sensory modalities and dominated by the visual domain (Williams et al., 1999, 2004; Abernethy, 2013). But on a closer view, motor behavior is a multimodal phenomenon: Motion can not only be observed visually but also perceived by the auditory and the tactile sense, and perception of one's own motion is just as well based on visual, auditory, kinesthetic, vestibular, and tactile information. Recent behavioral—as well as neurophysiological—research therefore focusses increasingly on audiomotor and multisensory contributions to the regulation of behavior (Frassinetti et al., 2002; Soto-Faraco et al., 2003; Calvert et al., 2004). Even though majority of work is localized in the field of basic research, also applied studies address the area of complex gross-motor behavior, often with a close link to biological motion perception (Barraclough et al., 2005; Bidet-Caulet et al., 2005; Mendonca et al., 2011).

Up to now, only a few studies are dealing with the multisensory influence on motor learning, and this is especially given for applied research related to gross motor motion, as being typical for sports. Therefore, the introduction will focus firstly on audiomotor information processing to identify the perceptual characteristics of audition, getting effective besides visual information on the regulation of behavior (Hauelsen and Knoesche, 2001; Bangert and Altenmüller, 2003; Haslinger et al., 2005; Lahav et al., 2007). Afterwards multisensory perception is taken into account, with the focus on mechanisms of audiovisual information processing and related behavioral benefits. Then some studies using sonification to support motor control and motor learning are introduced.

Findings on the emergence of audiomotor co-activations and the multisensory integration mechanisms will be taken in consideration to determine how additional movement acoustics could be shaped to address audiomotor functions as well as multisensory integration sites within the central nervous system (CNS). With other words: How could an effective movement sonification be tailored and how could it get effective on motor learning? Here some neurophysiological work will be consulted. Based on these findings the own method of movement sonification will be developed combining dynamic and kinematic movement parameters into a 4-dimensional sonification. Movement sonification was applied in the present study to support motor learning in technique acquisition of indoor-rowing. To evaluate the impact of an additional movement sonification three groups were treated with different kinds of instructions and feedback over a training period of 3 weeks: One group was treated with visual information (video instruction +

concurrent video feedback) and two groups with different kinds of audiovisual information (video/sonification instruction & real-time video/sonification feedback; video/motion attendant sound instruction & real-time video/motion attendant sounds feedback).

Music Making—Acoustic Information of Motor Behavior Generated by the Auditory System

In a first step, music related research will be focused. In the music domain research on motor learning is related to auditory, especially musical perception and thus is an appreciated supplement to visually dominated motor learning research in the fields of sport, ergonomics or motor rehabilitation. More than this: Music making is a domain of motor behavior with excellent subtle and unambiguous feedback about the precision and the quality of motor control. On music making, quality of motor control can be assessed immediately via the acoustical or musical result resp. When for example playing the keyboard, the spatial accuracy of actions can be assessed via the sound frequency or tone pitch sequence, the dynamical precision via the sound amplitude and the temporal exactness via the duration of sounds and pauses as well as via constancy of metrum and shape of rhythm. With other words: Auditory and especially musical perceptual skills are very well-suitable for analyzing several important qualities of motor control. But what are the appropriate perceptual and motor features established into the brain of an expert musician? The work of Hauelsen and Knoesche (2001) indicates that on expert pianists, when listening to a piece of one-handed piano music, motor areas primarily related to the analogous effector get co-activated.

Obviously, specific audiomotor networks get established in music experts—at least after years of exercise. But how long does it take till such sensorimotor networks get initially established? There already exists some empirical evidence that even on novices it does not take long time till auditory-motor co-activations appear within the human brain: Bangert and Altenmüller (2003) reported a fast emergence of audiomotor co-activation patterns on musical novices learning to play a simple melody on a keyboard. In an EEG-study, the authors reported audiomotor co-activations which develop within short temporal intervals of about 20 min of practice and get firmly established within a few weeks of exercise—nevertheless with a specific shape. Only the replay of the before trained simple melody led to motor co-activation, whereas new melodies, played on the same keyboard, did not.

Non-Musical Acoustic Information on Motor Behavior

When dealing with acoustic information related to motor behavior, also natural motion attendant sound and the information about the related action as well as about the related actor have to be taken into account. During last years, a growing body of empirical studies on the information coded within natural motion sounds has been acquired. Early

evidence about the impact of motion attendant sounds on motor performance has been presented by Takeuchi (1993) on tennis players indicating that auditory perception of tennis ball sounds (stroke and landing) supports a high performance and deprivation of auditory perception reduces the performance. The information coded within natural motion sounds was described by Effenberg (1996) and used as an initial point for the development of an ecological framework of movement sonification. In 2004, Agostini et al. demonstrated that athletes were able to use auditory models of hammer throws to improve their performance. The high amount of information mediated by natural movement sounds has also been illustrated by more recent studies using the own/other paradigm (Murgia et al., 2012; Kennel et al., 2014). Beyond that basic physiological parameters like the breath duration can be affected subconsciously by listening to the ecological sound of breathing even more compared to artificial sound and thereby indicating a demanding influence of natural sounds (Murgia et al., 2016). On the other hand complex artificial movement sound is powerful to allow subtle distinctions of own vs. other movement patterns as shown related to movement sonification of indoor rowing by Schmitz and Effenberg (2012). Meanwhile there exists numerous evidence about acoustically coded information in natural movement sounds and about the subtle impact on motor behavior (Sors et al., 2015). Supportive fMRI research has been published by Woods et al. (2014) e.g., indicating, that expertise in a certain sport is an important factor related to the way sport specific acoustic information is processed in the brain: Experts, familiar with the presented sport specific sounds showed greater neural activation in sensorimotor areas and areas responsible for auditory and motor planning (Agostini et al., 2004).

Music Making and Audiovisual Information Processing

Even though neurophysiological research on motor learning in the area of music is dominated by audiomotor relations, also audiovisual-motor interactions have been focused. Haslinger et al. (2005) found in an fMRI-study that pure observation of piano playing recruited auditory areas in experienced pianists and discussed the participation of mirror neurons within the inferior fronto-parieto-temporal network. Taken together with the findings of Bangert and Altenmüller (2003), it is getting obviously that learning to play an instrument results in relatively fixed co-activations between different perceptual—at least auditory and visual—and motor networks. If once established, visual stimulation is sufficient for co-activation of auditory areas as well as for co-activation of motor-related networks. Also auditory simulation leads to co-activation of motor-related networks. These findings got further support with an fMRI-study from Lahav et al. (2007) showing that such audiomotor as well as audiovisuomotor co-activations are case specific, that a certain co-activation pattern is referring to a certain musical pattern resp.: The established co-activation of motor subsystems only became evident if the melody, learned to play before, was hearable. Activation of the motor network

was much smaller when the order of the notes was changed and it disappeared by “motorically” unknown—untrained—melodies. The findings of Lahav et al. indicate that there is a common hearing-doing system, that is highly dependent on the individual’s motor repertoire, is getting established rapidly and is likely not limited to the field of music (Lahav et al., 2007).

Non-Musical Audiovisual Information Processing

For some years not only studies on audiomotor functions have been growing in number (Bangert et al., 2006), also crossmodal interactions related to the perceptual system as well as to motor control and motor learning are getting more and more into the focus of research (Seitz et al., 2006). The quality of perception is usually enhanced if distal events are perceived by at least two different senses compared to unimodal perception. Numerous studies on multisensory stimuli effectiveness on different aspects of behavior have been realized meanwhile and most of them deliver supporting evidence as realized by Vroomen and de Gelder (2000) showing with a stimulus detection task in a rapidly changing sequence of visual distractors. Further perceptual effects of audiovisual convergent stimuli have been described by Frassinetti et al. (2002) in terms of an increased detection rate of low intensity stimuli. Also Seitz et al. (2006) reported an enhanced ability for the detection and discrimination of coherent motion pattern for audiovisual trained subjects compared to unimodal trained ones. The discrimination ability of audiovisually presented objects was studied by Giard and Peronnet (1999), who reported more reliable discriminations for audiovisually coded transforming objects compared to unimodal presentation. Those enhanced activation correlates with better performance in detecting coherent motion.

Further studies in the field of applied research indicated enhanced effectiveness of convergent audiovisual information on motor perception and even on motor perception and motor control: Evidence on motor performance was delivered by Chiari et al. (2005) in terms of a real-time audio feedback on trunk kinematics enhancing the control of body sway. Rath and Rocchesso (2005) demonstrated that a tilting bar can be handled with a higher precision if additional motion acoustics are available. Furthermore, gross-motor sport movements can be assessed and reproduced with higher precision under an audiovisual condition compared to pure visual treatment as reported by Effenberg (2005).

Also on motor learning there exist some evidence on the supportive function of additional auditory information: Shea et al. (2001) have revealed supporting effectiveness of an additional auditory model. The authors observed besides a better performance of a simple motion pattern (rhythmic finger movement) even a more effective learning process: Required time was reduced and precision in terms of absolute as well as relative timing was enhanced. Later work of Kennedy et al. (2013) confirmed an enhanced temporal stability (retention) of a before learned 2:3 bimanual tapping task

especially for an audiovisual model. More recently Danna et al. (2013, 2015) reported first indications that also less rhythmic finemotor learning (handwriting acquisition of children) can be supported by additional auditory cues. Similar results had been presented recently by Effenberg et al. (2015): The acquisition of character handwriting of young children had been supported by an additional real-time sonification of the writing trace (“SoundScript”) in terms of an acceleration of the emergence of model-like character patterns. And the learning of gross motor movements—as they are typical on sports—has been supported with concurrent auditory feedback by Baudry et al. (2006). The authors found a time-stable benefit for the body segmental alignment on a circle movement performed on a pommel horse based on a training with concurrent auditory feedback. And most recent work from our workgroup delivered first evidence on the effectiveness of a 4-dimensional kinematic movement sonification in stroke rehabilitation on hemiparesis of the upper limbs after a 5-day training of everyday actions about only 20 min daily (Schmitz et al., 2014).

Neurophysiological Background

In addition to the behavioral findings primarily neurophysiologically intended research has been conducted in the last years focussing on the underlying mechanisms of multisensory integration responsible for performance enhancement. Cortical as well as subcortical multisensory integration sites are addressed by convergent multisensory stimuli in addition to unimodal visual and auditory functions and may be jointly responsible for perceptual and behavioral benefits (Calvert and Thesen, 2004; Beauchamp, 2005). Dedicated to gross-motor human motion perception there exists evidence on integrating motion sound within an area responsible for ‘visual’ biological motion perception, the posterior superior temporal sulcus (STSp; Bidet-Caulet et al., 2005). The first fMRI-study dedicated to the exploration of multisensory integration based on movement sonification was realized by our workgroup: Scheef et al. (2009) revealing, that an audiovisual stimulus of a counter-movement-jump (video and sonification of the ground reaction force) evokes real multisensory integration effects in terms of a supra-additive activation enhancement in Area V5/MT as well as enhanced activation in STS bilaterally, which both playing a role in audiovisual perception of biological motion.

Even if there exists only some evidence for direct connections between audiovisual perception and motor execution on humans—as described for the superior colliculus on cats by Stein and Meredith (1993) and Rowland and Stein (2008), the existence of audiovisual mirror neurons in the monkey brain has been demonstrated with single cell recordings realized by Kohler et al. (2002) and is also discussed on humans (Keyser et al., 2003). Baumann and Greenlee (2006) showed via fMRI on humans that convergent audiovisual motion stimuli evoke substantially larger activation in different brain areas (cluster within the superior temporal gyrus, supramarginal gyrus, superior parietal lobule, and cerebellum) compared to the addition of both unimodal activations.

Recent research from our workgroup (Schmitz et al., 2013) is indicating, that the activation of the mirror neuron system increases when convergent audiovisual information in terms of a movement sonification of a moving avatar is available to the perceptual systems. The superior temporal sulcus, inferior parietal cortex and premotor regions as well as subcortical structures showed enhanced activation of the action-observation-system in comparison to similar, but divergent audiovisual stimuli. Beyond this also key-players of the striato-thalamo-frontal motor loop got increasingly activated—which had been observed on untrained participants with no or nearly no experience in movement sonification. Though it has not been a new finding that short-term plasticity related to motor learning can appear even within minutes, it is important for our issue that based on artificial movement sound sensorimotor co-activation in terms of audiomotor or even audiovisual-motor co-activation emerges also on “novices” instantly. The term “novices” refers to participants without experience with movement sonification, but all of them were able to breaststroke. Taken together this is some functional evidence that enhancing motor execution with artificial movement acoustics is efficient to address additionally audiomotor as well as audiovisual-motor mechanisms within the context of motor learning.

Interim Resume

Music making and listening indicates the enormous capacity of the auditory system to generate and mediate information about movements. The observed prompt emergence of case specific audiomotor co-activation patterns indicated the rapid dynamics of neuronal plasticity with a high sensitivity for the specific shape of the referenced movement-acoustic—time-varying—stimuli. Also on non-musical behavior basal motion-related perceptual functions benefit from multisensory information processing as well as processes of perception and motor control of human movements. And even beyond that—motor learning is getting more effective and stable if based on multisensory information, as shown for fine-motor, rhythmic hand-/finger-movements and also for more complex writing-movements and even in first steps for a certain feature of a gross-motor sport-movement on a pommel horse as well as in stroke rehabilitation of hemiparesis. A sharp plea for multisensory training is given by Shams and Seitz (2008) referring to the additional integration of intermodal processing and multimodal integration functions in learning. Such findings are supported by a statistical approach from Ernst and Bühlhoff (2004) explaining the emergence of additional information from multisensory integration. And finally some more recent research from Shams et al. (2011) should be mentioned here indicating furthermore a retroaction of multisensory—audiovisual—training to the perception within each single modality.

Creating the Method of Real-Time Movement Sonification

The method of movement sonification is used by a growing number of researchers, when combining movement data with sound. We use this term, which we have created first in 2005 (Effenberg, 2005) related to motor perception and motor control,

for the acoustic transformation of kinematic and/or dynamic movement data. The fundamental idea is to tune in the ear into the process of motor perception, related to external motion of others (trainers, functioning as “models”) as well as to one’s own movements when functioning as an additional feedback channel to support, to enhance and to shape the emergent sensorimotor representations in terms of “internal models” (Wolpert et al., 1995, 2011). Considering that Sigrist et al. (2015) failed to generate a long-lasting learning benefit based on sonified acoustic error information on indoor rowing, which was consisting of only one dimension of one external movement feature during one phase of the movement cycle (the horizontal angle of the rowing oar during the recovery phase), we generate a continuous 4-dimensional movement acoustics based on two kinematic and two dynamic movement parameters.

This kind of real-time movement sonification is configured to be used for audiomotor processing as well as for enhancing and shaping multisensory representations efficiently. In contrast to the mode of “error-feedback,” as considered by Sigrist et al. (2013), the processing of our kind of movement acoustics is not dependent on conscious cognitive processing, because the processing—even multisensory integration—is mandatory if the stimulus is hearable and certain criteria of intermodal convergence are fulfilled. Resulting in an enhanced spectrum of movement information, usable as instruction as well as feedback, this kind of information supports the emergence of adequate sensorimotor/perceptuomotor representations (internal models). The used method is described in more detail in the following section. Here we conclude the first section with three research hypotheses:

- H1: Participants of both groups treated with convergent audiovisual information show better learning results in terms of a steeper learning curve/a faster approximation to the model technique.
- H2: Audiovisually treated groups are more precise in rhythmic demands of the indoor rowing technique.
- H3: Participants treated with complex sonification benefit in terms of a better coordination of the movement resulting in higher technical performance.

METHODS

Novices were asked to learn the basic technique of indoor rowing by visual or audiovisual instruction and feedback. When instructing the participants it was pointed out that the primary aim was not a maximum intensity of indoor rowing, but to reproduce the technical pattern as precisely as possible. The quality of the technical pattern was operationalized by four movement parameters: grip force and footrest forces as dynamic parameters and grip pull-out length and sliding seat position as kinematic parameters. The participants’ courses of these parameters during pull-out phase were compared to the model’s ones.

The similarity of the participants’ and the model’s technique was computed by using a DTW-algorithm designed for similarity calculation of two temporal sequences differing in length. Such

DTW-algorithms have already been used in different contexts like speech recognition and analysis of motion capture data as described in Rabiner and Juang (1993), Forbes and Fiume (2005) and Demuth et al. (2006).

Experimental Design

The experimental procedure extended to about 9 weeks, whereof the training period took 3 weeks. Participants were asked to acquire a basic technique of indoor rowing, demonstrated by a professional rowing athlete. The experimental procedure is visualized in **Figure 1**. With the initial two pretests data about the initial individual technique level as well as strength data were collected. Afterwards the total sample was divided up into three subsamples, parallelized on initial technique level and age. All three samples completed the same 3-week training period each obtaining different kinds of information in terms of instruction and real-time feedback. Participants trained two times a week. The procedure of a single training session is explained below. One week after the last training session a strength posttest was conducted. Three weeks after the last training session participants completed a technique retention test finally.

Participants

48 male volunteers participated in the experiment, all of them without any experience in rowing (mean age = 22.8 ± 5.0). All participants showed normal or corrected-to-normal vision and normal hearing¹. Though the Central Ethics Commission (CEC) at Leibniz University Hannover (LUH) was starting the assessment service for the first time not before 2012, the beginning of subject recruitment was started without a specific ethical approval but in accordance with the Ethics Guidelines of the “German Psychological Society” also including informed consent declaration, privacy and confidentiality and the final presentation of research results to all participants. All participants gave their written consent to participate in this psychological-behavioral study.

Experimental Conditions

All three subsamples ran through the same training procedure, each with a different kind of information in terms of instruction and real-time feedback.

- Visual condition (V): Treatment group V was only treated with video information.
- Natural audiovisual condition (AV_{nat}): Treatment group AV_{nat} was treated with video and natural motion attendant sounds.
- Sonified audiovisual condition (AV_{soni}): Treatment group AV_{soni} was treated with video and movement sonification.

Stimulus Material

Visual Stimuli

Instruction videos (rowing model) and feedback videos (participants’ performance) were taken from a lateral view (see **Figure 2**). Videos were projected on a big screen (260 × 195 cm) in front of the rowing ergometer type Concept II. For

¹Standard vision and hearing test: HTTS, Version 2.10, 00115.04711.

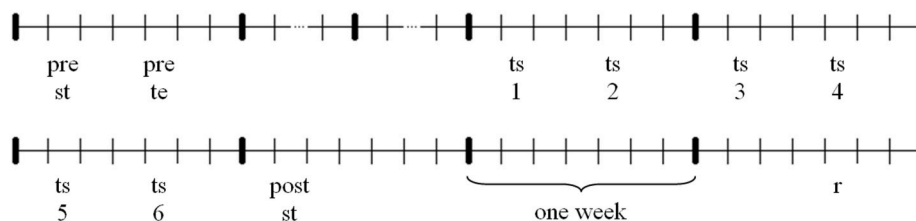


FIGURE 1 | Experimental procedure: pre st, strength pretest; pre te, technique pretest; ts1-6, training session 1-6; post st, strength posttest; r, technique retention test.

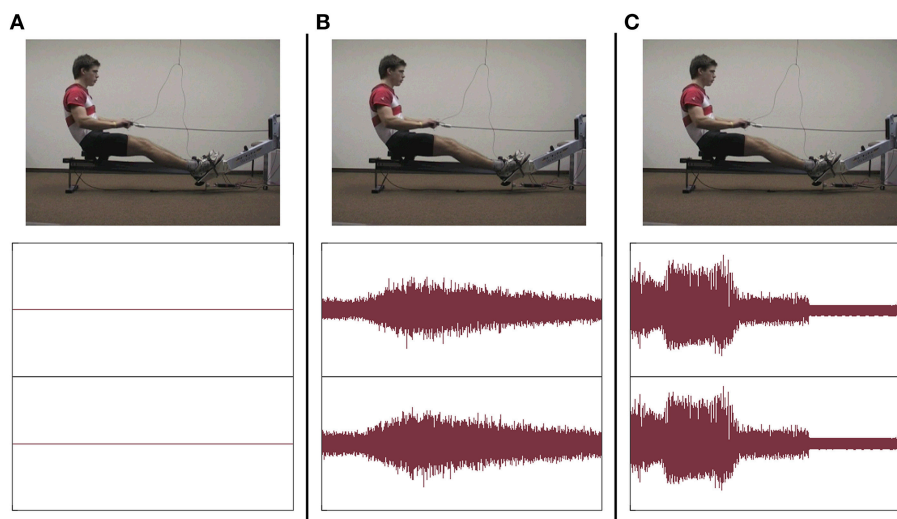


FIGURE 2 | Visual and auditory stimuli under the three conditions: frame of instruction video (above) + sound pressure level of soundtrack of a single rowing cycle (below); (A) group V, (B) group AV_{nat}, and (C) group AV_{soni}.

instruction and feedback, videos were presented with Sony Video Capture 6.0b.

Auditory Stimuli

Both kinds of auditory stimuli were presented via headphones. For group AV_{nat} the sound of the rowing ergometer flywheel and the sliding seat were taped with a directional microphone² and mediated via headphones³. For group AV_{soni} movement sonification based on two kinematic and two dynamic motion data streams that were transformed in real-time to multichannel continuous motion-sound. Data streams of four different sensors were acoustically represented: grip force, sum of footrest forces, grip pull-out length, and sliding seat position. In **Figure 2** visual and auditory stimuli under the three experimental conditions are depicted.

Kinematic and dynamic data were recorded using FES-Software⁴ and transmitted to Lab-View-Software⁵ and further on to sonification-software⁶. The sonification-software received

data of grip force, footrest forces, grip pull-out length and sliding seat position. Movement data were systematically mapped on sound features: each data stream was used to modulate frequency and partially also amplitude of a midi sound. Grip pull-out, grip force and footrest forces were represented continuously. For both force parameters a muting level was defined for values near around zero as well as for negative values. So forces could only be acoustically perceived when they were also kinesthetically clearly perceivable. By using a muting level oscillating sounds for fast changing forces near around zero were avoided. In contrast to the three continuously transformed parameters, sliding seat position was sonified event-related: it could be only heard at maximum and minimum position. Independently of effectively exerted force and realized grip pull-out length and sliding seat position, the frequency interval was chosen in a manner that maximum and minimum of a single data stream was related to the same frequency for each individual in each training session. Sonification of rowing model was produced the same way. This kind of normalization enables participants to produce the same sound pattern as the model, independently of individual absolute strength abilities and individual anthropometry. **Figure 3** shows the data curves of the kinematic (A) and the dynamic (B) parameters and the characteristics of the resulting sounds.

²Behringer, ECM 8000.

³beyer dynamic DT 100.

⁴FES, Ruderergometer Version 2.43.

⁵National Instruments, LabVIEW 7.1.

⁶MLmini, Universität Bonn, Institut für Informatik, AG Prof. Weber.

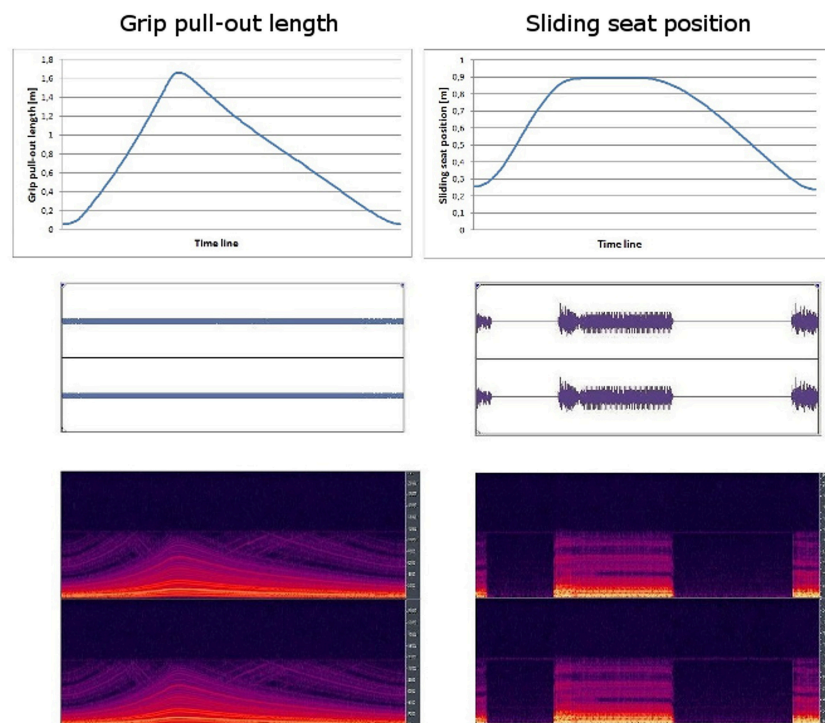
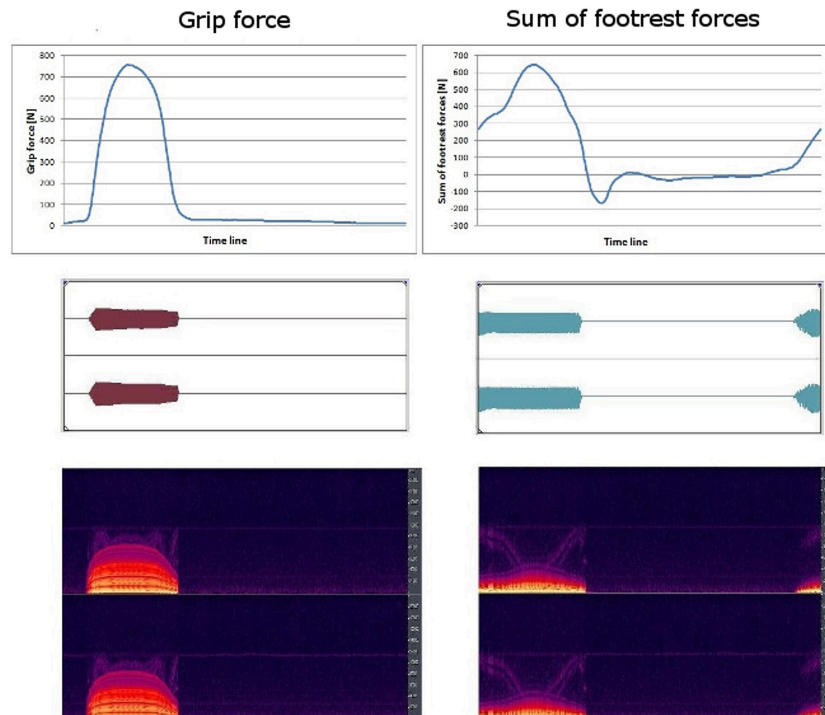
A Kinematic parameters**B Dynamic parameters**

FIGURE 3 | (A) Kinematic parameters; **(B)** Dynamic parameters. Data curve (up), amplitude level diagram (middle) and spectral graph (down) of the four data channels of a single rowing cycle (3.03 s).

Instruction

For the instruction-video the motion pattern of Eric Johannesen⁷ performing on the Concept II rowing ergometer (same rowing ergometer as used in the study, but higher drag factor) was videotaped. The videosequence contained 10 cycles of rowing, each lasting 3 s. Pull-out phase and recovery phase had a time ratio of 2:1. Depending on the treatment, there was no motion attendant soundtrack (V), the soundtrack contained natural motion attendant sounds of the rowing ergometer (AV_{nat}) or the model's movement sonification (AV_{soni}) resp.

Feedback

As feedback, participants observed their own rowing in real-time for ten cycles in the middle of each training block of 50 cycles. Depending on the treatment they heard no motion attendant sounds, their natural motion attendant sounds or their own movement sonification in real-time.

To mask natural motion attendant sounds all participants heard noise (sea rushing) via headphones while there was no auditory instruction or feedback. The chronology of sea rushing audio and audio-feedback within each block of training is illustrated in **Figure 4**: Audiovisual treatment groups heard 20 cycles of sea rushing followed by 10 feedback cycles (motion attendant sound or movement sonification resp.) followed by another 20 cycles of sea rushing.

Procedure

Pretest

After warming up, making themselves familiar with the rowing ergometer and watching the rowing model for ten cycles participants completed the technique pretest by rowing for 30 cycles without any feedback.

Training

One training session consisted of five training-blocks, each lasting for 2.5 min followed by 2.5 min break afterwards (see **Figure 4**). After the presentation of the model's technique as instruction, the participants started to row for 50 cycles. From cycles 21 to 30 the participants' own technique was presented as real-time feedback. Following the 50th cycle, the presentation of the instruction video was repeated.

Participants were instructed to align their rowing technique to the model's technique. In addition, they were told to tune their cycle frequency to 20 cycles per minute by watching the frequency display. To assure that participants do not tire—which would cause interferences with the quality of technique—they were instructed to row with comfortable effort. Group AV_{soni} was not informed in detail how sound was composed, they only knew that it was configured by their motion.

Retention

In retention participants completed one block of rowing without any instruction or feedback. Due to the long break between the last training session and retention of about 3 weeks, a warm-up of 50 cycles was realized before the retention block.

⁷Junior world champion 2005, 4+; world champion 2011 8+; Olympic Games champion 2012 8+.

Data Acquisition and Data Analysis

Personal data (name, sex, date of birth, health data etc.) were collected by questionnaire and made anonymous afterwards. In strength pretest and strength posttest isometric maximum strength of leg extension and arm flexion were measured with a leg-press (knee-angle: 90°) and a row machine (anteversion shoulder: 30°). Participants conducted the tests after warming up and making themselves familiar with the equipment. To determine the individual maximum strength, the best out of three trials was used. Because strength abilities generate an impact on motor performance, the development of strength was controlled by comparing data of strength pre- and posttests.

Four sensor systems were applied on the rowing ergometer: a resistance strain gauge for grip force (GF), two sensors for footrest forces (FF), and two incremental encoders each for grip pull-out length (GP) and sliding seat position (SS). All four parameters were recorded with 100 Hz, for footrest forces the sum of the two sensor streams was computed. For data analysis cycles 31–40 (21–30 for pretest) of each training block were selected. An average cycle was computed for each of the four raw data streams. In a second step data were normalized to eliminate differences in body size and individual strength. Grip pull-out and sliding seat data was normalized on values between 0 and 1, grip force and footrest forces data were only divided by the particular maximal value to maintain algebraic sign of measured values.

Using the dynamic-time-warping (DTW) algorithm (Müller, 2007) we calculated the distance values between the model's technique and participants' individual technique (normalized average pull-out phase) for each of the four parameters.

The corresponding procedures were performed for each of the regarded time series (normalized averaged curve of grip pull-out length, grip force, sum of footrest forces, sliding seat position) separately. As a result of the DTW algorithm we obtained an optimal alignment of the compared time series, a so called warping path. This warping path gives information how the time series have to be stretched or compressed to get an optimal matching. These temporal deformations do not have to be linear: some segments of a signal might be stretched while others are compressed to get the optimal alignment. We defined the accumulated costs along the warping path to be a distance measure to finally compare the time series.

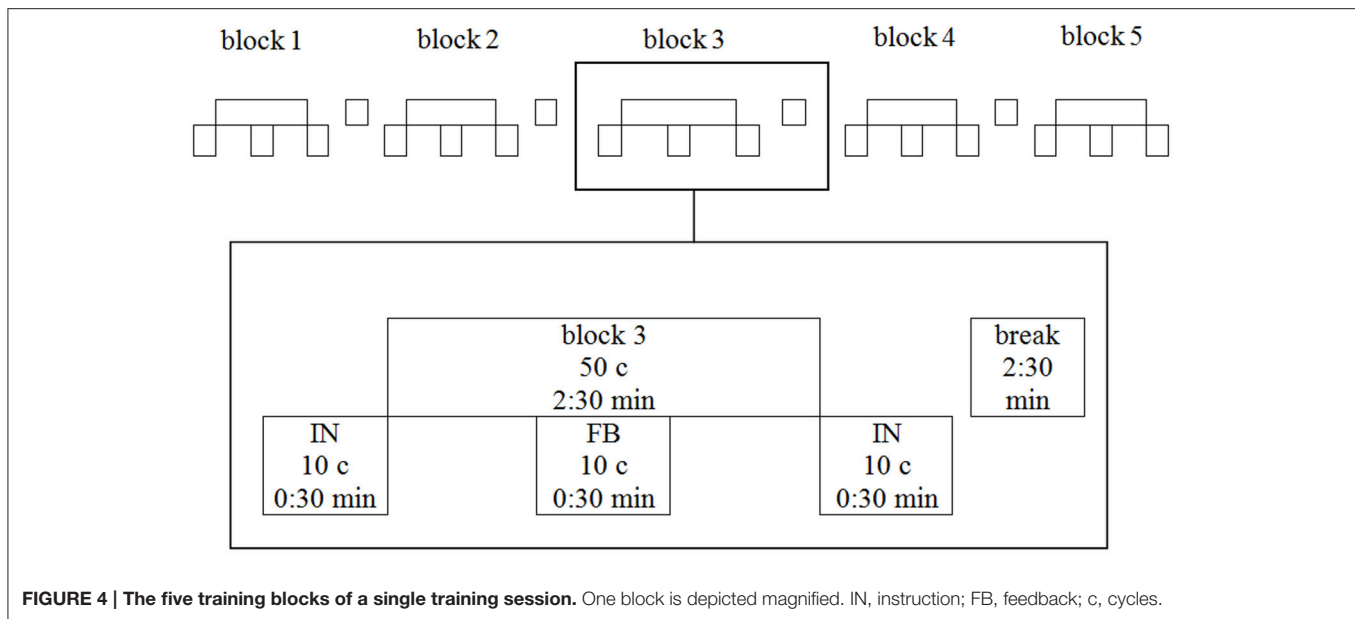
The computation of an alignment using DTW can be divided into three basic steps:

1. Computation of local distance matrix (LDM).
2. Computation of a global distance matrix (GDM).
3. Search the GDM for the warping path.

These steps are now described in detail:

1. Computation of the LDM

The local distance matrix is an $m \times n$ matrix, where m is the number of data points of the model's technique curve (s) and n is the number of data points of the participant's individual technique curve (t). Each entry of the matrix corresponds to



the absolute distance between one data point of the participant's technique curve and one data point of the model's technique curve. For example, the entry in the i -th row and the j -th column corresponds to the absolute distance between the i -th sample point of the participant's technique curve, and the j -th sample point of the model's technique curve.

```
for i: = 1 to n
  for j: = 1 to m
    LDM[i, j]: = d(s[i], t[j])
```

2. Computation of the GDM

Based on the LDM we are now able to compute the GDM. The GDM is an $m \times n$ matrix, too. It represents the accumulated or global costs between the time series regarded. Here the entry in the i -th row and j -th column is the minimal cost for an optimal alignment if the time series would end at these frames. The GDM is computed by accumulating local distances stored in the LDM according to a special scheme. The value of the entry in the i -th row and j -th column is computed as:

```
GDM[i, j]: = LDM[i, j] + minimum(GDM[i - 1, j],
  GDM[i, j - 1], GDM[i - 1, j - 1])
```

The computation starts at the entry [1, 1].

```
for i:= 1 to n
  for j:= 1 to m
    GDM[i, j]: = LDM[i, j]
      + minimum(GDM [i-1, j],
        GDM [i, j-1],
        GDM [i-1, j-1])
```

The entry [m, n] is taken as “vertical distance value.”

3. Search the warping path

After the computation of the GDM, we can now extract the optimal alignment. It leads from the entry [m, n] to the entry [1, 1], involving only single steps to the left, up or diagonally left up, leading to the entry with the lowest value, resp. The number of needed steps builds the “path length” and the ratio of “path length” and the minimal possible path length (in our case almost always $m-1$) builds the “horizontal distance value.” From vertical and horizontal distance value, a “distance value” was computed for each data stream. In order to respect the different dimensions of the two computed distance values, we decided to utilize Pythagorean theorem:

$$\text{distance value} = \sqrt{(\text{vertical distance value})^2 + (\text{horizontal distance value})^2}$$

To consider the rate of force, additionally a force index (FI) was built in terms of the average grip force during pull-out phase divided by the maximum strength value (MSV). Maximum strength value consisted of the mean of maximum strength data (sum of legpress & row machine) in strength pre- and posttest.

$$\text{MSV} = ((\text{maximum strength from legpress pretest} + \text{row machine pretest}) + (\text{maximum strength from legpress posttest} + \text{row machine posttest}))/2$$

$$\text{FI} = \text{average grip force during pull-out phase}/\text{MSV}$$

A combination of the four distances values with the force index allowed to build a general distance value (GDV, see **Figure 5**) for each block (five blocks a training session multiplied by six training sessions + pretest, retention test). GF was weighted fivefold due to its dominant importance on the acceleration of a rowing boat. To take into account the two technique influencing aspects, a formula was developed fusing a coordinative dimension with a fitness dimension.

$$\text{GDV} = 5 \times \text{distance value GF} + \text{distance value FF} + \text{distance value GP} + \text{distance value SS} - 0.5 \times \text{FI} \times 100$$

FIGURE 5 | Formula to compute the general distance value (GDV). GF, grip force; FF, footrest forces; GP, grip pull-out; SS, sliding seat; FI, force index. Distance value of the grip force is weighted five-fold, other distance values one-way, half of force index is subtracted.

To consider the influence of the subcomponents of the GDV (grip force, footrest forces, grip pull-out, sliding seat, and force index) which had not been normalized before training, for each component data of each of the three treatment groups was normalized to the group mean for pretest. Additionally, a variability coefficient was computed by dividing the standard deviation of mean energy expended during grip pull-out by its mean, to quantify the stability of motion. Also duration of pull-out phase was computed to check the difference of the model's (1s) and the participants' duration of pull-out phase.

RESULTS

Learning Effects

At the beginning of the study, a high variability of participants' rowing techniques was measurable. The average GDV of all participants was at 20.08 ± 10.46 (V: 21.29 ± 9.78 , AV_{nat} : 20.69 ± 10.31 , AV_{soni} : 21.27 ± 11.30) in pretest. In course of the study, participants approximated their rowing technique to the model's technique. The mean GDV of the last training session was 8.65 ± 3.99 (V: 9.38 ± 5.48 , AV_{nat} : 9.96 ± 4.45 , AV_{soni} : 6.62 ± 2.05). The learning curves of the three treatment groups are depicted in **Figure 6**.

A first strong reduction of GDV could be observed from the pretest to the first training session for all participants. The GDV in pretest was compared to the GDV at the first block in the first training session with an ANOVA r.m. revealing a significant main effect "time" [$F_{(1, 45)} = 16.086$, $p < 0.001$, $\eta_p^2 = 0.26$]. Neither main effect "treatment" [$F_{(2, 45)} = 0.145$, $p = 0.865$, $\eta_p^2 = 0.01$] nor interaction "time" \times "treatment" [$F_{(2, 45)} = 0.479$, $p = 0.623$, $\eta_p^2 = 0.02$] became significant.

For the whole training ANOVA r.m. (6*5 data points) on GDV revealed significant main effects on "training session" ("ts," six sessions in 3 weeks) [$F_{(5, 225)} = 33.111$, $p < 0.001$, $\eta_p^2 = 0.42$] and on "block" ("b," five blocks per training session) [$F_{(4, 180)} = 21.151$, $p < 0.001$, $\eta_p^2 = 0.32$].

The course of GDV during training is depicted in **Figure 7**. From training session to training session GDV was reduced. All differences except for the differences between training sessions 2 + 3, 4 + 5, 4 + 6 and 5 + 6 became significant with $p < 0.01$ (Post-Hoc: LSD).

Figure 8 shows the course of GDV averaged about all training sessions. From block to block GDV was reduced. All differences except for the difference between blocks 4 + 5 became significant with $p < 0.05$ (Post-Hoc: LSD).

The learning effect remained stable on retention (three weeks later), no differences between last training and retention became

evident, as well in GDV {ANOVA: [$F_{(1, 45)} = 1.062$, $p = 0.308$, $\eta_p^2 = 0.023$] as in its subcomponents: grip force [$F_{(1, 45)} = 2.129$, $p = 0.151$, $\eta_p^2 = 0.05$], footrest forces [$F_{(1, 45)} = 0.003$, $p = 0.953$, $\eta_p^2 < 0.001$], grip pull-out [$F_{(1, 45)} = 2.549$, $p = 0.117$, $\eta_p^2 = 0.05$], sliding seat [$F_{(1, 45)} = 1.450$, $p = 0.235$, $\eta_p^2 = 0.03$], and force index [$F_{(1, 45)} = 0.148$, $p = 0.702$, $\eta_p^2 < 0.001$].

Group Differences

A scheme about the group differences can be found in **Table 1**. Main effect "treatment" also became significant [$F_{(2, 45)} = 3.571$, $p < 0.05$, $\eta_p^2 = 0.14$]. In **Figure 9** the group differences are depicted. In training, AV_{soni} differed from AV_{nat} ($p = 0.037$) as well as from V ($p = 0.018$). AV_{nat} and V didn't differ ($p = 0.765$). (Post-Hoc: LSD)

Interactions

Interactions with "treatment" didn't become significant {"treatment" \times "training session": [$F_{(10, 225)} = 0.552$, $p = 0.852$, $\eta_p^2 = 0.02$]; "treatment" \times "block": [$F_{(8, 180)} = 0.658$, $p = 0.728$, $\eta_p^2 = 0.03$]; "treatment" \times "training session" \times "block": [$F_{(40, 900)} = 0.674$, $p = 0.940$, $\eta_p^2 = 0.03$], but there was a significant interaction between "training session" and "block" [$F_{(20, 900)} = 6.722$, $p < 0.001$, $\eta_p^2 = 0.13$].

In **Figure 10** the course of GDV within the six training sessions can be compared. The amount of reduction from training block to training block was reduced from training session to training session. In the first training session reduction was clearly notable, in the sixth training session the course was nearly horizontal.

Subcomponents

To explore the background of the reported significant main effects, the single components of the GDV (normalized to the group means for pretest) were regarded. For each of the five components, ANOVA revealed significant main effects for "training session" {grip force: [$F_{(5, 225)} = 21.154$, $p < 0.001$, $\eta_p^2 = 0.32$]; footrest forces: [$F_{(5, 225)} = 12.995$, $p < 0.001$, $\eta_p^2 = 0.22$]; grip pull-out: [$F_{(5, 225)} = 30.851$, $p < 0.001$, $\eta_p^2 = 0.41$]; sliding seat: [$F_{(5, 225)} = 11.088$, $p < 0.001$, $\eta_p^2 = 0.20$]; force index: [$F_{(5, 225)} = 48.813$, $p < 0.001$, $\eta_p^2 = 0.52$]} and for block {grip force: [$F_{(4, 180)} = 15.749$, $p < 0.001$, $\eta_p^2 = 0.26$]; footrest forces: [$F_{(4, 180)} = 8.864$, $p < 0.001$, $\eta_p^2 = 0.16$]; grip pull-out: [$F_{(4, 180)} = 29.992$, $p < 0.001$, $\eta_p^2 = 0.40$]; sliding seat: [$F_{(4, 180)} = 9.589$, $p < 0.001$, $\eta_p^2 = 0.18$]; force index: [$F_{(4, 180)} = 24.629$, $p < 0.001$, $\eta_p^2 = 0.35$]}, as well as for the

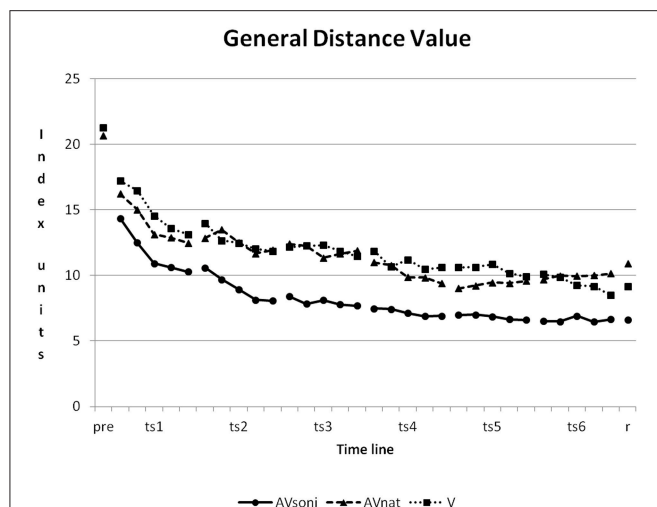


FIGURE 6 | Learning curves, showing the development of group means of GDV from pretest to retention test for all three treatment groups. Pre, pretest; ts, training session; r, retention test; AV_{soni}, treatment group AV_{soni}; AV_{nat}, treatment group AV_{nat}; V, treatment group V. Standard deviations are regarded subsequently.

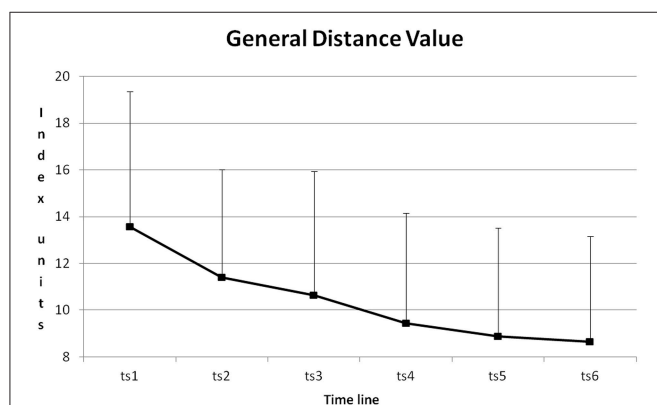


FIGURE 7 | Course of GDV and its standard deviation from training session one to six (ts1–6), averaged for all blocks and all participants.

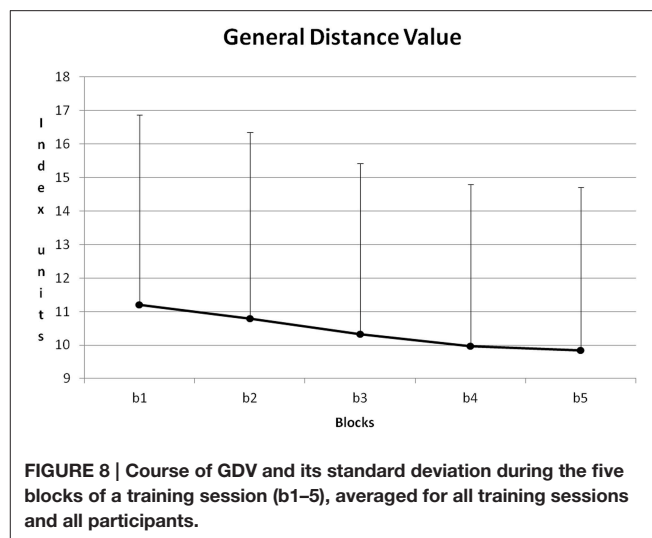


FIGURE 8 | Course of GDV and its standard deviation during the five blocks of a training session (b1–5), averaged for all training sessions and all participants.

TABLE 1 | Group differences.

Component	Prob. of error	Significant Post-Hoc results
General Distance Value	$p < 0.05$	AV _{soni} – AV _{nat} : $p < 0.05$ AV _{soni} – V: $p < 0.05$
Grip pull-out	$p < 0.001$	AV _{soni} – V: $p < 0.001$ AV _{nat} – V: $p < 0.001$
Sliding seat	$p < 0.001$	AV _{soni} – V: $p < 0.001$ AV _{soni} – AV _{nat} : $p < 0.001$
Grip force	$p = 0.096$	—
Footrest forces	$p < 0.001$	AV _{soni} – AV _{nat} : $p < 0.001$ AV _{soni} – V: $p < 0.05$ V – AV _{nat} : $p < 0.001$
Force index	$p = 0.166$	—
Duration of pull-out phase	$p < 0.001$	AV _{soni} – V: $p < 0.01$ AV _{nat} – V: $p < 0.01$
Variability coefficient	$p = 0.181$	—

Probabilities of error (ANOVA, 6*5 data points) for main effect treatment and the corresponding Post-Hoc probabilities of error (LSD) in GDV, its subcomponents normalized for pretest (grip pull-out, sliding seat, grip force, footrest forces, and force index) and the additional components normalized for pretest (duration of pull-out phase, variability coefficient). The former treatment group is in each case the one with the lower value.

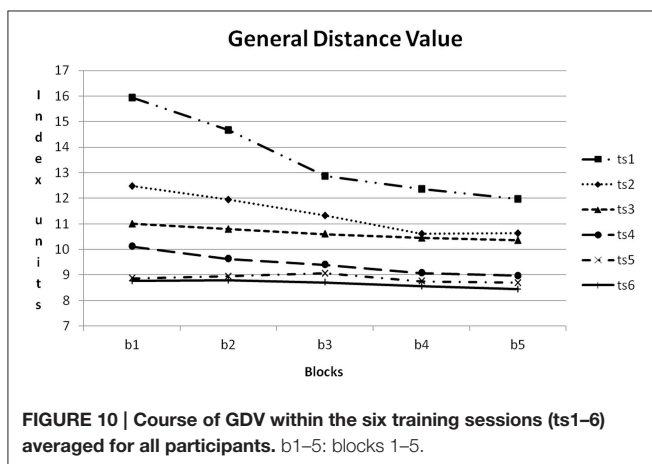
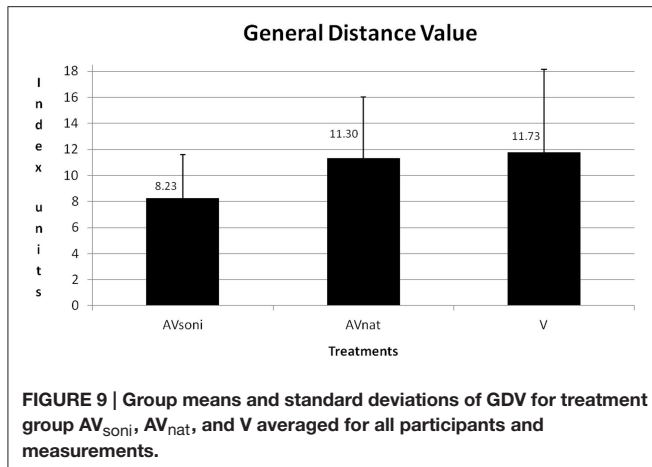
interaction “training session” × “block” {grip force: [$F_{(20, 900)} = 4.682$, $p < 0.001$, $\eta_p^2 = 0.09$]; footrest forces: [$F_{(20, 900)} = 3.519$, $p < 0.001$, $\eta_p^2 = 0.07$]; grip pull-out: [$F_{(20, 900)} = 5.801$, $p < 0.001$, $\eta_p^2 = 0.11$]; sliding seat: [$F_{(20, 900)} = 1.796$, $p < 0.05$, $\eta_p^2 = 0.04$]; force index: [$F_{(20, 900)} = 5.272$, $p < 0.001$, $\eta_p^2 = 0.10$]}, all with the same tendency as GDV.

Main effect “treatment” however differed between the components: for force index [$F_{(2, 45)} = 1.866$, $p = 0.166$, $\eta_p^2 = 0.08$] and grip force [$F_{(2, 45)} = 2.474$, $p = 0.096$, $\eta_p^2 = 0.10$] it didn’t reach significance, for footrest forces [$F_{(2, 45)} = 18.380$, $p < 0.001$ (Bonferroni adjusted p -value), $\eta_p^2 = 0.45$], grip pull-out [$F_{(2, 45)} = 19.453$, $p < 0.001$, $\eta_p^2 = 0.46$] and sliding seat [$F_{(2, 45)} = 23.065$, $p < 0.001$ (Bonferroni adjusted p -value), $\eta_p^2 = 0.51$] it reached level of significance. As we use both “footrest

forces” and “sliding seat” to test hypothesis H3, the Bonferroni-adjusted p -values are computed for these two components.

On footrest forces all three treatment groups differed from each other, AV_{soni} from AV_{nat} ($p < 0.001$), AV_{soni} from V ($p < 0.05$), and AV_{nat} from V ($p < 0.001$) (Post-Hoc: LSD). Footrest forces featured the strongest improvement in AV_{soni} and they featured the lowest improvement in AV_{nat}, as it is depicted in Figure 11.

On grip pull-out V differed from AV_{soni} ($p < 0.001$) as well as from AV_{nat} ($p < 0.001$) (Post-Hoc: LSD) with a worse performance for V and on sliding seat AV_{soni} differed from V ($p < 0.001$) as well as from AV_{nat} ($p < 0.001$) (Post-Hoc: LSD) with a better performance for AV_{soni}.

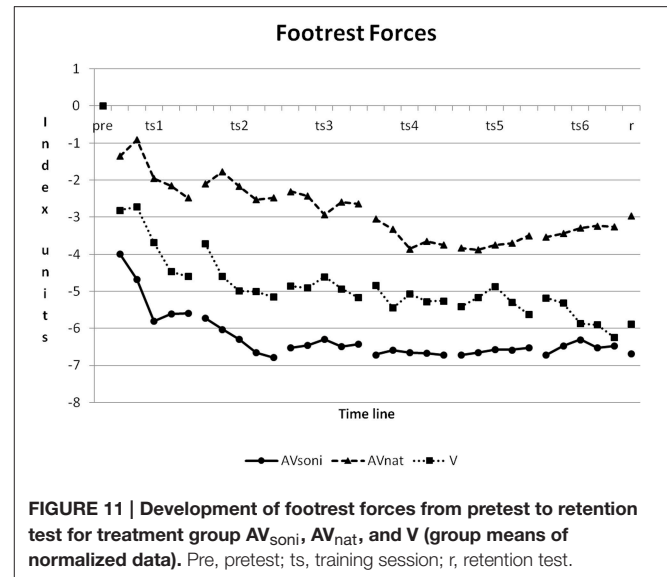


Additional Data

Additionally two components of the movement that were not integrated in the GDV were regarded: duration of pull-out phase and variability coefficient.

Duration of Pull-out Phase

In pretest, duration of pull-out phase had a mean of 1.47 s. This data was reduced in the course of the training to a minimum value of 1.25 s in the last block of the last training session and reached a value of 1.26 s in retention test. (The model's pull-out phase and recovery phase had a time ratio of 2:1 with 1 s for pull-out phase). For convenience, a lower value can be considered as better because only three participants ever reached a value of less than 1 s (the model's value). On duration of the pull-out phase (normalized to the group means for pretest) ANOVA revealed a significant main effect "training session" [$F_{(5, 225)} = 28.387, p < 0.001, \eta_p^2 = 0.39$], "block" [$F_{(4, 180)} = 16.707, p < 0.001, \eta_p^2 = 0.27$], and "treatment" [$F_{(2, 45)} = 6.347, p < 0.01, \eta_p^2 = 0.22$] and a significant interaction "training session" \times "block" [$F_{(20, 900)} = 1.867, p < 0.05, \eta_p^2 = 0.04$]. V differed from AV_{soni} ($p < 0.01$) as well as from AV_{nat} ($p < 0.01$) (*Post-Hoc*: LSD) with a worse performance for V as it is depicted in Figure 12.



Variability Coefficient

On variability coefficient (normalized to the group means for pretest) ANOVA revealed a significant main effect "training session" [$F_{(5, 225)} = 22.842, p < 0.001, \eta_p^2 = 0.34$] and "block" [$F_{(4, 180)} = 8.658, p < 0.001, \eta_p^2 = 0.16$]. Main effect "treatment" [$F_{(2, 45)} = 1.773, p = 0.181, \eta_p^2 = 0.07$] and interaction "training session" \times "block" [$F_{(20, 900)} = 1.504, p = 0.072, \eta_p^2 = 0.03$] didn't reach significance.

Retention Test

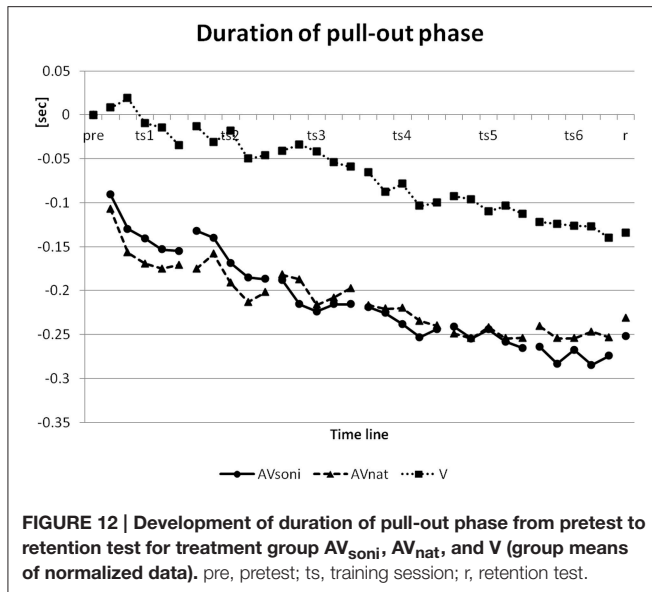
Main effect "treatment" persisted in retention test for GDV {ANOVA: [$F_{(2, 45)} = 3.707, p < 0.05, \eta_p^2 = 0.14$]} and also for the single components sliding seat [$F_{(2, 45)} = 19.875, p < 0.001, \eta_p^2 = 0.47$], footrest forces [$F_{(2, 45)} = 19.984, p < 0.001, \eta_p^2 = 0.47$], and grip pull-out [$F_{(2, 45)} = 10.209, p < 0.001, \eta_p^2 = 0.31$] but not for the components force index [$F_{(2, 45)} = 1.390, p = 0.259, \eta_p^2 = 0.06$] and grip force [$F_{(2, 45)} = 0.231, p = 0.795, \eta_p^2 = 0.01$].

Standard Deviations

As Figure 13 shows, standard deviations developed differently in the three treatment groups. Although group AV_{soni} started with highest value of SD, from training block two in training session two and also in retention test, it had smallest values compared to the other two groups. Group V shows the smallest reduction of SD. Levene's Test for Homogeneity of Variances was significant at 19 and not significant at 11 measuring points.

Force Index and Strength Ability

To assure, that the observed growth of the force index during training is not only determined by an improvement of the maximum strength ability, we controlled maximum strength values before and after training. Sums of maximum strength in leg extension and arm flexion before and after training were

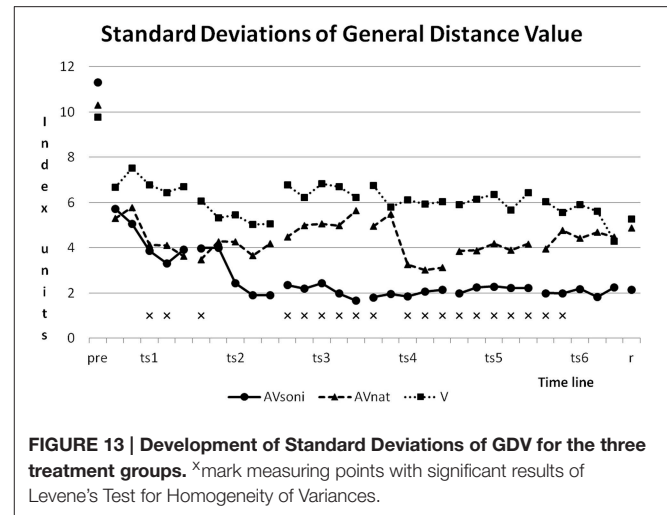


compared. ANOVA revealed no significant difference between the two measurements [$F_{(1, 45)} = 0.707$, $p = 0.405$, $\eta_p^2 = 0.02$]. **Figure 14** shows the changing of the maximum strength in relation to the development of FI. While FI was growing in the course of the training, maximum strength remained nearly unchanged.

DISCUSSION

The present study was conducted to explore the impact of a 4-dimensional movement sonification on the acquisition of the basic technique of indoor-rowing (motor learning). We hypothesized that an extension of visual instruction and feedback with an artificial movement-acoustics would support motor learning. All three experimental groups of male rowing novices observed a video of a rowing professional for instruction and their own movement execution on a projection screen for video feedback, but differed on quantity and quality of additional acoustic movement information. The first group (V) only got visual information. The second group additionally heard the respective natural motion attendant sounds (AV_{nat}) and the third group additionally received the movement sonification (AV_{soni}). For both audiovisual groups (AV_{soni}, AV_{nat}) an enhanced learning performance in terms of a sharper learning curve, indicating an increased approximation to the model's technique, was expected compared to the visual group (V).

For the whole sample consisting of male pupils and male students, measurement of initial technical level revealed large differences between participants. We attribute these differences to the wide age range and broad individual differences on coordinative skill level. Despite the heterogeneity of the sample, ANOVA r.m. for GDV revealed significant main effects "training session," "block" and "treatment". After 3-week training, a large enhancement in technical level from pretest to last training session was found, indicating for all three training modes a

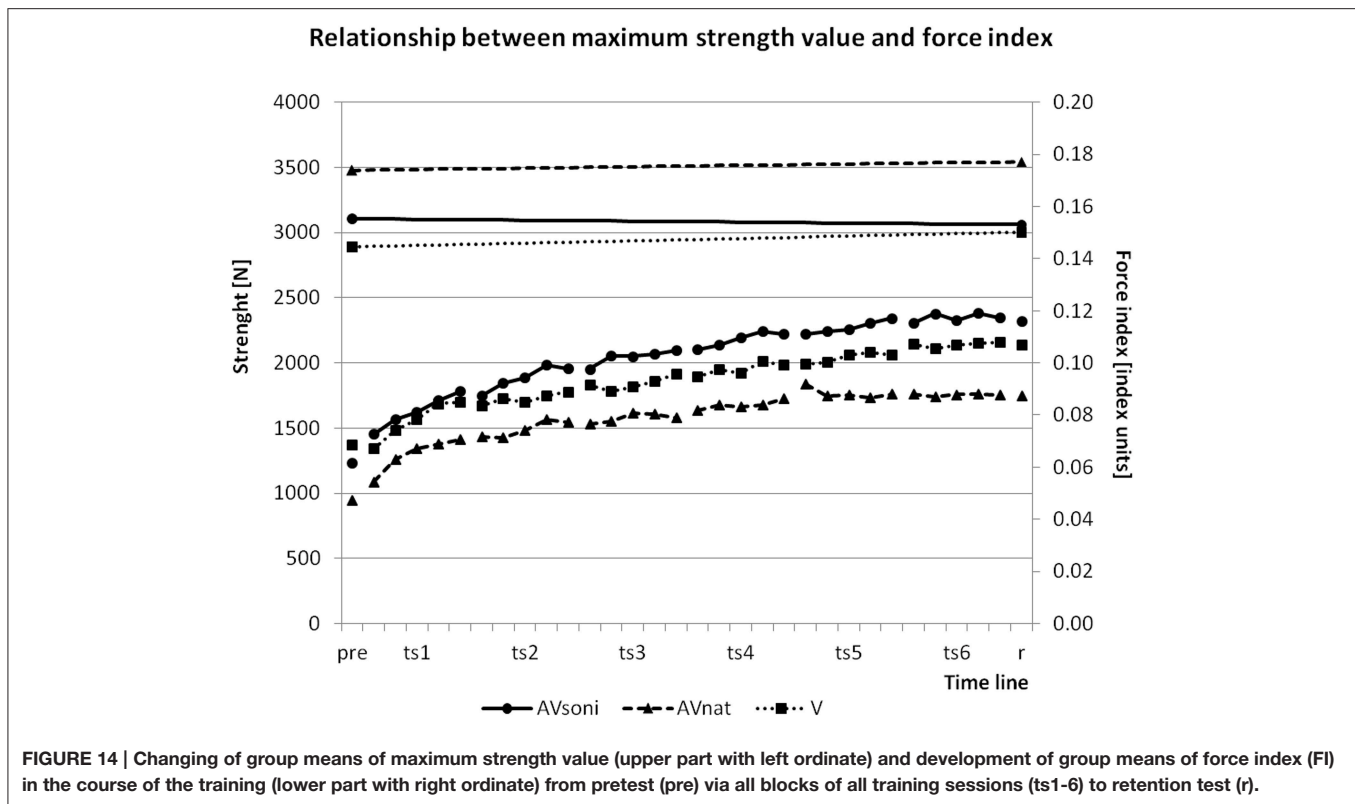


high efficiency on the autonomous acquisition of the rowing technique. Already a remarkable amount of enhancement from pretest to first training session became evident, confirming the novice status of the participants. But besides this, the "jump" of approximation between pretest and first training can be additionally interpreted as an initial effect of the available information on the internal forward dynamic model (Wolpert et al., 2001): Since no feedback was given in the pretest it is plausible, that the instructive information is used for the support of the forward-model. The inverse model should be additionally supported not before feedback was given, that means, not before the middle of the first training block.

Regarding the temporal course, differences between the single training sessions became evident, partly the improvement could be already observed between two consecutive training sessions. In the further study, the learning effect decreases, which might indicate that the realized method was appropriate for the participants. Also within single training sessions an improvement occurred, indicating short-term learning. The amount of the training effect was not constant: Within a single training session the training effect decreased from session to session probably indicating also a short-term learning ceiling effect (see **Figure 10** General Distance Value). Between training session five and six almost no improvement became observable any more, probably indicating a general ceiling effect of the chosen autonomous learning setting. Finally the retention measurement confirmed that the achieved learning effects remained stable and were not restricted to the 3-week-training period.

Subcomponents

The influence of single subcomponents of the movement on the results will be regarded to get additional insights into the development of rowing coordination. Main effects "training" and "block" on GDV can't be attributed to single subcomponents but are reflected in every single subcomponent (grip force, footrest forces, grip pull-out, sliding seat, force index) and also in the additionally regarded movement components, as the duration of



the pull-out and the variability coefficient. We interpret this result as an indication for an extensive support of the development of the technique specific coordination via available information. To explore which kind of treatment affects a certain technical feature to what extent, it will be regarded in a first step, how the different treatments are effective in mediating the basic rhythmic structure. Regarding the basic rhythmic structure of a rowing cycle, the expert rower or the model resp., who was performing ergometer rowing with a frequency of about 20 cycles/min resulting in a duration of about 3.0 s for each cycle, realized a basic phase structure of about 1:2, meaning the pull-out phase was about 1 s on average and the recovery phase was about 2 s (pull-out phase: Starting with the local minimum till the local maximum of the pull-out length is reached). The initial phase structure of the total sample showed a nearly 1:1 relation with about 1.47 s on average (initial values: 1.47 s AV_{soni}, 1.53 s AV_{nat}, 1.41 s V) for pull-out phase, meaning that novices took too much time for the pull-out and might not have been aware of the basic rhythmic structure of the demonstrated rowing motion.

The temporal development of this basic rhythmic movement structure over the training period is shown in **Figure 12** (Duration of pull-out phase) based on the normalized samples on the group mean value of the pretest. A clear decrease of pull-out phase duration became evident, resulting in about 1.25 sec on average (final values: 1.19 s AV_{soni}, 1.28 s AV_{nat}, 1.28 s V) for pull-out phase in the last training block. Furthermore, main effect “treatment” revealed differences between groups with lowest

decrease for V compared to both other groups AV_{soni} and AV_{nat}. The mean temporal decrease of the pull-out phase duration of all training measures was 0.21 s for AV_{soni}, 0.21 s also for AV_{nat} and 0.07 s for V which can be interpreted as a clear auditory or audiovisual benefit on the mediation of the basic movement rhythm compared to a merely visual condition. These findings do confirm hypothesis H2 (Audiovisually treated groups are more precise in rhythmic demands of the indoor rowing technique.).

Furthermore, the findings closely correspond with the ANOVA results concerning the course of the grip pull-out, exemplifying an enhanced approximation of both audiovisual groups to the model compared to the visual group. Natural motion attendant sounds as well as movement sonification in combination with visual movement information seem to provide rhythmic information, that is not available to this extend in the pure visual condition. Taken together, the findings on the duration of the pull-out phase as well as on the course of the grip pull-out clearly support a higher efficiency of an audiovisually based motor learning compared to a merely visual setting, but it cannot explain, on the other hand, the observed group differences of the technique acquisition between AV_{soni} and AV_{nat}.

Main Effect “Treatment”

When subsequently focusing on the technique acquisition as a whole by regarding the GDV, main effect “treatment” became significant. *Post-hoc* analysis indicated a better learning performance of the AV_{soni} group compared to both other groups

(AV_{nat} , V), as already described in the “Results”-section. This result pattern is not reflected in the previously regarded technical components “duration of grip pull-out” and “course of grip pull-out,” but in the technical components “sliding seat” and “footrest forces”: For both components the *post-hoc* analysis revealed a superiority for the AV_{soni} group compared to both other groups (AV_{nat} , V). Whereas the duration and the course of the grip pull-out do not directly refer to the leg motion and the transmission of the force from the footrest to the grip resp., “footrest forces” and “sliding seat” are both features explicitly referring to the kinematic chain “legs-trunk-arm” or to whole body coordination resp., which is the key feature of the rowing technique. The existence of a quite close binding of auditory and motor sequences has just been described by Rauschecker associated with the auditory dorsal stream (premotor-basal-ganglia circuits together with higher auditory centers), which “transforms musical into motor sequence information and vice versa, realizing what are known as forward and inverse models. The basal ganglia and the cerebellum are involved in setting up the sensorimotor associations, translating timing information into spatial codes and back again.” (2014, 1). Hypothesis H3 (Participants treated with complex sonification benefit in terms of a better coordination of the movement resulting in higher technical performance.) is confirmed based on this interpretation of the reported findings (Rauschecker, 2014).

Even though these observation might work as an explanation for parts of the learning effects observed in our study, it has not cleared yet what acoustic or musical features are related to which features of motor execution—which would be indeed a valuable framework for the designing of efficient mapping patterns of movement sonifications. Nevertheless, the findings of Rauschecker (2014) mentioned above can be interpreted to mean that complex musical structures can easily be learned and remembered, carrying a huge amount of sensorimotor information and are usually closely linked to inverse internal models. If a sonification is designed in a consistent and complex mode, as it was realized here in first steps with a mapping of amplitude, frequency and timbre to four certain discrete and continuous movement features, the emerging 4-dimensional movement sonification exhibits quasi-musical features like tempo and rhythm and also some simple melody. This complex and quasi-musical character of the sonification might be an explanation for the observed effects surpassing the effects of rhythmic adjustments.

This kind of complex information is visually obviously not available with comparable precision and it is also not included in the auditory information that natural motion attendant sounds provide. Obviously, a larger informational content is decoded from complex movement sonification compared to natural or more reduced forms of movement acoustics. This assumption is supported by two additional findings: (1) Group differences are also preserved in retention test 3 weeks after last training session, indicating an outlasting learning advantage for participants treated with additional sonification. (2) Interestingly, for the AV_{soni} group the approximation to the model’s technique during training seems to be accompanied by a reduction of interindividual heterogeneity, which seems to be given for the

AV_{soni} group as illustrated in **Figure 13** (Standard deviations of General Distance Value).

An alternative explanation could have been that participants did not improve their rowing technique but instead their strength capacity and thus were able to approximate increasingly to the model’s technique. But because maximum strength values did not change from pre- to posttest, the increase of used force (FI) can’t be attributed to an increase of the participants’ maximum strength capacities. A more plausible explanation would be that an improvement of a force demanding movement technique can be understood as a better utilization of already existing strength capacities by a smoother and more economical coordination when executing the technique, as illustrated in **Figure 14** (Relationship between maximum force value and force index).

To control if the expected differences between the merely visually treated group and the audiovisual sonification group can be explained by the addition of a further sense alone or by the specific shape of the sonification, we created an audiovisual control group with another 16 participants, who heard the natural motion attendant sounds besides seeing the video as instruction and feedback. This group also showed a worse learning performance compared to the audiovisually treated sonification group in terms of a larger technical distance to the model over the whole course of the study. This finding is somewhat surprising at a first sight, because it became evident (see section Non-Musical Acoustic Information on Motor Behavior) that also natural movement sounds carry a lot of information which can be decoded by the auditory system easily and used to enable or modulate movement perception as well as motor control. So how to explain the measured differences? Given that natural motion attendant sounds are integrated presumably not worse with visual information in humans than sonification is, sonification apparently provides more information about the movement than natural motion attendant sounds do. The latter consisted of the sound of the rowing ergometer flywheel and the sliding seat. In comparison to sonification, these sounds also contain information about the grip force (flywheel sound), the movement of the sliding seat and the temporal relation between both. But there was no acoustic information about the footrest force nor about the grip pull-out for this group, which might explain the reduced amount of information extractable from the natural motion attendant sounds. Hypothesis H1 (Participants of both groups treated with convergent audiovisual information show better learning results in terms of a steeper learning curve/a faster approximation to the model technique.) cannot be confirmed based on these findings.

Neurophysiological Framework

In summary, movement sonification in combination with video provides information that is neither available visually (video alone) nor with video combined with natural motion attendant sounds. For the alignment of one’s own action to a visual, auditory or audiovisual model action, the mirror neuron system is important. In our study, in that the rowing ergometer is tuned into a sound instrument, participants of the sonification group were able to perceive rowing action and sonification

for altogether 45 min synchronously. As the selected movement features were coupled to certain acoustic features in a fixed mode, it is plausible that movement specific audio-motor co-activation patterns emerge in the CNS: Such audio-motor co-activation networks can emerge even within about first 20 min of practice, as shown by Bangert and Altenmüller (2003) on music novices. The benefit of movement sonification on motor perception, re-enactment (Effenberg, 2005; Young et al., 2013) and synchronization (Schaffert et al., 2011) was established before. A motor learning study of Sigrist et al. (2015)—enhancing the horizontal angle of the rowing oar during the recovery phase in indoor rowing—failed to generate a long-lasting learning effect with error-feedback. In contrast to the use of sonification as error-feedback in rowing, usually requiring conscious cognitive processing, here a complex movement sonification is created to enhance and accelerate the emergence of adequate internal representations of the new movement technique. This should be achieved by a movement sonification which is comprehensively integrable with perceptual streams of other modalities as visual and kinesthetic as the most important ones. This way, the movement sonification should have supported the emergence of the forward model in a first step when used in the instructive mode. But though the rowing model as well as the participants' own rowing motion were sonified in an equivalent manner (all parameter sets were normalized before post-processing), also the inverse model should have been supported when sonification was used as additional real-time feedback during movement execution. These mechanisms might have been also responsible for the effects of real-time movement sonification we observed recently for the acquisition of character handwriting on children (Effenberg et al., 2015).

Even though multisensory integration efficiency of the generated audiovisual stimuli was not investigated in this behavioral study, former fMRI-research of our workgroup already confirmed the integration efficiency of such type of intermodal convergently shaped movement sonification based on dynamic (Scheef et al., 2009) and kinematic (Schmitz et al., 2013) movement parameters. Here, evidence was presented that learning of a complex gross motor movement can be improved by this type of movement information. Explicit knowledge about the mapping doesn't seem to be required for using this information on motor learning. It was shown that the effects were surpassing effects of rhythmic adaptation and that they were long-lasting. Additionally, there are some indications that the efficiency might be independent from directing conscious attention to it, which would be in line with the idea that at least bottom-up proportions of multisensory integration are dependent primarily on temporal and spatial stimulus convergence combined with structural analogy demands (content-related congruity) on perceived stimuli (Calvert et al., 2001). Intermodal convergent sonification seems to have an implicit informational effect, emerging from integration with visual and kinesthetic movement information, when matching to the observed movement just as natural motion attendant sounds do match.

Since movement sonification was configured with high-degree of convergence to visual percept, it could be suggested that beside audiomotor mechanisms, audio-visuo-motor mechanisms

were involved in copying the model's rowing technique. Kaplan and Iacoboni (2007) found a region in ventral premotor cortex that shows a specific response on the conjunction of visual and auditory action-related stimuli and thus might produce a modality-independent representation of the action. Since the ventral premotor cortex is involved in action planning, the findings indicate that the investigated region contributes to the representation of actions, independent of agency and sensory modality. Hence, although there is no direct linkage between the movement and the sonification, the visual event as mediating link might facilitate the establishment of such a linkage in the brain. This could be an explanation for the fact that a learning advantage of the sonification group occurred already in the first training session, that means after 60 s of exposure to sonification (30 s instruction and 30 s feedback).

Conclusion

Finally, it should be emphasized that the introduced method of intermodal convergent real-time movement sonification should be also adaptable to motor rehabilitation, such as for stroke rehabilitation (hemiparesis) or gait rehabilitation after endoprosthesis. If effectiveness of the method is at least partially based on direct multimodal integration as described here, it might work below the level of consciousness. For this reason, it should be also effective especially on patients with certain sensorimotor restrictions as in Parkinson's disease (Thaut et al., 1996) e.g., recently theoretically underpinned by Murgia et al. (2015) with strong references to "rhythmic auditory stimulation" established by Thaut et al. (1997)—and even beyond rhythmic adjustments. Additional empirical work of Young et al. (2014) indicates the efficiency of auditory step models on the gait of Parkinson patients. But multidimensional kinds of real-time movement sonification containing continuously mapped parameters even exceed rhythmic adjustments by addressing kinematic chains or whole body coordination, as discussed in section "Main effect Treatment." Besides the findings presented here, there is further initial evidence from our workgroup on motor rehabilitation of the upper limb on hemiparesis patients (Schmitz et al., 2014). In the future, it should be possible to further improve the efficiency of established methods by adding real-time movement sonification as described in this paper. What has not been proven yet is the effectiveness of auditory movement information alone—how movement sonification might support motor learning as a substitution of visual information. Further research should be directed to such questions as well as to different fields of motor rehabilitation, currently only sparsely supported by initial indications of intermodal information processing.

AUTHOR CONTRIBUTIONS

AE created the topic, developed the method and was concerned with the data collection and the statistical analysis of the data. He did write broad parts of the paper and was the responsible author together with UF. UF was supporting when creating the topic, she was collecting the data and was computing the statistical analysis of the data together with AE and GS. She did

also write broad parts of the paper and was the responsible co-author together with AE. Both authors (AE and UF) contributed equally to this work. GS was supporting the statistical analysis of the data. He gave significant input to some passages of the text. BK was developing parts of the sonification-system used in the study as well as the dtw-algorithm. HM was the supervisor of the project. He was supporting the development of the whole method as well as the organization of the data collection.

REFERENCES

- Abernethy, B. (2013). *Biophysical Foundations of Human Movement*. Champaign, IL: Human Kinetics.
- Agostini, T., Righi, G., Galmonte, A., and Bruno, P. (2004). "The relevance of auditory information in optimizing hammer throwers performance," in *Biomechanics and Sports. XI Winter Universiads 2003*, ed P. B. Pascolo (Wien; New York, NY: Springer), 67–74.
- Bangert, M., and Altenmüller, E. O. (2003). Mapping perception to action in piano practice: a longitudinal DC-EEG study. [Research Article]. *BMC Neurosci.* 4:26. doi: 10.1186/1471-2202-4-26
- Bangert, M., Peschel, T., Schlaug, G., Rotte, M., Drescher, D., Hinrichs, H., et al. (2006). Shared networks for auditory and motor processing in professional pianists: evidence for fMRI conjunction. *Neuroimage* 30, 917–926. doi: 10.1016/j.neuroimage.2005.10.044
- Barracough, N. E., Xiao, D., Baker, C. I., Oram, M. W., and Perrett, D. I. (2005). Integration of visual and auditory information by superior temporal sulcus neurons responsive to the sight of actions. *J. Cogn. Neurosci.* 17, 377–391. doi: 10.1162/0898929053279586
- Baudry, L., Leroy, D., Thouvenot, R., and Chollet, D. (2006). Auditory concurrent feedback benefits on the circle performed in gymnastics. *J. Sports Sci.* 24, 149–156. doi: 10.1080/02640410500130979
- Baumann, O., and Greenlee, M. W. (2006). Neural correlates of coherent audiovisual motion perception. *Cerebral Cortex*, 17, 1433–1443. doi: 10.1093/cercor/bhl055
- Beauchamp, M. S. (2005). See me, hear me, touch me: multisensory integration in lateral occipital-temporal cortex. *Curr. Opin. Neurobiol.* 15, 145–153. doi: 10.1016/j.conb.2005.03.011
- Bidet-Caulet, A., Voisin, J., Bertrand, O., and Fonlupt, P. (2005). Listening to a walking human activates the temporal biological motion area. *Neuroimage* 28, 132–139. doi: 10.1016/j.neuroimage.2005.06.018
- Calvert, G. A., Hansen, P. C., Iversen, S. D., and Brammer, M. J. (2001). Detection of audio-visual integration sites in humans by application of electrophysiological criteria to the BOLD effect. *Neuroimage* 14, 427–438. doi: 10.1006/nimg.2001.0812
- Calvert, G. A., Spence, C., and Stein, B. E. (Eds.). (2004). *The Handbook of Multisensory Processes*. Cambridge, MA: MIT Press.
- Calvert, G. A., and Thesen, T. (2004). Multisensory integration: methodological approaches and emerging principles in the human brain. *J. Physiol.* 98, 191–205. doi: 10.1016/j.jphysparis.2004.03.018
- Chiari, L., Dozza, M., Cappello, A., Horak, F. B., Macellari, V., and Giansanti, D. (2005). Audio-Biofeedback for Balance improvements: An Accelerometry-Based System. *IEEE Eng. Med. Biol. Mag.* 52, 2108–2111. doi: 10.1109/TBME.2005.857673
- Danna, J., Fontaine, M., Paz-Villagran, V., Gondre, C., Thoret, E., Aramaki, M., et al. (2015). The effect of real-time auditory feedback on learning new characters. *Hum. Mov. Sci.* 43, 216–288. doi: 10.1016/j.humov.2014.12.002
- Danna, J., Velay, J. L., Vietnam, V., Capel, A., Petroz, C., Gondre, C., et al. (2013). "Handwriting movement sonification for the rehabilitation of dysgraphia," in *Paper presented at the Sound, Music & Motion Proceedings of the 10th International Symposium on Computer Music Multidisciplinary Research (CMMR)* (Marseille), 1–9.
- Demuth, B., Röder, T., Müller, M., and Eberhardt, B. (2006). "An information retrieval system for motion capture data," in *Advances in Information Retrieval. ECIR 2006: 28th European Conference on IR Research*, eds M. Lalmas, A. MacFarlane, S. Rüger, A. Tombros, T. Tsikrika, and A. Yavlinsky (London, UK; Berlin; Heidelberg: Springer), 373–384.
- Effenberg, A. O. (1996). *Sonification – Ein akustisches Informationskonzept zur Menschlichen Bewegung*. [Sonification – A concept of acoustic information on human motion]. Hofmann, MN: Schorndorf.
- Effenberg, A. O. (2005). Movement sonification: effects on perception and action. *IEEE Multimedia Special Issue Interactive Sonification* 12, 53–59. doi: 10.1109/mmul.2005.31
- Effenberg, A. O., Schmitz, G., Baumann, F., Rosenhahn, B., and Kroeger, D. (2015). SoundScript-supporting the acquisition of character writing by multisensory integration. *Open Psychol. J.* 8, 230–237. doi: 10.2174/1874350101508010230
- Ernst, M. O., and Bühlhoff, H. H. (2004). Merging the senses into a robust percept. *Trends Cogn. Sci.* 8, 162–169. doi: 10.1016/j.tics.2004.02.002
- Forbes, K., and Fiume, E. (2005). "An efficient search algorithm for motion data using weighted PCA," in *Paper Presented at the Proceedings of the 2005 ACM SIGGRAPH/Eurographics Symposium on Computer Animation* (New York, NY: ACM), 67–76.
- Frassinetti, F., Bolognini, N., and Làdavas, E. (2002). Enhancement of visual perception by crossmodal visuo-auditory interaction. *Exp. Brain Res.* 147, 332–343. doi: 10.1007/s00221-002-1262-y
- Giard, M. H., and Peronnet, F. (1999). Auditory-visual integration during multimodal object recognition in humans: a behavioral and electrophysiological study. *J. Cogn. Neurosci.* 11, 473–490. doi: 10.1162/089892999563544
- Haslinger, B., Erhard, P., Altenmueller, E., Schroeder, U., Boecker, H., and Ceballos-Baumann, A. (2005). Transmodal sensorimotor networks during action observation in professional pianists. *J. Cogn. Neurosci.* 17, 282–293. doi: 10.1162/0898929053124893
- Hauelsen, J., and Knoesche, T. R. (2001). Involuntary motor activity in pianists evoked by music perception. *J. Cogn. Neurosci.* 13, 786–792. doi: 10.1162/08989290152541449
- Kaplan, J. T., and Iacoboni, M. (2007). Multimodal action representation in human left ventral premotor cortex. *Cogn. Process.* 8, 103–113. doi: 10.1007/s10339-007-0165-z
- Kennedy, D. M., Boyle, J. B., and Shea, C. H. (2013). The role of auditory and visual models in the production of bimanual tapping patterns. *Exp. Brain Res.* 224, 507–518. doi: 10.1007/s00221-012-3326-y
- Kennel, C., Hohmann, T., and Raab, M. (2014). Action perception via auditory information: Agent identification and discrimination with complex movement sounds. *J. Cogn. Psychol.* 26, 157–165. doi: 10.1080/20445911.2013.869226
- Keyers, C., Kohler, E., Umiltà, M. A., Nanetti, L., Fogassi, L., and Gallese, V. (2003). Audiovisual mirror neurons and action recognition. *Exp. Brain Res.* 153, 628–636. doi: 10.1007/s00221-003-1603-5
- Kohler, E., Keyers, 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
- Lahav, A., Saltzman, E., and Schlaug, G. (2007). Action representation of sound: audiomotor recognition network while listening to newly acquired actions. *J. Neurosci.* 27, 308–314. doi: 10.1523/JNEUROSCI.4822-06.2007

ACKNOWLEDGMENTS

We acknowledge support by European Commission H2020-FETPROACT-2014 No. 641321. Special thanks to Klaus Mattes (University of Hamburg) for realizing and providing the rowing model and validating the participants' movement quality. We also thank Michaela Girgenrath, Annette Rudolf, and Hannah Steingrebe for supporting preparation and realization of the learning intervention.

- Mendonca, C., Santos, J. A., and Lopez-Molnir, J. (2011). The benefit of multisensory integration with biological motion signals. *Exp. Brain Res.* 213, 185–192. doi: 10.1007/s00221-011-2620-4
- Müller, M. (2007). *Information Retrieval for Music and Motion*. Berlin: Springer.
- Murgia, M., Corona, F., Pili, R., Sors, F., Agostini, T., Casula, C., et al. (2015). Rhythmic Auditory Stimulation (RAS) and motor rehabilitation in Parkinson's disease: new frontiers in assessment and intervention protocols. *Open Psychol. J.* 8, 220–229. doi: 10.2174/1874350101508010220
- 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. doi: 10.1068/p7227
- Murgia, M., Santoro, I., Tamburini, G., Prpic, V., Sors, F., Galmonte, A., et al. (2016). Ecological sounds affect breath duration more than artificial sounds. *Psychol. Res.* 80, 76–81. doi: 10.1007/s00426-015-0647-z
- Newell, K. M., and Corcos, D. M. (1993). *Variability and Motor Control*. Champaign, IL: Human Kinetics.
- Rabiner, L., and Juang, B. H. (1993). *Fundamentals of Speech Recognition*. Upper Saddle River, NJ: PTR Prentice Hall.
- Rath, M., and Rocchesso, D. (2005). Continuous sonic feedback from a rolling ball. *IEEE Multimedia, Special Issue Interactive Sonification* 12, 60–69. doi: 10.1109/mmul.2005.24
- Rauschecker, J. P. (2014). Is there a tape recorder in your head? How the brain stores and retrieves musical melodies. *Front. Syst. Neurosci.* 8:149. doi: 10.3389/fnsys.2014.00149
- Rowland, B. A., and Stein, B. E. (2008). Temporal profiles of response enhancement in multisensory integration [Focused Review]. *Front. Neurosci.* 2, 218–224. doi: 10.3389/neuro.01.033.2008
- Schaffert, N., Mattes, K., Barrass, S., and Effenberg, A. O. (2011). "Exploring function and aesthetics in sonifications for elite sports," in *Human Communication Science: A Compendium*, eds R. Dale, D. Burnham, and C. J. Stevens (Sydney: ARC Research Network in Human Communication Science), 465–472.
- Scheef, L., Boecker, H., Daamen, M., Fehse, U., Landsberg, M. W., Granath, D. O., et al. (2009). Multimodal motion processing in area V5/MT: evidence from an artificial class of audio-visual events. *Brain Res.* 1252, 94–104. doi: 10.1016/j.brainres.2008.10.067
- Schmidt, R. A., and Lee, T. (2005). *Motor Control and Learning - A Behavioral Emphasis*. Champaign, IL: Human Kinetics.
- Schmitz, G., and Effenberg, A. O. (2012). "Perceptual effects of auditory information about own and other movements," in *Paper Presented at the 18th International Conference on Auditory Display* (Atlanta, GA), 89–94.
- Schmitz, G., Kroeger, D., and Effenberg, A. O. (2014). "A mobile sonification system for stroke rehabilitation," in *Paper presented at the The 20th International Conference on Auditory Display* (New York, NY), 1–7.
- Schmitz, G., Mohammadi, B., Hammer, A., Heldmann, M., Samii, A., Münte, T. F., et al. (2013). Observation of sonified movements engages a basal ganglia frontocortical network. *BMC Neurosci.* 14:32. doi: 10.1186/1471-2202-14-32
- Seitz, A. R., Kim, R., and Shams, L. (2006). Sound facilitates visual learning. *Curr. Biol.* 16, 1422–1427. doi: 10.1016/j.cub.2006.05.048
- Shams, L., and Seitz, A. R. (2008). Benefits of multisensory learning. *Trends Cogn. Sci.* 12, 411–417. doi: 10.1016/j.tics.2008.07.006
- Shams, L., Wozny, D. R., Kim, R., and Seitz, A. (2011). Influences of multisensory experience on subsequent unisensory processing. *Front. Psychol.* 2:264. doi: 10.3389/fpsyg.2011.00264
- Shea, C. H., Wulf, G., Park, J. H., and Gaunt, B. (2001). Effects of an auditory model on the learning of relative and absolute timing. *J. Mot. Behav.* 33, 127–138. doi: 10.1080/00222890109603145
- Sigrist, R., Rauter, G., Marchal-Crespo, L., Riener, R., and Wolf, P. (2015). Sonification and haptic feedback in addition to visual feedback enhances complex motor task learning. *Exp. Brain Res.* 233, 909–925. doi: 10.1007/s00221-014-4167-7
- Sigrist, R., Rauter, G., Riener, R., and Wolf, P. (2013). Augmented visual, auditory, haptic and multimodal feedback in motor learning: a review. *Psychon. Bull. Rev.* 20, 21–53. doi: 10.3758/s13423-012-0333-8
- Sors, F., Murgia, M., Santoro, I., and Agostini, T. (2015). Audio-based interventions in sport. *Open Psychol. J.* 8, 212–219. doi: 10.2174/1874350101508010212
- Soto-Faraco, S., Kingstone, A., and Spence, C. (2003). Multisensory contributions to the perception of motion. *Neuropsychologia* 41, 1847–1862. doi: 10.1016/S0028-3932(03)00185-4
- Stein, B. E., and Meredith, M. A. (1993). *The Merging of the Senses*. Cambridge: MIT Press.
- Takeuchi, T. (1993). Auditory Information In Playing Tennis. *Percept. Mot. Skills* 76, 1323–1328. doi: 10.2466/pms.1993.76.3c.1323
- Thaut, M. H., McIntosh, G. C., and Rice, R. R. (1997). Rhythmic facilitation of gait training in hemiparetic stroke rehabilitation. *J. Neurol. Sci.* 151, 207–121. doi: 10.1016/S0022-510X(97)00146-9
- Thaut, M. H., McIntosh, G. C., 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
- Vroomen, J., and de Gelder, B. (2000). Sound enhances visual perception: cross-modal effects of auditory organization on vision. *J. Exp. Psychol. Hum. Percept. Perform.* 26, 1583–1590. doi: 10.1037/0096-1523.26.5.1583
- Williams, A. M., Davids, K., and Williams, J. G. (1999). *Visual Perception and Action in Sport*. London; New York, NY: Taylor & Francis.
- Williams, A. M., Hodges, N., and Scott, M. (2004). *Skill Acquisition in Sport*. London; New York, NY: Routledge.
- Wolpert, D. M., Diedrichsen, J., and Flanagan, J. R. (2011). Principles of sensorimotor learning. *Nat. Rev. Neurosci.* 12, 739–751. doi: 10.1038/nrn3112
- Wolpert, D. M., Ghahramani, Z., and Flanagan, J. R. (2001). Perspectives and problems in motor learning. *Trends Cogn. Sci.* 5, 487–494. doi: 10.1016/S1364-6613(00)01773-3
- Wolpert, D. M., Ghahramani, Z., and Jordan, M. I. (1995). An internal model for sensorimotor integration. *Science* 269, 1880–1882. doi: 10.1126/science.7569931
- Woods, E. A., Hernandez, A. E., Wagner, V. E., and Beilock, S. L. (2014). Expert athletes activate somatosensory and motor planning regions of the brain when passively listening to familiar sports sounds. *Brain Cogn.* 87, 122–133. doi: 10.1016/j.bandc.2014.03.007
- Young, W., Rodger, M., and Craig, C. M. (2013). Perceiving and reenacting spatiotemporal characteristics of walking sounds. *J. Exp. Psychol. Hum. Percept. Perform.* 39, 464. doi: 10.1037/a0029402
- Young, W. R., Rodger, M. W., and Craig, C. M. (2014). Auditory observation of stepping actions can cue both spatial and temporal components of gait in Parkinson's disease patients. *Neuropsychologia* 57, 140–153. doi: 10.1016/j.neuropsychologia.2014.03.009

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.

Copyright © 2016 Effenberg, Fehse, Schmitz, Krueger and Mechling. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



More Feedback Is Better than Less: Learning a Novel Upper Limb Joint Coordination Pattern with Augmented Auditory Feedback

Shinya Fujii^{1,2}, Tea Lulic^{1,3} and Joyce L. Chen^{1,4*}

¹ Canadian Partnership for Stroke Recovery, Sunnybrook Research Institute, Toronto, ON, Canada, ² Graduate School of Education, The University of Tokyo, Tokyo, Japan, ³ Department of Kinesiology, McMaster University, Hamilton, ON, Canada, ⁴ Department of Physical Therapy and Rehabilitation Sciences Institute, University of Toronto, Toronto, ON, Canada

OPEN ACCESS

Edited by:

Diego Minicacchi,
University of Florence, Italy

Reviewed by:

Alireza Mousavi,
Brunel University, UK
Malte Schilling,
Bielefeld University, Germany

*Correspondence:

Joyce L. Chen
j.chen@sri.utoronto.ca

Specialty section:

This article was submitted to
Neuroprosthetics,
a section of the journal
Frontiers in Neuroscience

Received: 15 December 2015

Accepted: 20 May 2016

Published: 06 June 2016

Citation:

Fujii S, Lulic T and Chen JL (2016)
More Feedback Is Better than Less:
Learning a Novel Upper Limb Joint
Coordination Pattern with Augmented
Auditory Feedback.
Front. Neurosci. 10:251.
doi: 10.3389/fnins.2016.00251

Motor learning is a process whereby the acquisition of new skills occurs with practice, and can be influenced by the provision of feedback. An important question is what frequency of feedback facilitates motor learning. The guidance hypothesis assumes that the provision of less augmented feedback is better than more because a learner can use his/her own inherent feedback. However, it is unclear whether this hypothesis holds true for all types of augmented feedback, including for example sonified information about performance. Thus, we aimed to test what frequency of augmented sonified feedback facilitates the motor learning of a novel joint coordination pattern. Twenty healthy volunteers first reached to a target with their arm (baseline phase). We manipulated this baseline kinematic data for each individual to create a novel target joint coordination pattern. Participants then practiced to learn the novel target joint coordination pattern, receiving either feedback on every trial i.e., 100% feedback ($n = 10$), or every other trial, i.e., 50% feedback ($n = 10$; acquisition phase). We created a sonification system to provide the feedback. This feedback was a pure tone that varied in intensity in proportion to the error of the performed joint coordination relative to the target pattern. Thus, the auditory feedback contained information about performance in real-time (i.e., “concurrent, knowledge of performance feedback”). Participants performed the novel joint coordination pattern with no-feedback immediately after the acquisition phase (immediate retention phase), and on the next day (delayed retention phase). The root-mean squared error (RMSE) and variable error (VE) of joint coordination were significantly reduced during the acquisition phase in both 100 and 50% feedback groups. There was no significant difference in VE between the groups at immediate and delayed retention phases. However, at both these retention phases, the 100% feedback group showed significantly smaller RMSE than the 50% group. Thus, contrary to the guidance hypothesis, our findings suggest that the provision of more, concurrent knowledge of performance auditory feedback during the acquisition of a novel joint coordination pattern, may result in better skill retention.

Keywords: motor learning, augmented feedback, guidance hypothesis, auditory feedback, sonification

INTRODUCTION

Different joint coordination patterns can be used to achieve one specific motor action (Bernstein, 1967). For example, a person can reach for a cup of coffee on a table in front of them by extending the elbow and flexing the shoulder without moving the trunk. The same action can also be achieved by flexing the trunk with minimal or even no movement of the elbow and shoulder. Although in both cases, the goal of the movement is achieved, the biomechanical efficiency of the movements differs depending on how the joints are coordinated (Hirashima, 2011 for a review). Organized joint coordination patterns allow a person to use muscles efficiently and to prevent muscle fatigue (Furuya et al., 2009). Thus, achieving biomechanically and physiologically efficient movements requires the learning and execution of organized joint coordination patterns. The redundancy of joint coordination patterns can be an issue in motor rehabilitation as the restoration of normal joint coordination patterns is often a challenge for people with movement disorders (Levin, 1996; Cirstea and Levin, 2007). For example, individuals with stroke often implement compensatory strategies (i.e., reaching for a cup of coffee by bending forward with the trunk) given their impaired motor control. Although the movement goal is achieved, use of compensatory strategies may result in longer-term problems such as pain, discomfort, and joint contractures (Levin, 1996; Cirstea and Levin, 2007). Thus, improving movement quality through the re-learning of organized joint coordination patterns is of importance for people with movement disorders.

One strategy that can facilitate the motor (re)learning of organized joint coordination patterns is the use of *augmented feedback*. Augmented feedback is external information provided about the movement that is supplemental to *inherent feedback* (Schmidt and Lee, 2011). Inherent feedback is intrinsic sensory information that is naturally available to an individual during the movement (e.g., vision or proprioception of limbs). The provision of augmented feedback may be relevant when an individual is learning to execute a new skill such as the golf swing, or when a person with a stroke is re-learning how to reach for a cup of coffee. Augmented feedback can be classified into two types: *knowledge of results (KR)* and *knowledge of performance (KP)*. KR refers to feedback about the outcome of a movement, such as the score in a game of darts. KP refers to feedback about the nature of the movement pattern, such as whether the elbow was sufficiently extended when throwing a dart. Thus, KP feedback may be especially relevant if one wants to provide feedback about joint coordination patterns.

In a typical study that examines effects of augmented feedback on motor learning, participants first practice a task with augmented feedback during a period known as the *acquisition phase* (Schmidt and Lee, 2011). Performance during the acquisition phase is thought to represent a combination of effects derived from learning and the temporary guidance provided by augmented feedback. Therefore, to evaluate whether the skill has been learned, performance is tested during the *retention phase*, when the task is performed without augmented feedback. As such, acquisition and retention data are analyzed separately because the former may be conflated by the temporary

guidance effect of augmented feedback. In contrast, the retention data more clearly represents the degree to which a skill has been learned and retained (Winstein and Schmidt, 1990; Nicholson and Schmidt, 1991; Vander Linden et al., 1993; Tal, 1995; Wulf et al., 1998, 2010; Park et al., 2000). The retention phase can be further subdivided into immediate and delayed retention phases (Winstein and Schmidt, 1990). Immediate retention evaluates performance without feedback, shortly after skill acquisition on the same day. Delayed retention evaluates performance without feedback, usually on the following day, or even after a longer period.

Several studies have investigated what type or frequency of augmented feedback facilitates the retention of a motor skill (Winstein and Schmidt, 1990; Nicholson and Schmidt, 1991; Vander Linden et al., 1993; Tal, 1995; Wulf et al., 1998; Park et al., 2000). However, a yet to be resolved question is with what frequency should feedback be provided to facilitate retention. The influential guidance hypothesis (Salmoni et al., 1984; Schmidt et al., 1989) postulates that too much feedback is detrimental to motor skill learning. The guidance hypothesis makes three assumptions. First, frequent feedback, such as its provision on every training trial, is assumed to negatively affect learning because the learner comes to rely on augmented feedback at the expense of using his/her own inherent feedback. This reliance leads to the deterioration of performance when the augmented feedback is unavailable during the retention test.

Second, the guidance hypothesis assumes that a reduced frequency of augmented feedback (e.g., providing feedback on every other training trial) may facilitate learning because it promotes the learner to use their own inherent feedback during the no-feedback trials (Salmoni et al., 1984; Schmidt et al., 1989). The no-feedback trials also provide the learner with the opportunity to integrate information from previous feedback trials, with information derived from their own inherent feedback systems. The active use of inherent feedback systems during the no-feedback trials may help the learner form a motor command to execute a target movement without relying on the augmented feedback (Salmoni et al., 1984; Schmidt et al., 1989). Thus, when performance of the skill is tested at retention, there is no/less deterioration in performance because the learner is not reliant on the augmented feedback.

Third, the guidance hypothesis assumes that frequent augmented feedback may also increase movement variability (Salmoni et al., 1984; Schmidt et al., 1989). Movement variability is thought to increase because frequent augmented feedback encourages the learner to over-correct the movement (the so-called, *maladaptive short-term corrections*) even when performance is relatively close to the target (Schmidt, 1991). Therefore, taken together, the guidance hypothesis postulates that a reduced frequency of augmented feedback facilitates motor learning.

To our knowledge, the guidance hypothesis, or the optimal frequency of feedback, has not been tested in the context where individuals learn movements with augmented KP auditory feedback, provided concurrently with performance. Concurrent KP auditory feedback may be relevant for learning joint coordination patterns, especially in people

with movement disorders who may benefit from the re-learning of biomechanically and physiologically efficient movements. This is because feedback provided concurrently to the movement (as opposed to at the end, i.e., terminal feedback), may facilitate online motor planning of a joint coordination pattern. Furthermore, there may be a difference between the efficacy with which auditory and visual feedback facilitate motor learning. Auditory relative to visual feedback may guide movements in a temporally more efficient way given that the auditory system is generally better at resolving temporal information (Repp and Penel, 2002, 2004; Patel et al., 2005; Hove et al., 2010).

We developed our concurrent KP auditory feedback via the *sonification* of movements. Sonification refers to the use of sounds to convey information for the purposes of facilitating communication (Dubus and Bresin, 2013 for a review). For example, a sound variable such as loudness can be mapped onto a kinematic variable such as the vertical hand position. Here, the sound would get louder as the arm moves upwards, and quieter when the arm moves downwards. Recently, there has been increasing interest in understanding whether sonified feedback can facilitate motor (re)learning (Sigrist et al., 2013, 2015; Scholz et al., 2014, 2015). For example, a recent study mapped pitches of a violin sound onto oar movements for rowing, and showed that sonified feedback facilitates the learning of a target rowing velocity (Sigrist et al., 2015). Another study mapped pitch, brightness, and loudness of a synthesized sound onto the arm movements of stroke patients, in 3D space, and showed that re-training movements with sonified feedback improved arm motor functions (Scholz et al., 2015). However, there are at least two gaps in the literature that can be further explored to better understand the role of sonified feedback in motor learning. First, prior work discussed above sonified the endpoint movement. In the present study, we map a sound variable onto the error related to the performed joint coordination pattern. Thus, our novel work tests whether sonified feedback facilitates the motor (re)learning of joint coordination patterns. Second, these sonification studies compared the effect of sonified feedback with that of no feedback (Scholz et al., 2015) or visual and visuohaptic feedback (Sigrist et al., 2015). In the present study, we test the influential guidance hypothesis to investigate with what frequency sonified feedback should be provided to facilitate motor (re)learning. In accordance with the guidance hypothesis we postulate a reduced frequency of sonified feedback results in better retention of a learned joint coordination pattern.

We developed a sonification system that delivered a 440-Hz pure tone sound, which varied in intensity, in proportion to the error in the joint coordination pattern relative to a target pattern. We compared motor learning of the novel joint coordination pattern in two groups of healthy participants; one that received feedback on every training trial [i.e., 100% auditory feedback (AF)], and one that received feedback on every other training trial (i.e., 50% AF). According to the guidance hypothesis, we predicted: (1) The 50% AF group would show better retention of the learned joint coordination pattern compared to the 100% AF group because a reduced frequency of feedback encourages individuals to use inherent feedback naturally available to them; (2) The 50% AF group would show less variable joint

coordination patterns than the 100% AF group because a reduced frequency of feedback prevents individuals from making maladaptive corrections to their movement patterns.

METHODS

Participants

Twenty right-handed healthy individuals (16 females and 4 males) with normal hearing and no history of neurological or musculoskeletal disorders participated in this study. Participants were randomly assigned to one of two groups: 100 or 50% AF ($n = 10$ each). Age, handedness as assessed by the Edinburgh Handedness Inventory (Oldfield, 1971), number of years of education, and number of years of musical training are summarized in **Table 1**. The study was approved by the Institutional Review Board of Sunnybrook Health Sciences Centre and all participants provided written informed consent. Participants were compensated for their time and transportation.

Hearing Tests

In this study, participants learned a novel upper limb joint coordination pattern (see Section “Reaching Task” below) with augmented auditory feedback. The feedback was a sound that varied in intensity in proportion to the error of the joint coordination pattern relative to a target pattern. Therefore, to ensure participants were able to perceive these sounds, they underwent two hearing tests prior to the reaching task. A 440-Hz pure tone was used as the sound stimulus. The sounds were created with custom written C++ scripts and outputted as an analog signal with the analog input/output (AIO) board (ADA16-32/2CBF, Contec Co., Ltd., Japan). The analog signal was amplified by speakers (MM-SPWD2SV, Sanwa, Japan) and the sound was delivered to participants via headphones binaurally (MDR-NCB, SONY, Japan).

Hearing Test #1: Assessment of Hearing Threshold

This test determined the loudness threshold at which a participant heard a tone. This ensured all participants were able to perceptually hear the sounds, and ensured sound intensities were perceptually-equated across participants.

TABLE 1 | Participant demographics.

Group	Mean (SD)		P-values
	50% ($n = 10$, 2 males)	100% ($n = 10$, 2 males)	
Age (years)	33.9 (15.4)	34.0 (16.7)	0.631
Handedness*	89.0 (22.8)	92.0 (16.2)	1.000
Years of education	14.6 (3.8)	13.5 (2.8)	0.481
Years of musical training	7.0 (7.7)	8.4 (5.2)	0.739

*Laterality quotient assessed by the Edinburgh Handedness Inventory (Oldfield, 1971). Associated *p*-values for results from the independent-samples Mann-Whitney U-tests comparing 50 and 100% auditory feedback groups.

To determine the hearing threshold, we used the Bekesy audiometry method, which is a self-recording audiometer (Bekesy, 1947). In this method, an increase or decrease in sound intensity is controlled by the action of a switch that participants press on. When the switch is pressed, sound intensity begins to decrease. Once the participant no longer perceptually hears the tone, he/she releases the switch. At this point, the sound intensity begins to increase and when the participant perceptually hears the tone again, he/she presses the switch (see **Figure 1**). Therefore, by asking participants to press the switch when they hear the tone and to release the switch when they no longer hear the tone, an individual's hearing threshold can be determined via this self-recording approach (Bekesy, 1947).

The sound intensity of a 440-Hz pure tone either increased or decreased at a rate of 5 decibels (dBs) per second. One trial consists of 10 time-points at which the switch was pressed on and off (see the numbered turnaround points in **Figure 1**). The hearing threshold in a trial was calculated as the average of the middle six time points (see the turnaround points marked by red squares in **Figure 1**). Thus, we discarded two time points at the beginning and at the end, and only analyzed stable responses. Participants practiced until they became familiar with the task. After practice, four trials were recorded from which the mean hearing threshold was calculated.

Hearing Test #2: Detection of a Change in Sound Intensity

This test determined the perceptual threshold at which a participant detected a change in the sound intensity of a 440-Hz pure tone. In this study, sound intensity was manipulated such that the louder the sound, the larger the joint coordination error during reaching (see Section "Creation of Auditory Feedback" below). Therefore, this test ensured participants could perceptually detect changes in sound intensity so that they could use this information to minimize their joint coordination error during the reaching task.

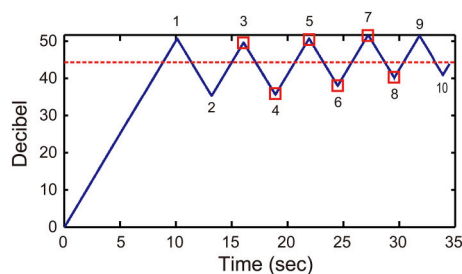


FIGURE 1 | Hearing test #1. An example audiogram obtained from self-recording audiometry. The vertical axis denotes the intensity of the 440-Hz pure tone in decibels relative to the intensity at the beginning of the trial. The turnaround points marked by 1, 3, 5, 7, and 9 denote the times during which a participant pressed a switch indicating they heard the tone; turnaround points 2, 4, 6, 8, and 10 denote the times during which the participant released the switch indicating they did not hear the tone. The hearing threshold (the dashed line) was calculated as the average of the middle six turnaround points, marked by open squares.

For this test, participants pressed a switch when they detected a change in sound intensity; they refrained from pressing the switch if there was no change. There were 48 trials in total comprising 40 test trials and 8 catch trials. Test trials involved a gradual change in sound intensity. Across the 40 test trials, there were eight different patterns in which sound intensity changed, with each pattern repeated five times. The eight patterns comprised four levels (e.g., 2, 4, 8, and 16 dBs) and two directions (increase/decrease) of sound intensity change (see **Figure 2A**). For example, the 440-Hz pure tone is sounded over 10-s while its intensity linearly increased or decreased at a fixed rate over a 4-s period in the mid-portion of the trial. The time point at which the sound intensity changed was jittered randomly between 2.0 and 2.5 s after the beginning of the trial. This prevented participants from anticipating the onset of a sound change.

There were no changes in sound intensity for catch trials; The tone was sounded over 10-s at a constant intensity (see **Figure 2B**). Four out of eight catch trials maintained the sound intensity at 0 dB while the other four catch trials maintained the intensity at one of the four levels (e.g., 2, 4, 8, or 16 dB). Note that sound intensity was set in decibels relative to the hearing threshold as determined in the first hearing test. Thus, 0 dB represents the hearing threshold and all other levels are calculated relative to this value.

First, participants practiced four trials comprising two test and two catch trials. Next, 48 trials were presented across two blocks of 24 trials with a break between the blocks. Trial order was randomized for each participant. We calculated percent of "changed intensity" response for each individual (see filled black circles in **Figure 2C**). In the figure, negative values on the horizontal axis denote conditions where sound intensity decreased, while positive values denote conditions where sound intensity increased (re: test trials). A zero value on the horizontal axis denotes no sound intensity change (re: catch trials). We fitted a binomial logistic regression model ("glmfit" function on Matlab with "binomial" and "logit" settings) to the data to draw a psychometric function. The perceptual thresholds were estimated for each participant by using the chance-level (50%) response value (see dashed vertical lines in **Figure 2C**). We calculated the mean perceptual thresholds for all participants.

Reaching Task

The reaching task was performed across two consecutive days (**Figure 3**). On Day 1, participants first performed 25 trials of baseline reaching (termed as "baseline phase"). Participants were seated with their forearm resting on the table and placed their hand on a start target (**Figure 4**). Participants reached to an ipsilateral end target in the sagittal plane, using shoulder flexion and elbow extension with no trunk displacement. A novel target joint coordination pattern was then created for each individual based on that person's kinematic data acquired during the baseline phase (see dashed lines in **Figure 5** and Section "Creation of Target Joint Coordination Pattern" below). The novel target joint coordination pattern can be described as follows: At the beginning of the reach, participants flexed the elbow and abducted the shoulder. In the middle portion of the reach, they extended the elbow while keeping the shoulder

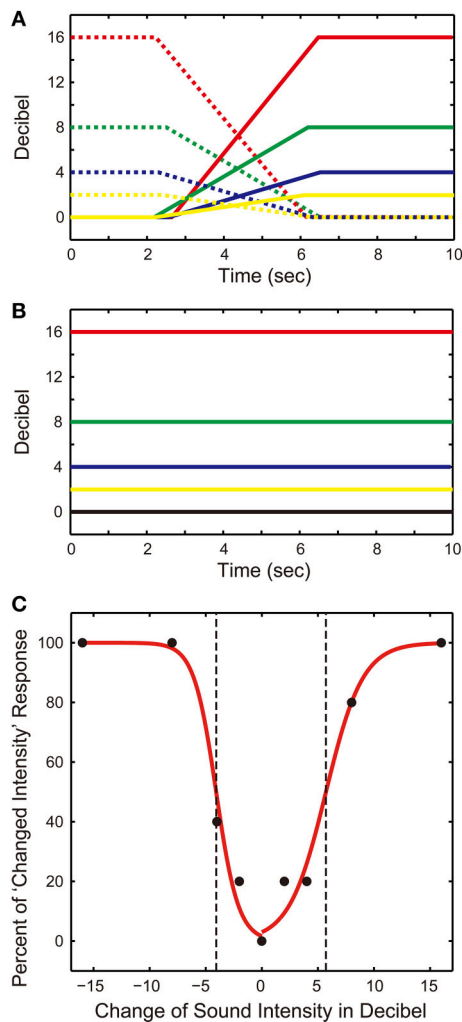


FIGURE 2 | Hearing test #2. (A) Patterns of sound stimuli in test trials. The intensity of a 440-Hz pure tone was increased (solid lines) or decreased (dashed lines) in the middle portion of a trial. There were four levels (2, 4, 8, and 16 dB) of sound intensity changes. **(B)** Pattern of sound stimuli in catch trials. The sound intensity was kept either at 0, 2, 4, 8, or 16 dB throughout the trial. **(C)** Psychometric function. Percent of “changed intensity” response was calculated from a participant’s responses (filled circles). A binomial logistic regression model was fitted to the data to draw a psychometric function. The perceptual thresholds were estimated with the chance-level (50%) response value (dashed vertical lines).

abducted. At the last portion of the reach, they flexed/adducted the shoulder to hit the end target. This novel target joint coordination pattern could be described like a “hook punch” movement in boxing. Importantly, participants were instructed to wear an eye mask while keeping their eyes closed throughout the task to prevent naturally available visual feedback from influencing motor learning.

Participants attempted to learn the novel joint coordination pattern across 100 trials of practice (“acquisition phase”). Participants in the 100% auditory feedback (AF) group received feedback on every trial while those in the 50% AF group received feedback on every other trial (Figure 3). The experimenter

informed participants that the tone would become louder as a performed joint coordination pattern deviated from the target joint coordination, while it would become quieter as performance became closer to the target joint coordination (see Section “Error in Joint Coordination Pattern” below). Thus, participants were instructed to minimize the sound intensity to the best of their ability. During the no feedback trials for the 50% AF group, participants were instructed to perform the target joint coordination pattern to the best of their ability (since no feedback was guiding them).

Immediately after the acquisition phase on Day 1, participants performed 25 trials of reaching without feedback to assess immediate retention of the learned joint coordination pattern (“immediate retention phase”). On Day 2, participants performed 25 trials of reaching without feedback to assess delayed retention of the learned joint coordination pattern (“delayed retention phase”).

Setup for Reaching Task

To measure movement kinematics during arm reaching, we used three goniometer sensors (Biometrics Ltd. UK). The sensors were attached using double-sided medical adhesive tape across three joints (elbow, shoulder, and trunk; see Figure 4). The proximal endblock of the elbow goniometer was attached to the arm with its center axis coincident with the center axis of the arm; the distal endblock was attached to the forearm with its center axis coincident with the center axis of the forearm. The proximal endblock of the shoulder goniometer was attached over the belly of the trapezius muscle aligning the distal end of the proximal endblock with the acromion, while the distal endblock was attached to the humerus with its center axis coincident with the center axis of the lateral side of the humerus with the inter-endblock distance of 14 cm. The lower endblock of the trunk goniometer was attached to the lumbar spine with its center axis coincident with the center of the spine aligning the top level of the endblock to the level of L5; the upper endblock was attached to the thoracic spine with its center axis coincident with the center of the spine with the inter-endblock distance of 7 cm. The sensor positions were marked on the skin with a pen to ensure identical placement of goniometer sensors across the 2 days. We used single-axis goniometers to collect data from the elbow (flexion/extension) and trunk (flexion/extension; shown as “Elbow” and “Trunk” in Figure 5A). We used a twin-axis goniometer to record data from the shoulder (abduction/adduction: “Shoulder 1,” flexion/extension: “Shoulder 2”).

The height of the chair was fixed at 46 cm. We adjusted the height of the table for each participant to ensure the forearm was in a comfortable position. The average height of the table was 69.1 ± 2.5 cm (mean \pm standard deviation). The start target was embedded into the surface of the table and consisted of an electric switch (1.5 V battery) that recorded the start time of a reach. The end target was placed in front of the participant at arm’s length and at shoulder level (with the shoulder at 90° of flexion and elbow at 0° of flexion). The end target also consisted of an electric switch (1.5 V battery) that recorded the time when the end target was hit. The center-to-center distance from the start

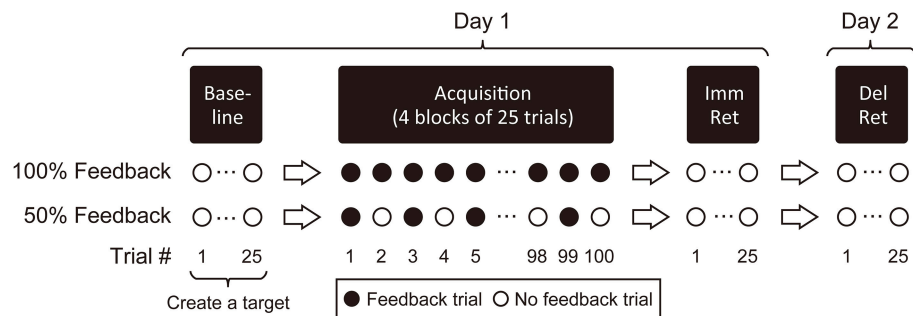


FIGURE 3 | Reaching task paradigm. Twenty-five trials of normal reaching movements with no feedback were performed during the baseline phase. These kinematic data were used to create a target joint coordination pattern. Participants then performed four blocks of 25 trials (100 trials) to acquire the target joint coordination pattern (acquisition phase); they received auditory feedback on every (100% feedback) or every other practice trial (50% feedback). Twenty-five no-feedback trials were performed immediately after the acquisition phase (Imm Ret: immediate retention phase on Day 1), and an additional 25 no-feedback trials were performed the next day (Del Ret: delayed retention phase on Day 2).

to the end targets was 40.2 ± 4.5 cm. The start and end targets were aligned in the same sagittal plane, ipsilateral to the reaching arm. The position of the start and end target switches were fixed throughout the reaching task. The distance from the front legs of the chair to the end target was 28.7 ± 5.0 cm. Participants were seated with their initial arm position in 97.4 ± 11.4 degrees of elbow flexion, 31.4 ± 6.6 degrees of shoulder abduction, with the hand closed in a fist, resting on the start target.

Signals of the goniometer sensors were amplified with the K800 amplifier (Biometrics Ltd., UK). Signals of the electric switches and goniometer sensors were synchronized and converted from analog to digital at a frequency of 200 Hz with the AIO board, and recorded on a personal computer with a custom written program in C++. Goniometer data was low-pass filtered offline using a 4th order Butterworth filter with a cut off frequency of 10 Hz by custom written scripts in Matlab software (Mathworks, USA).

Creation of Target Joint Coordination Pattern

The joint coordination pattern for reaching during the baseline phase was used to create a novel target joint coordination pattern to be learned by participants. A typical data set from the 25 baseline trials of a participant is shown in **Figure 5A**. The left panel shows data during the forward portion of the reach (start to end target) while the right panel shows data during the backward portion of the reach (end to start target). The x-axis represents time, normalized as percentage of reach (% reach) based on duration recorded from the start and end target switches. The average of 25 trials for each joint is shown as solid lines in **Figure 5B**.

The average joint coordination pattern is represented as a trajectory in three-dimensional joint coordination space, consisting of the averaged Elbow, Shoulder 1, and Shoulder 2 signals (average of 25 trials for each joint, see solid black line in **Figure 5C**). The Trunk signal is not included since no participant moved the trunk.

To create the novel target joint coordination pattern, we “deflected” the Elbow (E), Shoulder 1 (S1), and Shoulder

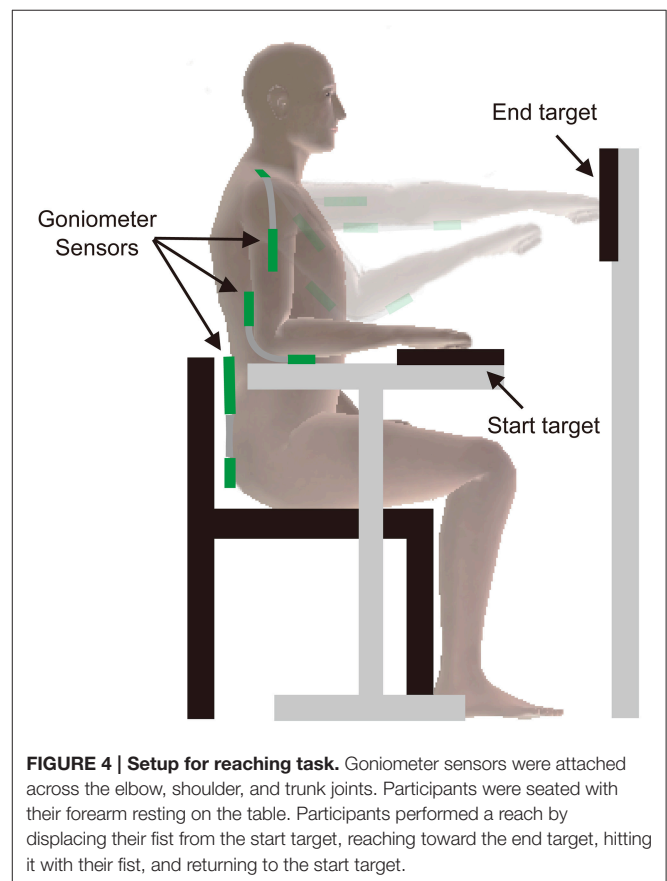


FIGURE 4 | Setup for reaching task. Goniometer sensors were attached across the elbow, shoulder, and trunk joints. Participants were seated with their forearm resting on the table. Participants performed a reach by displacing their fist from the start target, reaching toward the end target, hitting it with their fist, and returning to the start target.

2(S2) signals (see dashed lines in **Figures 5B,C**; see also the Supplementary Material for details on how to deflect the trajectory). The idea to deflect the movement pattern was based on a previous reaching study that deflected each individual's baseline trajectory to create a novel target reaching pattern (Wu et al., 2014). We applied this idea because it allowed us to create a novel and unfamiliar upper limb coordination pattern for each individual to learn using augmented auditory feedback.

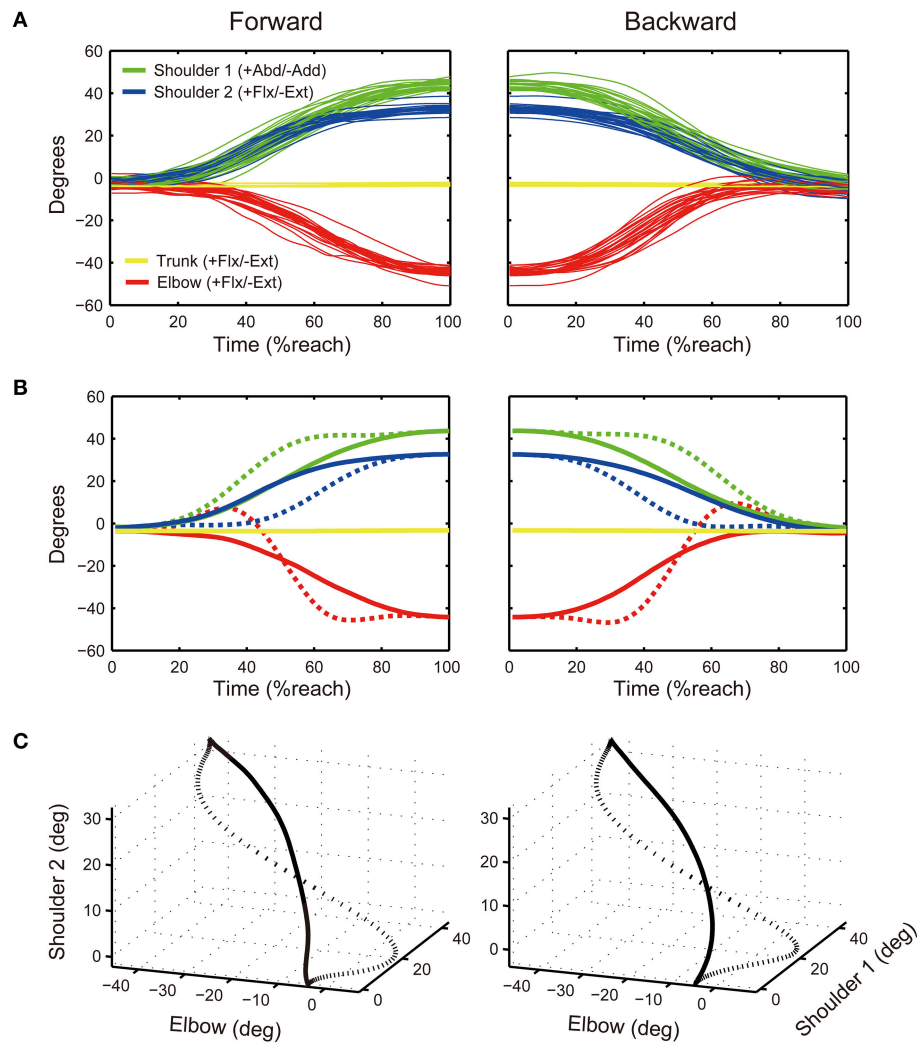


FIGURE 5 | Processing of baseline data. (A) A typical example of data recorded during the baseline phase from a participant (right arm, 25 trials). The left panel shows data during the forward reach while the right panel shows data during the backward reach. For the shoulder, a twin-axis goniometer was used that enabled the recording of shoulder abduction/adduction (Abd/Add; Shoulder 1, green line) and shoulder flexion/extension (Flx/Ext; Shoulder 2, blue line). For the elbow and trunk goniometer sensors, a single-axis goniometer was used that recorded elbow flexion/extension (Flx/Ext) and trunk forward flexion and extension (Flx/Ext; see red and yellow lines). **(B)** Averaged time series across the 25 trials (solid lines) and deflected time series (dashed lines). **(C)** Three-dimensional plot consisting of Elbow, Shoulder 1, and Shoulder 2 signals. Solid line denotes the average joint coordination pattern during the baseline, while dashed line denotes the deflected target joint coordination pattern.

Error in Joint Coordination Pattern

To create concurrent KP auditory feedback, data from the goniometer sensors were processed every 10 ms in the C++ program. To assess how a performed joint coordination pattern deviates from the target joint coordination (defined from the deflection of the baseline reach), we define an error at i -th sampled time frame (e_i) as,

$$e_i = \sqrt{(E_i - E_j)^2 + (S1_i - S1_j)^2 + (S2_i - S2_j)^2}; \quad (1)$$

where E_i , $S1_i$, and $S2_i$ are i -th sampled time frame of performed joint angles measured by the goniometer sensors at the elbow and shoulder. For example, a performed joint coordination at the i -th

sampled time frame can be drawn as a point (P_i) in the three-dimensional joint coordination space (**Figure 6**). E_j , $S1_j$, and $S2_j$ are j -th sampled time frame of the target joint angles where the distance from P_i to the target trajectory becomes minimum. Thus, an error (e_i) can be drawn as the minimum distance from P_i to the target trajectory (**Figure 6**). The intensity of the feedback sound at the i -th time frame (I_i) is then set in decibels to be twice as large as the amount of error in degrees;

$$I_i = 2 \times e_i \quad (2)$$

Thus, sound intensity in decibels was set as zero ($I_i = 0$) if a performed joint coordination perfectly matched the target ($e_i =$

0), while for example, a participant hears a 20 dB tone if the error is 10° from the target. Here, zero decibels correspond to the hearing threshold as determined in hearing test #1.

Measures to Assess Joint Coordination Pattern

To assess the degree to which participants achieved the target joint coordination pattern, we calculated the root-mean squared error (RMSE) between the performed and target trajectories in the three-dimensional joint coordination space:

$$RMSE = \frac{1}{n} \sum_{i=1}^n e_i; \quad (3)$$

where e_i is the error of joint coordination at the i -th sampled time frame (see **Figure 6**) and n is the total number of data points. Note, that root of the squared error (RSE) was already calculated in Equation (1) and therefore the mean over the time points (RMSE) was calculated in Equation (3).

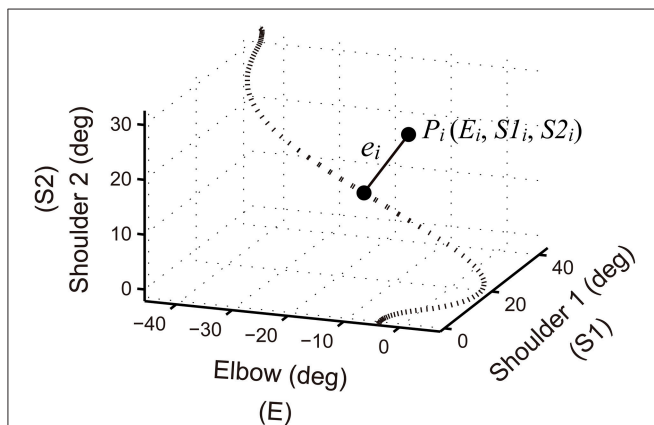


FIGURE 6 | Schematics of joint coordination error. A performed joint coordination pattern at the i -th sampled time frame can be drawn as a point (P_i) in the three-dimensional joint coordination space. The pattern is derived from the Elbow (E), Shoulder 1 (S1), and Shoulder 2 (S2) signals measured with the goniometer sensors. The dashed line denotes the target joint coordination pattern. An error at the i -th sampled time frame (e_i) can be visualized as the minimum distance from P_i to the target trajectory. Auditory feedback was created by changing the sound intensity of a pure tone in proportion to the amount of error.

The RMSE assesses the degree to which the performed joint coordination pattern deviates from the target joint coordination pattern. To assess the consistency of joint coordination across trials, we calculated the variable error (VE; Schmidt and Lee, 2011). The calculation of VE is similar to that of RMSE but differs in the reference trajectory used to quantify the error. To calculate the RMSE, the target joint coordination pattern was used as the reference trajectory. On the other hand, the mean joint coordination trajectory across the 25 trials in a block was used as the reference trajectory to evaluate the error in the VE measure (see **Figure 7**). In **Figure 7**, an example of inconsistent and consistent joint coordination patterns is shown in the left and right panels, respectively. The data from 25 trials of reaching are plotted in each of the left and right panels. The black line in the figure shows the average across the 25 trials. (Note, that the black line is not the target trajectory.) For each trial, the VE was calculated as the RMSE between a performed trajectory (a red line) and the mean trajectory (the black line). The VE is larger in the left example compared to the right one.

The RMSE and VE were calculated from data acquired between 20 and 80% of the reach where the main deflection was made to create the target trajectory. For each individual, the RMSE and VE were calculated for each trial during the acquisition and retention phases then averaged across 25 trials in each block. Note that the RMSE and VE were calculated for acquisition and retention data but not for baseline data. This was because participants had no “target” during baseline since they performed 25 trials of normal reaching during this phase.

Statistics

Consistent with previous studies (Winstein and Schmidt, 1990; Nicholson and Schmidt, 1991; Vander Linden et al., 1993; Tal, 1995; Wulf et al., 1998, 2010; Park et al., 2000), acquisition and retention data were analyzed separately. For the acquisition phase, the RMSE and VE were subjected to a two-way repeated-measures analysis of variance (ANOVA) with the within-participant factor of Block (Block 1, 2, 3, and 4) and the between-participant factor of Group (100 and 50% AF groups). For the retention phase, the RMSE and VE were subjected to a two-way repeated-measures ANOVA with the within-participant factor of Day (immediate retention on Day 1 and delayed retention on Day 2) and the between-participant factor of Group (100 and

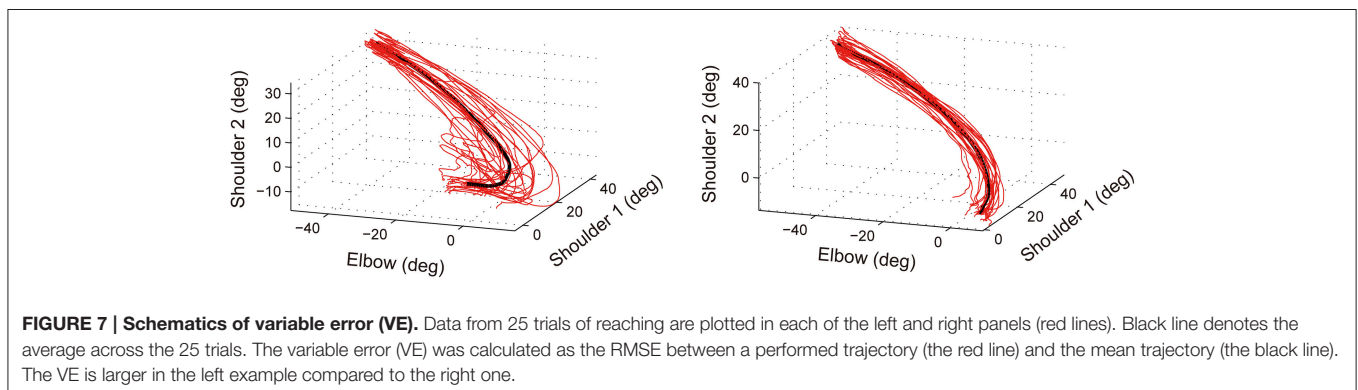


FIGURE 7 | Schematics of variable error (VE). Data from 25 trials of reaching are plotted in each of the left and right panels (red lines). Black line denotes the average across the 25 trials. The variable error (VE) was calculated as the RMSE between a performed trajectory (the red line) and the mean trajectory (the black line). The VE is larger in the left example compared to the right one.

50% AF groups). The perceptual threshold to detect a change in sound intensity (measured in hearing test #2) was subjected to an independent-samples *t*-test to compare thresholds between the 100 and 50% AF groups. We used the Mann-Whitney *U* tests to compare the 100 and 50% AF groups on age, handedness, number of years of education, and number of years of musical training. Significance was set at $P < 0.05$ (two-tailed) for all statistical tests.

RESULTS

Demographics

There was no Significant Difference in Age, Handedness, Number of Years of Education, and Number of Years of Musical Training Between the Two Groups ($P > 0.143$, see Table 1).

Perceptual Threshold

The perceptual thresholds for the detection of changes in sound intensity (as measured in hearing test #2) are summarized in Table 2. There was no significant difference in perceptual thresholds between the two groups ($P > 0.35$).

A Typical Joint Coordination Pattern during Acquisition and Retention Phases

A typical example of the performed elbow and shoulder joint coordination pattern for the forward portion of the reach, during acquisition, and retention phases are shown in Figure 8. The RMSE and VE become smaller over the course of practice blocks during the acquisition phase. That is, the performed trajectories (red lines) become closer to the target (black line) and less variable. The RMSE at the delayed retention phase was larger than that at the immediate retention phase. That is, the performed trajectories (red lines) deviated more from the target (black) compared to those at the immediate retention phase.

Acquisition Phase

For the RMSE, there was no significant interaction between the Block and Group factors in the two-way ANOVA [$F_{(3, 54)} = 0.31$, $P = 0.82$, $\eta^2 = 0.02$]. The main effect of Block was significant [$F_{(3, 54)} = 7.92$, $P < 0.001$, $\eta^2 = 0.31$] whereas, that of Group was

not [$F_{(1, 18)} = 0.17$, $P = 0.69$, $\eta^2 = 0.01$], showing that both 100 and 50% groups reduced the joint coordination error relative to the target across practice in the acquisition phase (see Figure 9).

For the VE, there was no significant interaction between the Block and Group factors in the two-way ANOVA [$F_{(3, 54)} = 0.47$, $P = 0.70$, $\eta^2 = 0.03$]. The main effect of Block was significant [$F_{(3, 54)} = 12.80$, $P < 0.001$, $\eta^2 = 0.42$] whereas, that of Group was not [$F_{(1, 18)} = 3.90$, $P = 0.06$, $\eta^2 = 0.18$], showing that the VE became smaller in both 100 and 50% groups across practice in the acquisition phase (see Figure 10). Taken together, both RMSE and VE were significantly reduced over the course of training in both 100 and 50% feedback groups, suggesting that auditory feedback guided the joint coordination pattern to the target with less variability during the acquisition phase.

Retention Phase

For the RMSE, there was no significant interaction between the Day and Group factors in the two-way ANOVA [$F_{(1, 18)} = 0.99$, $P = 0.33$, $\eta^2 = 0.05$] (Figure 9). The main effect of Day was significant [$F_{(1, 18)} = 7.18$, $P < 0.05$, $\eta^2 = 0.29$], showing that RMSE was smaller at the immediate retention phase compared with the delayed retention phase. The main effect of Group was also significant [$F_{(1, 18)} = 4.86$, $P < 0.05$, $\eta^2 = 0.21$], showing that the RMSE of the 100% AF group was significantly smaller than that of the 50% AF group at both retention phases.

For the VE, there was no significant interaction between the Day and Group factors in the two-way ANOVA [$F_{(1, 18)} = 1.15$, $P = 0.30$, $\eta^2 = 0.06$] (Figure 10). No main effect of Day nor Group was found in the ANOVA [Day: $F_{(1, 18)} = 1.40$, $P = 0.25$, $\eta^2 = 0.07$; Group: $F_{(1, 18)} = 2.89$, $P = 0.11$, $\eta^2 = 0.14$, respectively]. Taken together, the 100% AF group showed smaller RMSE than the 50% AF group while VE was comparable between the groups at both retention phases.

DISCUSSION

The purpose of this study was to test the guidance hypothesis (Salmoni et al., 1984; Schmidt et al., 1989) in the context of learning a novel joint coordination pattern with concurrent KP auditory feedback. According to the guidance hypothesis (Salmoni et al., 1984; Schmidt et al., 1989), we predicted the following: First, the 50% AF group would show better retention of learned joint coordination patterns after the removal of auditory feedback, compared to the 100% AF group. Second, the 50% AF group would show less variable joint coordination patterns than the 100% AF group. Contrary to the first prediction, the 100% AF group showed better retention of the learned joint coordination pattern (i.e., smaller RMSE) at both immediate and delayed retention phases, than the 50% AF group. Contrary to the second prediction, there was no significant difference in VE between the 50 and 100% AF groups for either acquisition or retention phases. Thus, the guidance hypothesis was not supported in this study using our specific type of sonified feedback manipulation. Our results suggest that concurrent KP auditory feedback facilitates learning of a novel joint coordination pattern when the feedback is presented more frequently.

TABLE 2 | Hearing test #2: perceptual threshold to detect a change in sound intensity (decibels).

Group	Mean (SD)		T_{18}	P -values
	50% ($n = 10$)	100% ($n = 10$)		
Perceptual threshold (Intensity decreased)	5.2 (4.2)	5.9 (3.6)	0.396	0.696
Perceptual threshold (Intensity increased)	6.3 (2.7)	7.7 (3.8)	0.970	0.345
Perceptual threshold (Mean)	5.7 (2.9)	6.8 (3.5)	0.736	0.471

Associated p -values for results of the independent-samples *t*-tests comparing 50 and 100% auditory feedback groups.

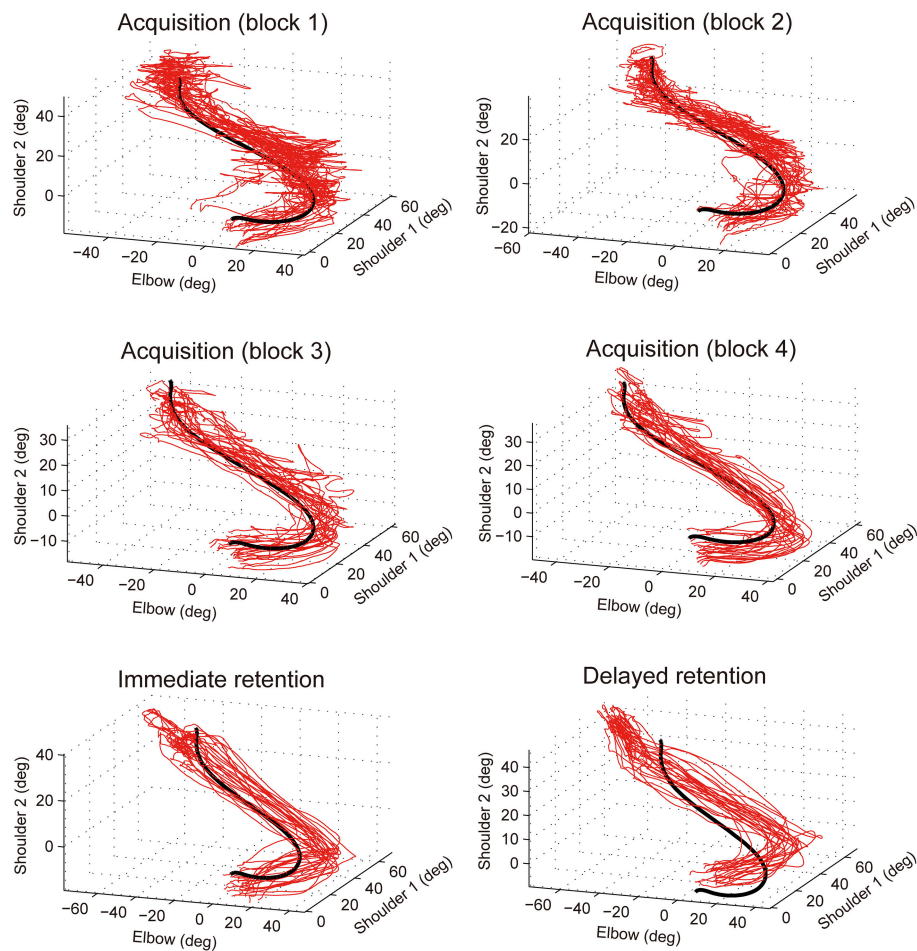


FIGURE 8 | Examples of joint coordination patterns in acquisition and retention phases. Data from 25 trials representing the forward reach from a participant in the 100% auditory feedback group. Red thin lines show performed joint coordination patterns while black solid lines show the target joint coordination pattern.

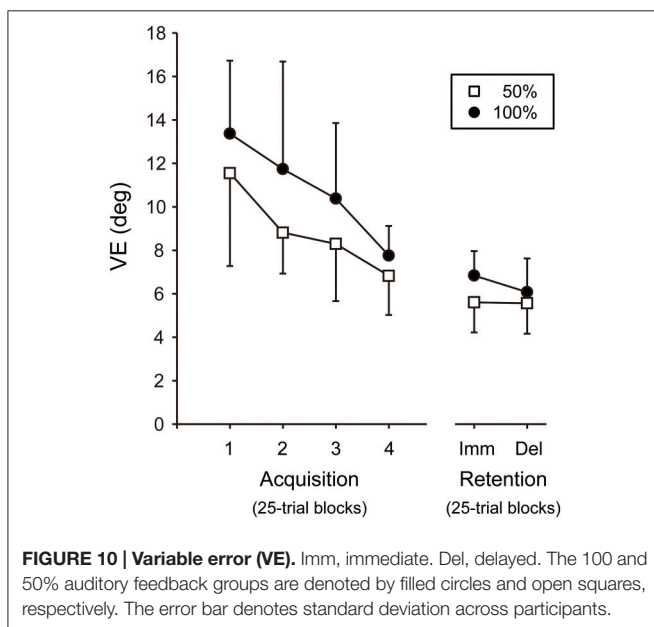
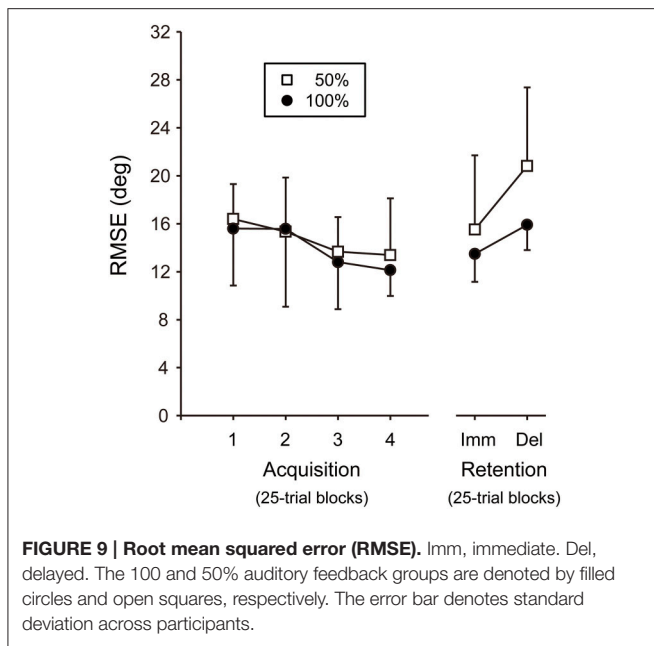
Why More is Better

In our study, the 100% AF group showed better retention of the learned joint coordination pattern compared to the 50% AF group. These findings are in contrast with those from prior research that showed better retention of performance when the skill was learned with a reduced frequency of feedback (Winsten and Schmidt, 1990; Nicholson and Schmidt, 1991; Vander Linden et al., 1993; Park et al., 2000).

We suggest that task complexity could be a main reason why more feedback led to better retention in this study. While some studies support the guidance hypothesis (Winsten and Schmidt, 1990; Nicholson and Schmidt, 1991; Vander Linden et al., 1993; Park et al., 2000), others do not (Wulf et al., 1998, 2010). Studies that do not support the guidance hypothesis showed that frequent augmented feedback resulted in better retention of the learned skill. These findings are consistent with our results. Wulf and Shea (2002) pointed out in their review that studies supporting the guidance hypothesis used relatively simple tasks seen in typical laboratory settings (e.g., the lever patterning task in the study by Winsten and Schmidt (1990)). In contrast, studies that

do not support the guidance hypothesis use more complex tasks such as those that mimic real-life learning situations (e.g., the ski simulator task in the study by Wulf et al. (1998) and the bimanual soccer throw-in task in the study by Wulf et al. (2010)). The reaching task in the present study could be regarded as relatively complex. In fact, some participants in this study reported that the task was very demanding. Thus, there may be an interaction between task complexity and feedback frequency (Wulf et al., 1998; Wulf and Shea, 2002). The learning of simple motor skills may benefit from a reduced frequency of feedback while the learning of more complex motor skills may benefit from a higher frequency of feedback. Taken together, task complexity could explain why more feedback led to better retention.

One might assume that the modality of feedback could also be a factor that explains the discrepancy. To test the guidance hypothesis, previous studies used augmented visual feedback (Winsten and Schmidt, 1990; Nicholson and Schmidt, 1991; Vander Linden et al., 1993; Swinnen et al., 1997; Wulf et al., 1998; Park et al., 2000). To our knowledge, a limited number of studies used augmented auditory feedback and showed that it



did not deteriorate performance of the learned skill at retention (Ronsse et al., 2011; Sigrist et al., 2015). Interestingly, the provision of visual feedback did negatively affect performance (Ronsse et al., 2011). Moreover, audiovisual feedback better facilitated the learning of a target rowing velocity compared to visuohaptic feedback (Sigrist et al., 2015). Thus, the use of augmented auditory feedback (in contrast to other modalities of feedback) may also be one reason why more feedback led to better retention in this study. However, at this point, it is not clear how the modality of feedback and frequency of feedback interact. Future studies are needed to clarify this issue.

Potential Confounders: Perception and Age

First, one might assume that the difference between the 100 and 50% AF groups could be simply attributed to a difference in auditory perception. However, our results from hearing test #2 rule out this possibility. There was no significant difference between the two groups in their perceptual thresholds to detect changes in sound intensity. In addition, there was no significant difference in the number of years of musical training, showing that musical background was also comparable between groups. Thus, the observed group difference in this study could not be attributed to a difference in auditory perceptual capability.

Second, one might assume that the age of participants could be a confounder. The mean age of participants in this study (i.e., 34.0 years of age) was relatively higher compared with those of previous studies that showed the advantage of reduced frequency of feedback for motor learning (Winstein and Schmidt, 1990; Nicholson and Schmidt, 1991; Vander Linden et al., 1993; Park et al., 2000; these studies tested mostly undergraduate students as participants). Therefore, one might think that more feedback would lead to better skill retention in an older population because older adults may demonstrate slower motor learning (Fernandez-Ruiz et al., 2000) and require more information to help them learn. To test this possibility, we performed an additional analysis to investigate the relationship between age and RMSE at retention (i.e., averaged RMSE across the immediate and delayed retention phases). We found no significant correlation for the 100% AF group (Spearman's $\rho = -0.55$, $P = 0.10$) or for the 50% AF group ($\rho = -0.10$, $P = 0.79$). Thus, there was no significant relationship between age and retention of performance in this study. In addition, previous studies show that both younger and older adults process feedback similarly (Swanson and Lee, 1992; Wishart and Lee, 1997), and benefit from reduced frequency of feedback to learn (Tal, 1995). This suggests that an older population does not necessarily need more feedback to learn a skill. Accordingly, age may not be a factor to explain why more feedback led to better retention in this study.

The Role of Movement Variability in Motor Learning

During the acquisition phase, the 100% AF group showed a trend toward increased variability in the joint coordination pattern (i.e., larger VE) compared to the 50% AF group ($P = 0.06$). The guidance hypothesis views increased variability as a negative outcome, defining it as maladaptive short-term corrections (Schmidt, 1991). In contrast, more recent motor-control studies view movement variability as an essential ingredient that facilitate motor learning (Herzfeld and Shadmehr, 2014; Wu et al., 2014). Given that the 100% AF group showed better retention of the novel joint coordination pattern, the tendency toward an increased variability in the joint coordination pattern observed during acquisition may be viewed as adaptive (and not maladaptive) corrections in this study. That is, the tendency for increased movement variability in the 100% AF group may have a functional role, helping the learner adapt to a new situation. Specifically, the reaching task in this study required participants to map changes in sound intensity onto changes

in joint coordination patterns. Therefore, if joint coordination patterns were more variable, participants would be able to acquire more information about the auditory-motor mapping. This might then help the learner to develop a more advanced internal model, which may lead to better retention of the learned joint coordination pattern.

If the above assumption is correct, there should be a tight relationship between VE during the acquisition phase and RMSE at the retention phase. A learner, who experiences more variable joint coordination patterns during the acquisition phase, would show better retention of the target joint coordination pattern at the immediate and delayed retention phases. Thus, a significant correlation between the two measures is expected. We therefore performed an analysis to investigate the relationship between the VE averaged across the four acquisition blocks and the mean RMSE across the immediate and delayed retention phases. However, there was no significant correlation for the 100% AF group (Pearson's $r = -0.31$, $P = 0.39$) or for the 50% AF group ($r = -0.55$, $P = 0.10$). Thus, we cannot make a strong case regarding a potential relationship between increased movement variability and learning. Future studies are needed to clarify the role of movement variability for the learning of a novel joint coordination pattern with augmented auditory feedback.

CONCLUSION

Our study demonstrates that concurrent KP auditory feedback may facilitate the learning of a novel upper-limb joint

coordination pattern when it is provided during all practice trials as opposed to during half of the trials. Our finding will help us better understand how to facilitate the (re)learning of organized joint coordination patterns with auditory feedback during motor skill acquisition and rehabilitation.

AUTHOR CONTRIBUTIONS

SF, TL, and JC conceived and designed the study. SF and TL performed the experiment. SF analyzed the data. SF, TL, and JC interpreted the data and wrote the paper.

ACKNOWLEDGMENTS

We thank Dr. Masaya Hirashima for his technical help with development of the augmented auditory feedback. We thank Payal Gandhi, Ayesha Tasneem, Alvina Siu, and Kanako Sugita for their help with the experimental setup, testing, and data preprocessing across different phases of this study. SF was supported by a fellowship from the Japan Society for the Promotion of Science (JSPS). This study was supported by funding from the Heart and Stroke Foundation (HSF) Canadian Partnership for Stroke Recovery (CPSR).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fnins.2016.00251>

REFERENCES

- Bekesy, G. V. (1947). A new audiometer. *Acta Otolaryngol.* 35, 411–422.
- Bernstein, N. (1967). *The Coordination and Regulation of Movements*. New York, NY: Pergamon Press.
- Cirstea, M. C., and Levin, M. F. (2007). Improvement of arm movement patterns and endpoint control depends on type of feedback during practice in stroke survivors. *Neurorehabil. Neural Repair* 21, 398–411. doi: 10.1177/1545968306298414
- Dubus, G., and Bresin, R. (2013). A systematic review of mapping strategies for the sonification of physical quantities. *PLoS ONE* 8:e82491. doi: 10.1371/journal.pone.0082491
- Fernandez-Ruiz, J., Hall, C., Vergara, P., and Diaz, R. (2000). Prism adaptation in normal aging: slower adaptation rate and larger aftereffect. *Brain Res. Cogn. Brain Res.* 9, 223–226. doi: 10.1016/S0926-6410(99)00057-9
- Furuya, S., Osu, R., and Kinoshita, H. (2009). Effective utilization of gravity during arm downswing in keystrokes by expert pianists. *Neuroscience* 164, 822–831. doi: 10.1016/j.neuroscience.2009.08.024
- Herzfeld, D. J., and Shadmehr, R. (2014). Motor variability is not noise, but grist for the learning mill. *Nat. Neurosci.* 17, 149–150. doi: 10.1038/nn.3633
- Hirashima, M. (2011). "Induced acceleration analysis of three-dimensional multi-joint movements and its application to sports movements," in *Theoretical Biomechanics*, ed V. Klika (Rijeka: InTech), 303–318.
- Hove, M. J., Spivey, M. J., and Krumhansl, C. L. (2010). Compatibility of motion facilitates visuomotor synchronization. *J. Exp. Psychol. Hum. Percept. Perform.* 36, 1525–1534. doi: 10.1037/a0019059
- Levin, M. F. (1996). Interjoint coordination during pointing movements is disrupted in spastic hemiparesis. *Brain* 119, 281–293.
- Nicholson, D. E., and Schmidt, R. A. (1991). "Scheduling information feedback to enhance training effectiveness," in *Proceedings of the Human Factors Society 35th Annual Meeting* (San Francisco, CA), 1400–1402.
- Oldfield, R. C. (1971). The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia* 9, 97–113.
- Park, J. H., Shea, C. H., and Wright, D. L. (2000). Reduced-frequency concurrent and terminal feedback: a test of the guidance hypothesis. *J. Mot. Behav.* 32, 287–296. doi: 10.1080/00222890009601379
- Patel, A. D., Iversen, J. R., Chen, Y., and Repp, B. H. (2005). The influence of metricality and modality on synchronization with a beat. *Exp. Brain Res.* 163, 226–238. doi: 10.1007/s00221-004-2159-8
- Repp, B. H., and Penel, A. (2002). Auditory dominance in temporal processing: new evidence from synchronization with simultaneous visual and auditory sequences. *J. Exp. Psychol. Hum. Percept. Perform.* 28, 1085–1099. doi: 10.1037/0096-1523.28.5.1085
- Repp, B. H., and Penel, A. (2004). Rhythmic movement is attracted more strongly to auditory than to visual rhythms. *Psychol. Res.* 68, 252–270. doi: 10.1007/s00426-003-0143-8
- Ronsse, R., Puttemans, V., Coxon, J. P., Goble, D. J., Wagemans, J., Wenderoth, N., et al. (2011). Motor learning with augmented feedback: modality-dependent behavioral and neural consequences. *Cereb. Cortex* 21, 1283–1294. doi: 10.1093/cercor/bhq209
- Salmoni, A. W., Schmidt, R. A., and Walter, C. B. (1984). Knowledge of results and motor learning: a review and critical reappraisal. *Psychol. Bull.* 95, 355–386.
- Schmidt, R. A. (1991). "Frequent augmented feedback can degrade learning: evidence and interpretations," in *Tutorials in Motor Neuroscience*, eds J. Requin and G. E. Stelmach (Dordrecht: Kluwer), 59–75.

- Schmidt, R. A., and Lee, T. D. (2011). *Motor Control and Learning*. Champaign: Human Kinetics.
- Schmidt, R. A., Young, D. E., Swinnen, S., and Shapiro, D. C. (1989). Summary knowledge of results for skill acquisition: support for the guidance hypothesis. *J. Exp. Psychol. Learn. Mem. Cogn.* 15, 352–359.
- Scholz, D. S., Rhode, S., Grossbach, M., Rollnik, J., and Altenmüller, E. (2015). Moving with music for stroke rehabilitation: a sonification feasibility study. *Ann. N.Y. Acad. Sci.* 1337, 69–76. doi: 10.1111/nyas.12691
- Scholz, D. S., Wu, L., Pirzer, J., Schneider, J., Rollnik, J. D., Grossbach, M., et al. (2014). Sonification as a possible stroke rehabilitation strategy. *Front. Neurosci.* 8:332. doi: 10.3389/fnins.2014.00332
- Sigrist, R., Rauter, G., Marchal-Crespo, L., Riener, R., and Wolf, P. (2015). Sonification and haptic feedback in addition to visual feedback enhances complex motor task learning. *Exp. Brain Res.* 233, 909–925. doi: 10.1007/s00221-014-4167-7
- Sigrist, R., Rauter, G., Riener, R., and Wolf, P. (2013). Augmented visual, auditory, haptic, and multimodal feedback in motor learning: a review. *Psychon. Bull. Rev.* 20, 21–53. doi: 10.3758/s13423-012-0333-8
- Swanson, L. R., and Lee, T. D. (1992). Effects of aging and schedules of knowledge of results on motor learning. *J. Gerontol.* 47, P406–P411. doi: 10.1093/geronj/47.6.P406
- Swinnen, S. P., Lee, T. D., Verschueren, S., Serrien, D. J., and Bogaerds, H. (1997). Interlimb coordination: learning and transfer under different feedback conditions. *Hum. Mov. Sci.* 16, 749–785.
- Tal, J. (1995). Is more always better? Optimal amounts of feedback in learning to calibrate sensory awareness. *Occup. Ther. J. Res.* 15, 181–197.
- Vander Linden, D. W., Cauraugh, J. H., and Greene, T. A. (1993). The effect of frequency of kinetic feedback on learning an isometric force production task in nondisabled subjects. *Phys. Ther.* 73, 79–87.
- Winsten, C. J., and Schmidt, R. A. (1990). Reduced frequency of knowledge of results enhances motor skill learning. *J. Exp. Psychol. Learn. Mem. Cognit.* 16, 677–691.
- Wishart, L. R., and Lee, T. D. (1997). Effects of aging and reduced relative frequency of knowledge of results on learning a motor skill. *Percept. Mot. Skills* 84, 1107–1122. doi: 10.2466/pms.1997.84.3.1107
- Wu, H. G., Miyamoto, Y. R., Gonzalez Castro, L. N., Olveczky, B. P., and Smith, M. A. (2014). Temporal structure of motor variability is dynamically regulated and predicts motor learning ability. *Nat. Neurosci.* 17, 312–321. doi: 10.1038/nn.3616
- Wulf, G., Chiviacowsky, S., Schiller, E., and Avila, L. T. (2010). Frequent external-focus feedback enhances motor learning. *Front. Psychol.* 1:190. doi: 10.3389/fpsyg.2010.00190
- Wulf, G., and Shea, C. H. (2002). Principles derived from the study of simple skills do not generalize to complex skill learning. *Psychon. Bull. Rev.* 9, 185–211. doi: 10.3758/BF03196276
- Wulf, G., Shea, C. H., and Matschiner, S. (1998). Frequent feedback enhances complex motor skill learning. *J. Mot. Behav.* 30, 180–192. doi: 10.1080/00222899809601335

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.

Copyright © 2016 Fujii, Lulic and Chen. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Sonification of Arm Movements in Stroke Rehabilitation – A Novel Approach in Neurologic Music Therapy

Daniel S. Scholz¹, Sönke Rohde¹, Nikou Nikmaram¹, Hans-Peter Brückner¹, Michael Großbach¹, Jens D. Rollnik² and Eckart O. Altenmüller^{1*}

¹ Institute of Music Physiology and Musicians' Medicine, University of Music, Drama and Media, Hannover, Germany,

² Institute for Neurorehabilitational Research (InFo), BDH-Clinic Hessisch Oldendorf, Teaching Hospital of Hannover Medical School (MHH), Hessisch Oldendorf, Germany

OPEN ACCESS

Edited by:

Diego Minciacci,
University of Florence, Italy

Reviewed by:

Jose Luis Contreras-Vidal,
University of Houston, USA
Takako Fujioka,
Stanford University, USA

*Correspondence:

Eckart O. Altenmüller
eckart.altenmueller@hmtm-hannover.de

Specialty section:

This article was submitted to
Neuroprosthetics,
a section of the journal
Frontiers in Neurology

Received: 17 August 2015

Accepted: 20 June 2016

Published: 30 June 2016

Citation:

Scholz DS, Rohde S, Nikmaram N, Brückner H-P, Großbach M, Rollnik JD and Altenmüller EO (2016) Sonification of Arm Movements in Stroke Rehabilitation – A Novel Approach in Neurologic Music Therapy. *Front. Neurol.* 7:106. doi: 10.3389/fneur.2016.00106

Gross motor impairments are common after stroke, but efficient and motivating therapies for these impairments are scarce. We present an innovative musical sonification therapy, especially designed to retrain patients' gross motor functions. Sonification should motivate patients and provide additional sensory input informing about relative limb position. Twenty-five stroke patients were included in a clinical pre–post study and took part in the sonification training. The patients' upper extremity functions, their psychological states, and their arm movement smoothness were assessed pre and post training. Patients were randomly assigned to either of two groups. Both groups received an average of 10 days ($M = 9.88$; $SD = 2.03$; 30 min/day) of musical sonification therapy [music group (MG)] or a sham sonification movement training [control group (CG)], respectively. The only difference between the two protocols was that in the CG no sound was played back during training. In the beginning, patients explored the acoustic effects of their arm movements in space. At the end of the training, the patients played simple melodies by coordinated arm movements. The 15 patients in the MG showed significantly reduced joint pain ($F = 19.96$, $p < 0.001$) in the Fugl–Meyer assessment after training. They also reported a trend to have improved hand function in the stroke impact scale as compared to the CG. Movement smoothness at day 1, day 5, and the last day of the intervention was compared in MG patients and found to be significantly better after the therapy. Taken together, musical sonification may be a promising therapy for motor impairments after stroke, but further research is required since estimated effect sizes point to moderate treatment outcomes.

Keywords: sonification, stroke, neurorehabilitation, neuroplasticity, music-supported therapy

INTRODUCTION

Stroke is a major cause of mortality and morbidity in both the developed and developing world (1). In Germany, stroke is one of the most common disorders with an estimated 200,000 first events and 66,000 recurrent events in 2008 (2). The World Health Organization stresses the need to collect high quality longitudinal data on rehabilitation and to improve the comparability between studies (3).

The rehabilitation of stroke patients remains a challenge, although there are currently several new training programs under development that aim at improved efficiency and sustainability of stroke rehabilitation (4). Some of the traditional rehabilitation programs lack general acceptance by patients, due to the required endurance and high demands on the patients' cooperation, which sometimes is perceived as a frustrating experience (5). Yet, even the well-established standard physiotherapies do not unambiguously provide evidence of efficacy when it comes to improvement of skilled motor behavior (6–8). Therefore, there is an urgent need for innovative, motivating, and goal-directed training protocols in stroke rehabilitation.

In this article, we present an innovative approach to rehabilitation by retraining the gross motor functions of the affected upper limbs using musical sonification. In an earlier clinical feasibility study (9), we showed how a musical sonification therapy could be applied. The data presented herein were obtained with this method from a larger number of patients. Sonification stands for the usage of non-speech sound representing otherwise not audible information (10). One of the first sonification devices was the Geiger–Müller counter, which detects electromagnetic radiation and communicates a decay by a click sound. In the present study, arm movements were translated into sound. In two earlier studies, we demonstrated the efficacy of a music-supported stroke rehabilitation training utilizing a MIDI drumset and a MIDI piano (11, 12). Stroke patients with some residual abilities to move the arm and the fingers were instructed to play simple tunes (nursery rhymes or folk songs) on either instrument. We could show that auditory sensorimotor circuits established *via* this form of music-supported therapy (MST) promotes beneficial neuroplasticity in stroke patients (13, 14). One of the few constraints of MST was that it was mainly designed to retrain fine-motor skills on MIDI instruments. And it did not provide continuous real-time feedback for the gross motor functions of the arm, which are more frequently impaired in early rehabilitation stages. A real-time movement feedback may be beneficial since it informs the patients about the way they move, not only whether they hit the target or not. With the musical sonification therapy presented here, patients repeatedly train movements with their affected arm in a predefined space. They form associations of their relative armposition in space and the corresponding sound at this specific position. At the end, they play familiar melodies by moving their arm. This musical sonification therapy, therefore, broadens the scope to train stroke patients from an earlier stage on, when still suffering from gross motor dysfunction. Musical sonification will not only contribute to the motivation of the patients due to its playful and positive emotional character, but may also improve motor control, since auditory real-time feedback of the patient's arm movements can be substituted for potentially lost proprioception. There are several preliminary studies with healthy participants that apply non-musical sonification in motor control and the perception of movements (15–17). Schmitz et al. found that sonifying breast stroke movements led to more precise perceptual judgments of movement velocity. They showed that sonification of movements amplifies the human action observation system as indicated by more

pronounced fMRI connectivity patterns between the activation peaks of the left superior and medial posterior temporal regions with the basal ganglia, the thalamus, and frontal regions for movement congruent sonification stimuli. Thus, sonification may be an important method to enhance training and therapy effects in neurological rehabilitation. Chen et al. developed a real-time, multimodal feedback system for stroke rehabilitation (18). This sonification system was tested with stroke patients and showed promising results (19). However, in their design, music was only a passive byproduct of arm movements. That means participants did not play with the sonification sound intentionally. They moved their arms and harmonic music progressions were played back to them. In contrast to that, we developed a musical sonification therapy to train stroke patients to explicitly and consciously play music through intended movements of their affected upper extremity. Thus, we hoped to be able to use the beneficial effects of music on neuroplasticity to facilitate the recovery after stroke (13). Since in other studies repetitive exercise has been shown to be effective (8, 20), our training is of a repetitive nature too. We hypothesize that the auditory cues provided by the sonification may make multimodal associative learning possible where otherwise mere visual and motor learning would have taken place. We assume that patients will benefit in their rehabilitation process from guided attention, necessary concentration, and long-term motivation to play music. Rohrer et al. (21) (see also references therein) describe an increase of several movement smoothness indices in both acute and chronic stroke patients during movement therapy. Hence, the present study additionally investigated changes in movement smoothness over the course of the therapy. After having evaluated an optimal two-dimensional sonification mapping (22), we now present a more detailed analysis of our three-dimensional musical sonification therapy with a larger sample (9).

MATERIALS AND METHODS

Patients

Twenty-five inpatients (11 women, see **Table 1** for details) at the BDH Neurological Rehabilitation Hospital in Hessisch Oldendorf, Germany, participated after giving informed consent. They suffered from a moderate impairment of motor function of the upper extremity after stroke. Inclusion criteria were (a) patients had to have residual function of the affected extremity (i.e., the ability to move the affected arm and the index finger without help from the healthy side), furthermore, (b) an overall Barthel index higher than 50 was required, and (c) patients had to

TABLE 1 | Demographic details of the 25 patients.

	Music group	Control group
Number of subjects	15	10
Gender (m/f)	8/7	6/4
Age (M and SD) range (years)	68.8 ± 13.6 (32–86)	72.2 ± 8.4 (57–85)
Affected arm: right	15	10
Right-handed	15	10
Days after stroke (Median)	32.5	28
Barthel index (M and SD)	56.5 ± 25.3	47 ± 35.9

be right-handed. Patients with other neurological or psychiatric disorders were excluded.

Patients were pseudorandomly assigned to the experimental or to the control group (CG) by the supervisor of the study who was not the experimenter. The experimental group received conventional physiotherapy plus an average of 10 days of a musical sonification training [music group (MG), henceforth].

The CG also received conventional physiotherapy plus a sham sonification movement training with exactly the same movements required as in the sonification study, but with no sound being played back. All patients were native German speakers. The study was approved by the Ethics Review Board of the Hannover Medical School (MHH).

Evaluation of Motor Functions, Stroke Impact, and Movement Smoothness Procedure

Patients were tested pre and post training with a battery of clinical motor function tests and a psychological questionnaire. The test battery consisted of (a) the upper extremity part of the Fugl-Meyer assessment (FMA), still considered the gold standard in evaluating motor recovery after stroke (23, 24). The FMA consists of four bigger subsections. FM.A–D assesses the motor function of the affected arm by checking reflexes, volitional movements, wrist and hand function, and the coordination of the upper extremity. In FM.H, the tactile sensation compared to the non-affected other extremity is assessed. In FM.I, passive joint motion is assessed and FM.J passively measures joint pain. (b) The action research arm test (ARAT) rates upper limb functioning by using observational methods and collecting behavioral data (25, 26). (c) The box and block test (BBT) assesses unilateral gross manual dexterity (27, 28). (d) The nine-hole pegboard test (NHPT) measures finger dexterity (29) and (e) the stroke impact scale (SIS) evaluates the health status following a stroke, including subscales for emotional well-being, memory, thinking, and social participation. The subscales are SIS.1 that asks for physical problems, which may have occurred as a result of the stroke; SIS.2 investigates memory and thinking of the patient; SIS.3 assesses mood and emotions; SIS.4 checks for the communication skills in speaking, reading, and writing; SIS.5 determines how impaired the patient is during daily activities; in SIS.6, the mobility of the patient is investigated; SIS.7 assesses the remaining function of the affected hand; in SIS.8, the patient is asked to which extent he or she is impaired in their social activities; SIS.9 is a self-rating of the patient on how far the stroke recovery has progressed (30, 31).

Movement data were recorded in the MG only using a custom made computer program and two inertial sensors (Xsens, X-MB-XB3), one attached to the fore arm close to the wrist and one attached to the upper arm of the patients. It took approximately 1 h to complete the test battery at the beginning and at the end of the study. Regular training sessions lasted approximately 30 min.

Sonification Training Training

After the pretests the patients received either an average of 10 days ($M = 9.88$; $SD = 2.03$) of musical sonification training (MG), or 10 days of sham sonification training (CG), following the same

protocol as MG but with loudspeakers switched off. The whole procedure followed a standardized protocol to train gross motor functions of the affected right upper extremity in a repetitive manner. Patients were seated as close as possible to the desk with the wooden 3D space frame atop so that the board nearly touched their stomach. Depending on whether in a wheelchair or not, the desk with the 3D space on was adjusted to the individual needs of the patients before starting the training. To get acquainted with the sonification system and the acoustic effects produced by their own arm movements, patients first had to freely move their arm in a three-dimensional sonification space, a wooden cubic frame of 51 cm side length, confined by four vertical beams in the corners of the bottom board (Figure 1). The beams were labeled with the note pitches; the board was subdivided into nine labeled fields for ease of instructions.

Movement sonification was implemented so that upward movements resulted in an ascending C major scale from c' (256 Hz, in Helmholtz pitch notation) to the sixth interval a' (440 Hz). Vertical movements in this space resulted in a change in brightness of sound and, thus, mimicking real musical instrument timbres (modeled by varying the number and amplification of overtones in the sound synthesis; SynthesisToolKit – STK; Cook and Scavone (32); from a rather dull clarinet sound at the very left to saxophone in the middle and a bright sounding bowed instrument at the very right). Movements along the z-axis caused an increase in loudness from proximal to distal. After a first exploration phase to allow for implicitly learning the rules of the musical sonification, more complex exercises followed, demanding incremental degrees of difficulty: At the beginning of each training session patients had to play four upward and downward legato C major scales at position 1 (Figure 1). The same exercise was then repeated at positions 2, 3, 7, and 9. [You can listen to the legato scale playing of a patient at day 1 (Audio S1 in Supplementary Material), day 5

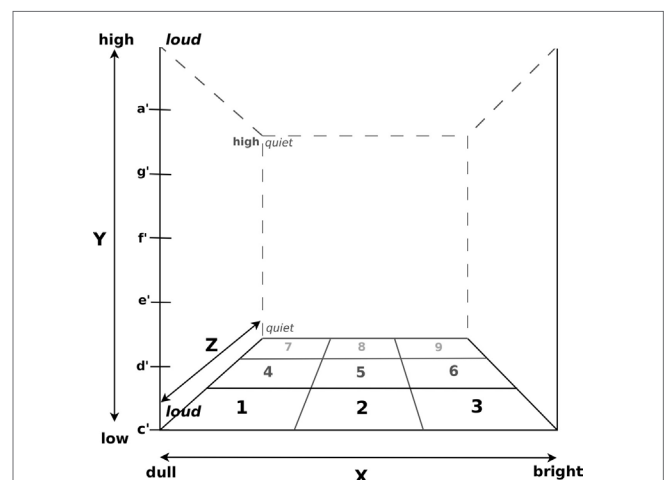


FIGURE 1 | The 3D space defined by a wooden frame. Pitch was mapped onto the y-axis ranging from c' at the bottom to a' at the top. Brightness was mapped onto the x-axis from the left (dull) to the right (bright) and volume onto the z-axis with loudness increasing with increasing proximity. Positions in the x-z plane were labeled for ease of instruction.

(Audio S2 in Supplementary Material), and the last training day (Audio S3 in Supplementary Material)]. These exercises were followed by playing musical intervals by moving the arm faster but as precisely as possible, from c' to d', from c' to e', from c' to f', from c' to g', and from c' to a'. This exercise was repeated four times at position 1 and then likewise at positions 2, 3, 7, and 9. The final goal of the training was to teach patients to play several simple nursery rhymes or other familiar tunes only by moving their affected right arm in the three-dimensional sonification space.

The experimenter gave verbal instructions for the training procedure. Additionally, the experimenter pointed at the visual cues written at the positions on the wooden frame of the 3D space (Figure 1). When playing the melodies, patients could read the required “coordinates” from a sheet provided. All melodies were played vertically, i.e., along the y-axis, at position 1 (Figure 1). Tones could be repeated by dipping the hand horizontally in one direction while maintaining vertical position. Patients always moved their impaired arms by themselves. Arm movements were never guided nor physically supported by the experimenter.

Patients' arm movements were sonified in real time using two small inertial sensors (Xsens, X-MB-XB3) placed at the wrist and the upper arm of the affected limb. The continuous data stream comprising acceleration, rotation, and gravity were transferred via Bluetooth to a laptop and stored for later evaluation in the MG only. The spatial information of the arm movements in 3D space were sonified in real time. The only difference in the training procedure for the sham sonification group (CG) was the muted playback system. Otherwise, exactly the same exercises were carried out during the training sessions.

Data Analysis

Statistical analysis was conducted using R¹ (version 3.2.1) in RStudio Server² (version 0.99.467) on data of the motor function tests and the SIS. Motor test and questionnaire data were preprocessed and tested whether they complied with ANCOVA assumptions. Pretest scores were then used as covariate either in separate ANCOVAs for each response variable, or in the Johnson–Neyman test when applicable: when comparing two groups with respect to their performance before and after treatment, and the assumption of homogenous regression slopes is violated, the Johnson–Neyman technique is used instead of ANCOVA. It allows for the two groups to have different slopes, and tests whether they differ. Additionally, it determines an “area of significance” (33) where the two groups show a statistically meaningful difference in their posttreatment score, after controlling for the differing slopes.

Arm movement data from the four trials at position 1 (Figure 1) from each MG patient were collected and transformed into Cartesian coordinates using a custom-made computer program. Three-dimensional movement trajectories from upward and downward legato C major scales were manually selected and Butterworth lowpass filtered (cut-off 8 Hz) to eliminate tremor

movements. Movement smoothness was calculated as the curvature index κ :

$$\kappa^2 = \frac{(\dot{x}^2 + \dot{y}^2 + \dot{z}^2)(\ddot{x}^2 + \ddot{y}^2 + \ddot{z}^2)(\dot{x}\ddot{x} + \dot{y}\ddot{y} + \dot{z}\ddot{z})^2}{(\dot{x}^2 + \dot{y}^2 + \dot{z}^2)^3}$$

Osu et al. (34) (see *ibid.* for advantages of curvature over other measures for smoothness, like jerk, or snap), representing the inverse of the movement radius for each trajectory point. To synchronize an increase in movement smoothness with an increase in κ and to account for positive skewness of κ , the median negative logarithm for each movement segment was taken.

These values were then compared between the MGs' first, fifth, and last therapy sessions using Friedman's test, followed by Wilcoxon's signed rank test as *post hoc* test to determine an increase in smoothness during (day 5) or after the end of the sonification training. The kernel density estimation of the transformed κ (see Figure 2) was calculated with a data-driven kernel suggested by Sheather and Jones (35) [cited in Ref. (36)].

RESULTS

Motor Tests

The main results of this study are depicted in Table 2. Although the ANCOVA group comparisons for ARAT, BBT, and NHPT were non-significant, MG patients showed a significantly higher improvement compared to the CG in the subscale FMJ of the

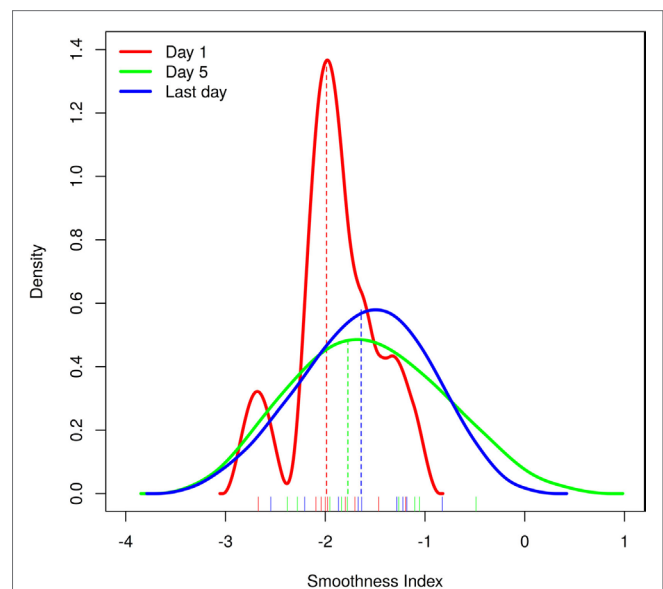


FIGURE 2 | Movement smoothness at day 1, day 5, and the last day of the intervention was compared in the treatment group (MG) patients and found to be significantly better after the therapy. Shown here are the kernel density estimates of the MG smoothness measures at day 1 (red line), day 5 (green line), and the last day of the intervention (blue line; see Materials and Methods and Results for details) with the smoothness index shown on the x-axis and the density estimation on the y-axis. Medians of the corresponding distribution are depicted as dashed lines in corresponding color.

¹<http://www.r-project.org>

²<http://www.rstudio.com>

TABLE 2 | ANCOVA group comparison results for the motor tests and the stroke impact scale.

Test (pretest: group interaction)	Sum Sq	Mean Sq	F	p
ARAT	6.635	6.635	0.1102	0.7432
BBT	0.7481	0.7481	0.01645	0.8994
NHPT	1694	1694	2.206	0.1558
FM.A-D (motor function)	0.3989	0.3989	0.01179	0.9146
FM.H (sensation)	0.8866	0.8866	1.641	0.2155
FM.I (passive joint motion)	1.775	1.775	0.363	0.5536
FM.J (joint pain)	–	–	–	see Results and Figure 2
SIS (total)	2091	2091	4.63	0.0445*
SIS.1 (physical problems)	0.7875	0.7875	0.002578	0.9601
SIS.2 (memory and thinking)	69.03	69.03	0.2237	0.6426
SIS.3 (mood and emotions)	159.3	159.3	0.4738	0.5011
SIS.4 (communication skills)	48.42	48.42	0.1189	0.7347
SIS.5 (daily activities)	168.6	168.6	0.6058	0.4477
SIS.6 (mobility)	39.39	39.39	0.1303	0.7229
SIS.7 (affected hand function)	878	878	4.278	0.0552
SIS.8 (social activities)	833.5	833.5	0.8634	0.3666
SIS.9 (stroke recovery rating)	207.4	207.4	0.8629	0.3667

ARAT, action research arm test; BBT, box and block test; NHPT, nine-hole pegboard test; FM, Fugl-Meyer assessment; SIS, stroke impact scale.

Significant group differences at the $\alpha = 0.05$ – level are indicated by *.

FMA [$F(1,21) = 21.23$, $p < 0.05$], as shown by the Johnson–Neyman technique, suggesting that they perceived reduced joint pain after the training. The two groups' corrected pre–post differences differed significantly for pretest scores below 20, suggesting an advantage for the sonification therapy over sham treatment for patients scoring low in the FM.J subscale (Figure 3).

Stroke Impact Scale

The SIS total value was significantly higher for the MG as compared to the CG, after the training ($F = 4.63$, $p = 0.0445$). But, after correcting for the random group difference prior to the training, there was no more pretest: group interaction to be found [$F(1,21) = 3.83$, $p > 0.05$]. A pre test group difference for SIS.2 [two-sided paired t -test ($t = 2.229$, $df = 19.92$, $p = 0.0375$)] was detected. For that reason this specific subscale was not included into further evaluation. The subscale SIS.7 showed a trend ($F = 4.278$, $p = 0.0552$) toward better hand function of the MG patients after the training. All other ANCOVA group comparisons were non-significant.

Arm Movement Smoothness

Movement data from two MG patients could not be retrieved due to technical failure of the recording system. For the remaining 13 patients, trajectory smoothness was derived at day 1, day 5, and the last day of their training. A significant difference between training days was shown with Friedman's test ($\chi^2 = 6.222$, $df = 2$, $p = 0.0445$; see Figure 2), and Wilcoxon *post hoc* tests showed a significantly higher movement smoothness at the last day as compared to the first training day ($V = 7$, $p = 0.037$). But after applying the Bonferroni–Holms correction this effect was not significant anymore ($p = 0.074$). This could be due to a very subtle effect for which the Wilcoxon test is not powerful enough to detect. Also the small number of subjects should be taken into account.

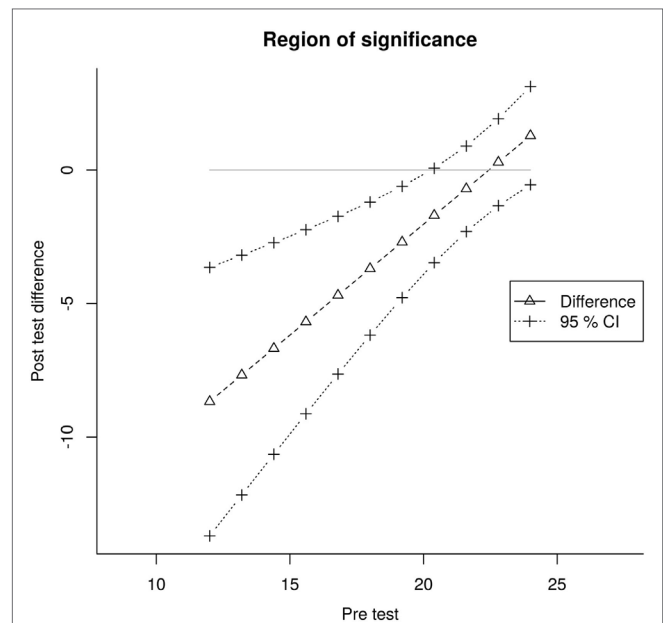


FIGURE 3 | Fugl-Meyer joint pain subscale results. Taking into account the unequal regression slopes of CG and MG, the Johnson–Neyman technique estimates the two groups' post-test difference (triangles, dashed straight line) from the joint pretest values and determines upper and lower confidence intervals (CI; crosses, dotted lines). The triangles and crosses do not represent individual data points but estimated values. The CIs and the solid straight line at $y = 0$, confine the area where the two groups differed significantly in their post-test Fugl-Meyer Joint pain score. Compared to CG patients, MG patients with a pretest score below 20 seemed to benefit from the additional sonification therapy.

DISCUSSION

The results of this clinical sonification study show that a musical sonification therapy may be a promising new way of treating motor impairments after stroke. Musical sonification therapy may even improve psychological well-being after stroke. The 15 patients of the musical sonification group improved significantly compared to the movement training group in the Fugl-Meyer subscale assessing joint pain. Reduced pain after a motor training or a mechano-acoustic vibration therapy for stroke patients was also shown by Lee and Kim (37) and Constantino et al. (38). The MG patients of our study also showed a trend to regain a better hand function in the SIS after the training. In addition to the motor domain, the SIS assesses the emotional state of the patient, memory, and social participation. Movement smoothness pre and post intervention was found to be significantly better after the therapy in the MG. This is in line with the findings of Rohrer et al. (21) who showed an increased movement smoothness with a robotic therapy device in four of five measures in 31 patients recovering from stroke. In contrast, the patients of our CG, receiving only a “sham” movement training without musical feedback improved very little and non-significantly in some of the tests. Thus, we assume that the musical aspect plays an important role in the sonification therapy. However, in this study, we did not control whether it is the musical aspect of sonification or just any sound information provided by the sonification. Furthermore,

different motor tests should be included in future research in order to prevent floor (NHPT) and ceiling effects (ARAT), which were found in some of the tests in this study. Of course, as we only present a small clinical trial with limited statistical power, results need to be verified with a larger group of patients.

The novel aspect of our approach is that we encourage the patients in the musical sonification group to actively play and create music by moving their arms. This way, music was not only a byproduct of, e.g., a grasping motion. Instead, movements resembled more a novel musical instrument patients were starting to play. This musical instrument was sometimes compared to a theremin by professional musicians. Hence, our sonification training was designed to resemble a music lesson rather than shaping a movement during sound playback. Furthermore, we used musical stimuli, such as a musical major scale with discrete intervals and timbre parameters derived from the sound characteristics of acoustical musical instruments, as opposed to the widely used sound mappings where tone pitch is scaled continuously and rather artificial sounds are applied (39). The main idea underlying our hypothesis was that participants could improve control of arm positions in space *via* associative learning, leading to associating a given relative arm position with a specific musical sound. This sound-location association may then substitute the frequently declined, or even lost, proprioception. Additionally, the arm movement trajectories from outset to the target position were audible as well. Thus, multimodal learning might have taken place because patients received sound as an additional parameter supplying information. One could speculate that this multimodal learning could help to close the sensorimotor loop, which may have been affected by the stroke.

In view of the clinical application, reduced gross motor functions of the arm and reduced proprioception are common disabilities in stroke patients (40). Hence, the advantages of continuous real-time musical feedback are obvious: the therapy, therefore, aims at retraining gross motor movements of the arm, which are the most disabling challenges in early rehabilitation of stroke. Second, real-time sonification may substitute deficits in proprioception of the arm, which frequently are a consequence of stroke.

Finally, this form of therapy is highly motivating and could thus enhance motor functions and the emotional well-being in

some patients, maybe through the creative, playful character of this musical sonification device (41–44).

To summarize, we have developed and tested a novel musical sonification therapy in a group of patients, which supports learning effects in auditory sensory–motor integration. Now, multimodal learning of spatial, motor, auditory, and proprioceptive information in rehabilitation of arm motor control in stroke patients needs to be evaluated in a larger multicentered representative randomized controlled clinical trial.

AUTHOR CONTRIBUTIONS

All authors listed have made substantial, direct, and intellectual contribution to the work and approved it for publication.

ACKNOWLEDGMENTS

This work was supported by the cortexplorer program of the Hertie Foundation for Neurosciences (<http://www.ghst.de/en/neurosciences/cortexplorer/>). This paper is part of a Ph.D. of DS at the Center for Systems Neurosciences (ZSN), Hanover, Germany. The authors wish to thank Prof. Alfred Effenberg, Dr. Gerd Schmitz (both at Leibniz University Hanover, Institute of Sport Science), and Prof. Holger Blume (Leibniz University Hanover, Institute of Microelectronic Systems) for important insights while collaborating in an earlier joint European Regional Development Fund project on movement sonification. We also wish to thank Benjamin Krüger and Dr. H-PB for implementing the real-time 3D sonification for the clinical study and Prof. Jörg Remy, Till Engert, and Steffen Rummel (SRH Hochschule der populären Künste GmbH, Berlin, Germany, University of Applied Sciences) for further development. Finally, we would like to thank the patients and the staff of the BDH Klinik Hessisch Oldendorf for their ongoing support and patience.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at <http://journal.frontiersin.org/article/10.3389/fneur.2016.00106>

REFERENCES

- Black M, Wang W, Wang W. Ischemic stroke: from next generation sequencing and GWAS to community genomics? *OMICS* (2015) 19:451–60. doi:10.1089/omi.2015.0083
- Wiedmann S, Heuschmann PU, Hillmann S, Busse O, Wiethölter H, Walter GM, et al. The quality of acute stroke care – an analysis of evidence – based indicators in 260 000 patients. *Dtsch Arztebl Int* (2014) 111:759–65. doi:10.3238/arztebl.2014.0759
- Sauzet O, Kleine M, Menzel-Begemann A, Exner A-K. Longitudinal randomised controlled trials in rehabilitation post-stroke: a systematic review on the quality of reporting and use of baseline outcome values. *BMC Neurol* (2015) 15:99. doi:10.1186/s12883-015-0344-y
- Lohse KR, Hilderman CGE, Cheung KL, Tatla S, der Loos HFMV. Virtual reality therapy for adults post-stroke: a systematic review and meta-analysis exploring virtual environments and commercial games in therapy. *PLoS One* (2014) 9:e93318. doi:10.1371/journal.pone.0093318
- Pulman J, Buckley E, Clark-Carter D. A meta-analysis evaluating the effectiveness of two different upper limb hemiparesis interventions on improving health-related quality of life following stroke. *Top Stroke Rehabil* (2013) 20:189–96. doi:10.1310/tsr2002-189
- Hakkenes S, Keating JL. Constraint-induced movement therapy following stroke: a systematic review of randomised controlled trials. *Aust J Physiother* (2005) 51:221–31. doi:10.1016/S0004-9514(05)70003-9
- Peurala SH, Kantanen MP, Sjögren T, Paltamaa J, Karhula M, Heinonen A. Effectiveness of constraint-induced movement therapy on activity and participation after stroke: a systematic review and meta-analysis of randomized controlled trials. *Clin Rehabil* (2012) 26:209–23. doi:10.1177/0269215511420306
- Stevenson T, Thalman L, Christie H, Poluha W. Constraint-induced movement therapy compared to dose-matched interventions for upper-limb dysfunction

- in adult survivors of stroke: a systematic review with meta-analysis. *Physiother Can* (2012) 64:397–413. doi:10.3138/ptc.2011-24
9. Scholz DS, Rohde S, Großbach M, Rollnik J, Altenmüller E. Moving with music for stroke rehabilitation: a sonification feasibility study. *Ann N Y Acad Sci* (2015) 1337:69–76. doi:10.1111/nyas.12691
 10. Kramer G, Walker B, Bonebright T. *Sonification Report: Status of the Field and Research Agenda*. Santa Fe, NM: National Science Foundation, International Community for Auditory Display (ICAD) (1999).
 11. Schneider S, Schönle PW, Altenmüller E, Münte TF. Using musical instruments to improve motor skill recovery following a stroke. *J Neurol* (2007) 254:1339–46. doi:10.1007/s00415-006-0523-2
 12. Altenmüller E, Marco-Pallares J, Münte TF, Schneider S. Neural reorganization underlies improvement in stroke-induced motor dysfunction by music-supported therapy. *Ann N Y Acad Sci* (2009) 1169:395–405. doi:10.1111/j.1749-6632.2009.04580.x
 13. Rojo N, Amengual J, Juncadella M, Rubio F, Camara E, Marco-Pallares J, et al. Music-supported therapy induces plasticity in the sensorimotor cortex in chronic stroke: a single-case study using multimodal imaging (fMRI-TMS). *Brain Inj* (2011) 25:787–93. doi:10.3109/02699052.2011.576305
 14. Amengual JL, Rojo N, de Las Heras MV, Marco-Pallares J, Grau-Sánchez J, Schneider S, et al. Sensorimotor plasticity after music-supported therapy in chronic stroke patients revealed by transcranial magnetic stimulation. *PLoS One* (2013) 8:e61883. doi:10.1371/journal.pone.0061883
 15. Scheef L, Boecker H, Daamen M, Fehse U, Landsberg MW, Granath D-O, et al. Multimodal motion processing in area V5/MT: evidence from an artificial class of audio-visual events. *Brain Res* (2009) 1252:94–104. doi:10.1016/j.brainres.2008.10.067
 16. Schmitz G, Mohammadi B, Hammer A, Heldmann M, Samii A, Münte TF, et al. Observation of sonified movements engages a basal ganglia frontocortical network. *BMC Neurosci* (2013) 14:32. doi:10.1186/1471-2202-14-32
 17. Schmitz G, Kroeger D, Effenberg AO. A mobile sonification system for stroke rehabilitation. Sonification as concurrent augmented feedback for motor skill learning and the importance of mapping design. *The 20th International Conference on Auditory Display (ICAD-2014)*. New York (2014).
 18. Chen Y, Huang H, Xu W, Wallis R, Sundaram H, Rikakis T, et al. The design of a real-time, multimodal biofeedback system for stroke patient rehabilitation. *Proceedings of the 14th ACM Multimedia Conference (MM '06)*. Santa Barbara, CA (2006).
 19. Wallis I, Ingalls T, Rikakis T, Olsen L, Chen Y, Xu W, et al. Real-time sonification of movement for an immersive stroke rehabilitation environment. *Proceedings of the 13th International Conference on Auditory Display*. Montréal, Canada (2007).
 20. Taub E, Uswatte G, Pidikiti R. Constraint-induced movement therapy: a new family of techniques with broad application to physical rehabilitation – a clinical review. *J Rehabil Res Dev* (1999) 36:237–51.
 21. Rohrer B, Fasoli S, Krebs HI, Hughes R, Volpe B, Frontera WR, et al. Movement smoothness changes during stroke recovery. *J Neurosci* (2002) 22:8297–304.
 22. Scholz DS, Wu L, Schneider J, Pirzer J, Grossbach M, Rollnik JD, et al. Sonification as a possible stroke rehabilitation strategy. *Front Neurosci* (2014) 8:332. doi:10.3389/fnins.2014.00332
 23. Crow JL, van der Wel BCH. Hierarchical properties of the motor function sections of the Fugl-Meyer assessment scale for people after stroke: a retrospective study. *Phys Ther* (2008) 88:1554–67. doi:10.2522/ptj.20070186
 24. Woodbury ML, Velozo CA, Richards LG, Duncan PW, Studenski S, Lai S-M. Longitudinal stability of the Fugl-Meyer assessment of the upper extremity. *Arch Phys Med Rehabil* (2008) 89:1563–9. doi:10.1016/j.apmr.2007.12.041
 25. Platz T, Pinkowski C, van Wijk J, Kim I-H, di Bella P, Johnson G. Reliability and validity of arm function assessment with standardized guidelines for the Fugl-Meyer test, action research arm test and box and block test: a multicentre study. *Clin Rehabil* (2005) 19:404–11. doi:10.1191/0269215505cr832oa
 26. Nijland R, van Wegen E, Verbunt J, van Wijk R, van Kordelaar J, Kwakkel G. A comparison of two validated tests for upper limb function after stroke: the wolf motor function test and the action research arm test. *J Rehabil Med* (2010) 42:694–6. doi:10.2340/16501977-0560
 27. Canny ML, Thompson JM, Wheeler MJ. Reliability of the box and block test of manual dexterity for use with patients with fibromyalgia. *Am J Occup Ther* (2009) 63:506–10. doi:10.5014/ajot.63.4.506
 28. Chen H-M, Chen CC, Hsueh I-P, Huang S-L, Hsieh C-L. Test-retest reproducibility and smallest real difference of 5 hand function tests in patients with stroke. *Neurorehabil Neural Repair* (2009) 23:435–40. doi:10.1177/1545968308331146
 29. Grice KO, Vogel KA, Le V, Mitchell A, Muniz S, Vollmer MA. Adult norms for a commercially available nine hole peg test for finger dexterity. *Am J Occup Ther* (2003) 57:570–3. doi:10.5014/ajot.57.5.570
 30. Duncan PW, Bode RK, Lai SM, Perera S; Glycine Antagonist in Neuroprotection Americans Investigators. Rasch analysis of a new stroke-specific outcome scale: the stroke impact scale. *Arch Phys Med Rehabil* (2003) 84:950–63. doi:10.1016/S0003-9993(03)00035-2
 31. Lin K-C, Fu T, Wu C-Y, Hsieh Y-W, Chen C-L, Lee P-C. Psychometric comparisons of the stroke impact scale 3.0 and stroke-specific quality of life scale. *Qual Life Res* (2010) 19:435–43. doi:10.1007/s11136-010-9597-5
 32. Cook PR, Scavone G. The synthesis toolkit (stk). *Proceedings of the International Computer Music Conference* (1999). p. 164–6.
 33. Kowalski CJ, Schneiderman ED, Willis SM. Assessing the effect of a treatment when subjects are growing at different rates. *Int J Biomed Comput* (1994) 37:151–9. doi:10.1016/0020-7101(94)90131-7
 34. Osu R, Ota K, Fujiwara T, Otake Y, Kawato M, Liu M. Quantifying the quality of hand movement in stroke patients through three-dimensional curvature. *J Neuroeng Rehabil* (2011) 8:62. doi:10.1186/1743-0003-8-62
 35. Sheather SJ, Jones MC. A reliable data-based bandwidth selection method for kernel density estimation. *J R Stat Soc Ser B Methodol* (1991) 53:683–90. doi:10.2307/2345597
 36. Venables WN, Ripley BD. In: Chambers J, Eddy W, Härdle W, Sheather SJ, Tierney L, editors. *Modern Applied Statistics with S*. New York: Springer (2002).
 37. Lee HS, Kim JU. The effect of self-directed exercise using a task board on pain and function in the upper extremities of stroke patients. *J Phys Ther Sci* (2013) 25:963–7. doi:10.1589/jpts.25.963
 38. Constantino C, Galuppo L, Romiti D. Efficacy of mechano-acoustic vibration on strength, pain, and function in poststroke rehabilitation: a pilot study. *Top Stroke Rehabil* (2014) 21:391–9. doi:10.1310/tsr2105-391
 39. Dubus G, Bresin R. A systematic review of mapping strategies for the sonification of physical quantities. *PLoS One* (2013) 8:e82491. doi:10.1371/journal.pone.0082491
 40. Sacco RL, Bello JA, Traub R, Brust JC. Selective proprioceptive loss from a thalamic lacunar stroke. *Stroke* (1987) 18:1160–3. doi:10.1161/01.STR.18.6.1160
 41. Koelsch S. Investigating emotion with music: neuroscientific approaches. *Ann N Y Acad Sci* (2005) 1060:412–8. doi:10.1196/annals.1360.034
 42. Eschrich S, Münte TF, Altenmüller EO. Unforgettable film music: the role of emotion in episodic long-term memory for music. *BMC Neurosci* (2008) 9:48. doi:10.1186/1471-2202-9-48
 43. Koelsch S. A neuroscientific perspective on music therapy. *Ann N Y Acad Sci* (2009) 1169:374–84. doi:10.1111/j.1749-6632.2009.04592.x
 44. Bood RJ, Nijssen M, van der Kamp J, Roerdink M. The power of auditory-motor synchronization in sports: enhancing running performance by coupling cadence with the right beats. *PLoS One* (2013) 8:e70758. doi:10.1371/journal.pone.0070758

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.

Copyright © 2016 Scholz, Rohde, Nikmaram, Brückner, Großbach, Rollnik and Altenmüller. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Sensori-Motor Learning with Movement Sonification: Perspectives from Recent Interdisciplinary Studies

Frédéric Bevilacqua^{1*}, Eric O. Boyer^{1,2}, Jules Françoise¹, Olivier Houix¹, Patrick Susini¹, Agnès Roby-Brami² and Sylvain Hanneton³

¹ STMS Ircam-Centre National de la Recherche Scientifique-UPMC, Paris, France, ² UMR7222 ISIR - Université Pierre et Marie Curie, Paris, France, ³ UMR 8242 Centre National de la Recherche Scientifique - Université Paris Descartes, Paris, France

OPEN ACCESS

Edited by:

David Rosenboom,
California Institute of the Arts, USA

Reviewed by:

Rolando Grave De Peralta Menendez,
Geneva Electrical Neuroimaging
Group, Switzerland
Martin Lotze,
University of Greifswald, Germany

*Correspondence:

Frédéric Bevilacqua
frederic.bevilacqua@ircam.fr

Specialty section:

This article was submitted to
Neuroprosthetics,
a section of the journal
Frontiers in Neuroscience

Received: 30 April 2016

Accepted: 08 August 2016

Published: 25 August 2016

Citation:

Bevilacqua F, Boyer EO, Françoise J, Houix O, Susini P, Roby-Brami A and Hanneton S (2016) Sensori-Motor Learning with Movement Sonification: Perspectives from Recent Interdisciplinary Studies. *Front. Neurosci.* 10:385. doi: 10.3389/fnins.2016.00385

This article reports on an interdisciplinary research project on movement sonification for sensori-motor learning. First, we describe different research fields which have contributed to movement sonification, from music technology including gesture-controlled sound synthesis, sonic interaction design, to research on sensori-motor learning with auditory-feedback. In particular, we propose to distinguish between *sound-oriented* tasks and *movement-oriented* tasks in experiments involving interactive sound feedback. We describe several research questions and recently published results on movement control, learning and perception. In particular, we studied the effect of the auditory feedback on movements considering several cases: from experiments on pointing and visuo-motor tracking to more complex tasks where interactive sound feedback can guide movements, or cases of sensory substitution where the auditory feedback can inform on object shapes. We also developed specific methodologies and technologies for designing the sonic feedback and movement sonification. We conclude with a discussion on key future research challenges in sensori-motor learning with movement sonification. We also point out toward promising applications such as rehabilitation, sport training or product design.

Keywords: sonification, movement, learning, sensori-motor, sound design, interactive systems

1. INTRODUCTION

The idea of using auditory feedback in interactive systems has recently gained momentum in different research fields. In applications such as movement rehabilitation, sport training or product design, the use of auditory feedback can complement visual feedback. It reacts faster than the visual system and can continuously be delivered without constraining the movements. In particular, movement sonification systems appear promising for sensori-motor learning in providing users with auditory feedback of their own movements. Generally, sonification is defined as the use of non-speech audio to convey information (Kramer et al., 1999). Nevertheless, research on movement sonification for sensori-motor learning has been scattered in totally different research fields. On the one hand, most neuroscience and medical experiments have made use of very basic interactive systems, with little concern for sound design and the possible types of sonification. On the other hand, novel sound/music interactive technologies have been developed toward artistic practices, gaming or sound design, with little concern for sensori-motor learning.

Clearly, there has been a lack of overlap between all these different disciplines, which would each benefit from more exchanges on tools, methods and knowledge. This rationale motivated us to initiate an interdisciplinary project that focused on sensorimotor learning in movement-based sound interactive systems¹. Overall, this body of work, that we partially present in this *Perspective* paper, allowed us to establish general principles on movement sonification and to formalize fundamental questions that should be addressed in future research.

The paper is structured as follows. First, we recall related works in interactive music systems, human-computer design, sonic interaction design, and movement sonification for sport and rehabilitation. Second, we report on the questions and results we obtained in our project. Third, we discuss key research questions that should open a broad discussion.

2. INTERSECTING RESEARCH ON SOUND, MOVEMENT, AND INTERACTION

Different types of interactive systems can produce sound based on human movement. Movement parameters are typically obtained from motion capture systems—such as optical motion capture, cameras, or inertial measurement units—and the sound can be rendered continuously using various types of real-time sound synthesis methods. In this paper, we restrain the discussion to interactive systems built with a deterministic *mapping* between movement and sound parameters (Dubus and Bresin, 2013). As described in the next section, these technologies have been developed in different contexts focusing either on sound or on movement aspects.

2.1. Movement-Based Interfaces for Sound Production and Expression

The music technology research community has long been concerned with gestural and bodily control of sound². Technologies for movement capture, analysis, recognition and interaction design have been developed and reported in the *sound and music computing* literature. In particular, the so-called *mapping* between movement parameters and sound synthesis parameters has been formalized and categorized (Hunt et al., 2003; Wanderley and Depalle, 2004). Methods and tools have been developed and are available for research communities (Leman, 2008; Schnell et al., 2009; Fiebrink and Cook, 2010; Bevilacqua et al., 2011). Surprisingly, though, sensori-motor learning has been rarely studied explicitly in such electronic or digital musical instruments.

In musical applications, the goal of the interaction is often to produce a specific *sound*. Therefore, we propose to refer to such tasks as *sound-oriented tasks*, during which the focus of the user's attention is drawn toward the sound produced by the interactive system. In general, the users must adapt their movement to the interface and gain expertise to achieve high control of sound production and musical expressivity. We explicitly used the

concept of *sound-oriented task* to demonstrate how auditory feedback can be used in sensori-motor adaptation (Boyer et al., 2014). This important point will be further discussed in Section 3.

2.2. Movement Sonification for Sensori-Motor Learning

On the other side of the spectrum lie works on sensori-motor learning *per se*. The large majority of neuroscience papers on the human motor system only deals with visual, haptic and vestibular sensory inputs and rarely mentions the auditory modality. Historically, most papers reporting on auditory-motor mechanisms are concerned speech learning and production. Due to promising applications in movement learning, mostly in sport and rehabilitation, there has recently been an increasing number of studies showing the potential interest of auditory feedback (see Sigrist et al., 2013 for a review). Nevertheless, the technology used in these studies remains generally rudimentary, considering only simple movement-to-sound mapping using parameters such as audio energy and pitch (Dubus and Bresin, 2013).

Generally, the tasks described in such research correspond to what we call *movement-oriented tasks*, where the attention (and the instruction) is put on the movement itself. Movements are thus designed to exhibit specific characteristics (e.g., exercises in rehabilitation) or fully constrained by the application (e.g., specific movements that must be mastered in a given sport). The auditory exteroceptive concurrent feedback either informs whether the movement is properly executed (KR: Knowledge of Results) or how it is executed (KP: Knowledge of Performance) (Schmidt, 1988; Cirstea et al., 2006).

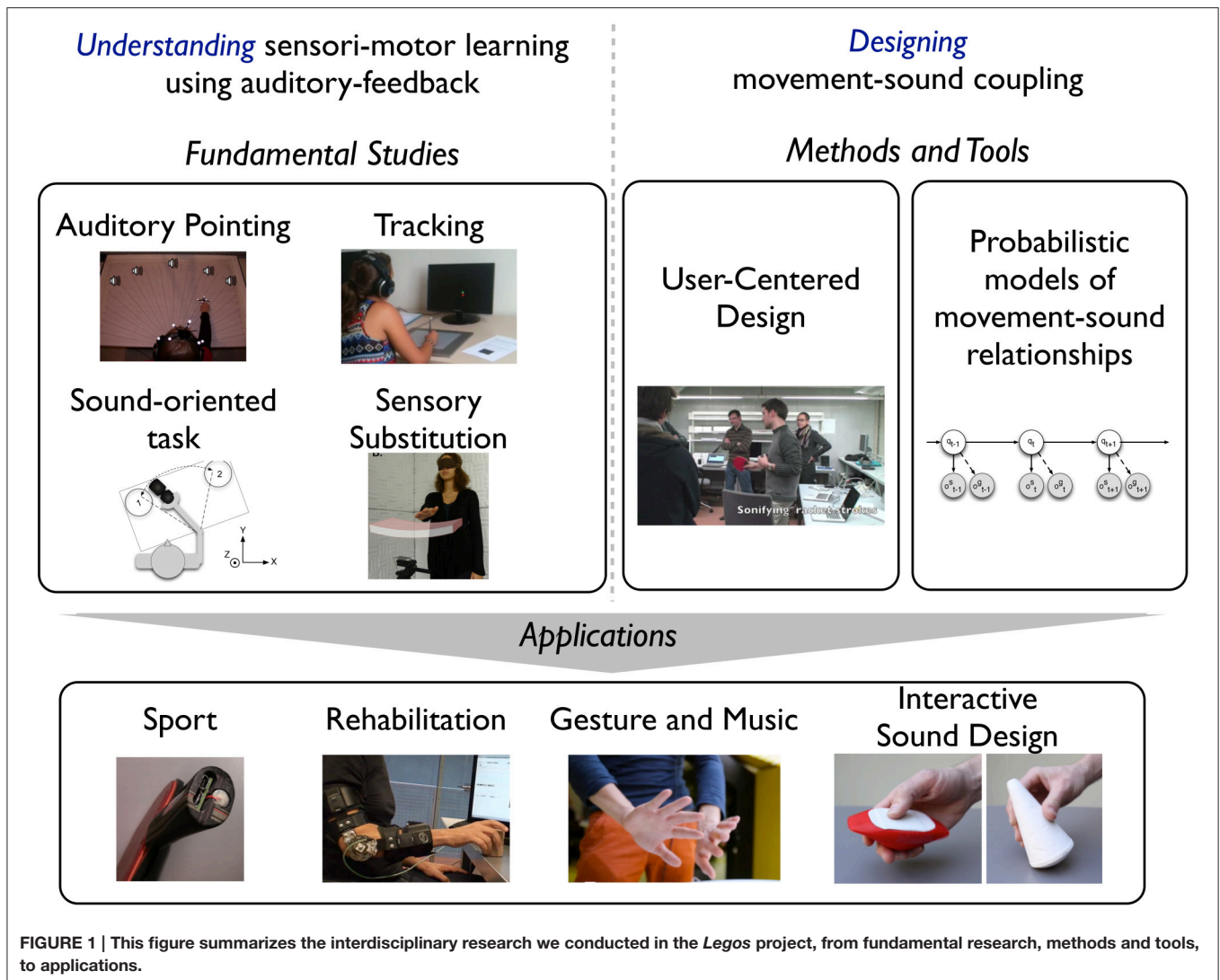
It is worth noting that the beneficial effect of music therapy for sensori-motor rehabilitation is now well recognized, particularly in stroke patients (Ripollés et al., 2015) and in other neurological diseases such as Parkinson (Thaut, 2015) where the synchronization of rhythmic auditory cues is proven to improve gait and motor activity (Schiavio and Altenmüller, 2015). The effect of music training is probably not only due to motivation and psychosocial factors linked with community practicing but also to the multisensory feedback linked to musical motor actions and the brain plasticity it induces (Schlaug, 2015). Rhythmic cues are an important support during music execution (Schneider et al., 2007). Less is known about the effect of continuous sound or music feedback on discrete movements of the upper-limb. Recent evidence suggests that such tasks performed with continuous sound feedback could improve the performance and facilitate learning (Rosati et al., 2012). Thus, sonification has been proposed during rehabilitation, in isolation or to augment other exercise based methods (Scholz et al., 2014).

3. RESEARCH QUESTIONS AND FUNDAMENTAL STUDIES

We report below the different research questions we have investigated (see **Figure 1**), covering fundamentals studies, methods and tool development. In particular, we describe in this section the fundamental and methodological aspects of

¹Legos project, see <http://legos.ircam.fr>.

²See for example the community related to the NIME conferences (New Interfaces for Musical Expression) (Bevilacqua et al., 2013).



our experimental studies on the influence of continuous and concurrent auditory feedbacks.

3.1. Can Auditory Feedback Modify and/or Improve Movement Performance?

We investigated movement sonification in a visuo-motor tracking task (Boyer, 2015). In this case, we compared the sonification of three different variables: the visual target presented on a screen, the participant's pointer (i.e., movement hand) and the online error between the target and the pointer. In the three conditions, we found a positive effect of the auditory feedback for improving the tracking accuracy. Interestingly, the sonification of the hand movement seems in this case to favor an increase of the average movement energy, even after a long exposure to the task, and to improve retention.

Another study focused on a pointing task to an unseen spatialized auditory target, in which we evaluated the role of the target sound duration and the movement sonification (Boyer et al., 2013a). A long duration target presentation improved

the pointing accuracy, highlighting the contribution of neuronal integration processes of the auditory information. The hand movement sonification was not found useful in this case, which might be explained by the complexity of the perception of two different spatialized sound sources (target and sonified hand).

Tajadura-Jiménez et al. (2014) also showed that in a touch task, interactive auditory feedback could modify the user's behavior, precisely the hand velocity and finger pressure. Finally, we found that movement sonification could be used to stabilize the performance of newly learned gestures (Françoise et al., 2016).

3.2. Can the Presentation of a Specific Sound Serve to Specify a Movement?

We investigated whether auditory feedback can be designed for guiding users in the performance of a specific movement. For example, we built an interactive system where participants had to discover how to move an object on a table using solely the auditory feedback (Boyer et al., 2014). They were asked to pay attention to specific sound features, which corresponds to

what we define as a *sound-oriented task*. The whole movement was continuously sonified with sound properties depending on the error between the performed and targeted velocity profiles. Globally, we found that such an auditory feedback was effective to guide participants to learn to perform a predefined velocity profile. Also, after a first stage of exposure with a fixed velocity profile, movement adaptation was also observed when modifying the target profile (without informing the participants). This confirmed similar results obtained by Rath and Rocchesso (2005) and Rath and Schleicher (2008).

Importantly, a large variability was found between participants, which could be partially explained by the fact that such a task (i.e., performing a specific movement being guided by sound feedback) was totally unfamiliar to the participants. It is also likely that each subject exhibits different audio-motor ability.

3.3. Sensory Substitution: Can Sound Replace Another Modality?

We explored a case of sensory substitution where participants had to estimate the curvature of a virtual shape from auditory feedback (Boyer et al., 2015; Hanneton et al., 2015). In the experiment, users received continuous auditory feedback when ‘touching’ the virtual surface. While the accuracy of participants’ estimation of the curvature was inferior to published results with tangible surfaces, we found that the auditory feedback can be effective for such a task, especially when the sound responds to the hand velocity. Most interestingly, different strategies on the use of the movement-sound interaction were observed between users: some persons tend to gently tap perpendicularly to the surface, while others prefer to explore the surface with large lateral movements. This also here indicates large discrepancies between participants in transferring movement sonification information.

3.4. Can (Interactive) Sound Alter Perception and Emotion?

As we just reported, people can use the auditory channel to adapt their movements. Nevertheless, little is known about the subjective changes (for the users’ perception) of the sound and the movement, as well as possible change in their emotional state. In a tapping task with an artificial auditory feedback, the emotional response has been found to be affected by the congruence between the sound energy and the tapping dynamics (Tajadura-Jimenez et al., 2015). In particular, audio-motor incongruences can lead to unpleasant experiences, which shows that expectation of the user for the audio feedback might be crucial for integrating the feedback. The artificial sound feedback of touch can also alter the texture perception, such as the coldness or material type (Tajadura-Jiménez et al., 2014).

Beyond fundamental neuroscience research, such investigations—that confirm other studies on multimodal sensory integration (Zampini and Spence, 2004; Maes et al., 2014),—have high impact potential applications for diminishing pain (Singh et al., 2016) or effort perception (Fritz et al., 2013).

4. DESIGNING MOVEMENT-SOUND INTERACTION

The various results we gather indicates that the effect of the sonification might depend on specific aspects of the interaction design, which confirmed previous studies. In particular, the sound, and more specifically the congruence between the movement and sound, can strongly modify the user experience and therefore the effectiveness of the feedback. In Castiello et al. (2010), it was shown that the effect on the action of reaching and grasping an object is favored (in terms of movement duration) for congruent conditions, when the sound corresponds to the material covering the object to grasp, compared to incongruent conditions. In Susini et al. (2012), congruent sound-action conditions in terms of naturalness were found to be determinant in the appraisal of the sounds of everyday objects. These findings call for improving methodologies for the design of such sound interactive systems.

Building upon previous results (Rocchesso et al., 2009; Franinović and Serafin, 2013), we developed user-centered methodologies based on participatory design workshops. A central idea was to explore strategies combining the analysis of various objects’ affordances with established sound and action taxonomies (Houix et al., 2014, 2015). The design of the movement-sound can be leveraged by taking advantages of users expectancy on the auditory feedback. In such a case, we refer to *ecological* relationships between action and sound.

The notion of object affordances can also be extended to sound, by questioning reversely which movement could be performed to match a given sound (Caramiaux et al., 2014a,b). Following such premises, we developed a method called *mapping by demonstration*, that allows to program an interactive systems based on movement performed while listening to a sound (Françoise, 2015). Such an approach can leverage known association between movement and sound feedback, and is particularly adapted for user-centered methodology in the design of interactive systems.

5. DISCUSSION AND FUTURE RESEARCH CHALLENGES

We discuss here some of the research questions we mentioned in the previous sections, and propose new steps that we think as central for future research.

First, auditory feedback can be designed to convey different type of information. A first approach is to inform continuously on the error between the performed movement and a “normal” movement. In this case, the learning or adaptation is explicit. The alternative approach is to provide users with a real-time movement sonification independently of a reference to a “normal movement.” In this case, implicit learning is in play. The comparison between these two approaches remains to be carefully investigated, both in term of learning speed and retention. Our results (Boyer et al., 2014; Boyer, 2015) show that these two approaches are in fact complementary and the

combination of both can be beneficial. Nevertheless, more studies are necessary to clarify the different neural mechanisms that are implied:

- *Are the neural mechanisms for error-related auditory feedback different to those when movement sonification is used as a feedback on users' own movements?* In particular, can the movement sonification be considered as “enhancing proprioception” in integrating congruent information?
- *What is the role of the guidance hypothesis in these cases?* Can the constant use of the auditory feedback be detrimental for learning (i.e., the effect disappears when the feedback is removed). Such a point needs to be clarified for both approaches and has still insufficiently been studied.
- *Is the standard distinction between KR (knowledge of results) or KP (knowledge of performance) relevant when considering movement sonification?* Since both feedback types can occur simultaneously during real-time movement sonification (Boyer et al., 2014), specific formalization of auditory feedback types should be developed.

Second, the role of the sound characteristics remains elusive for quantifying the learning efficiency or learning rate. Reported results have been sometimes contradictory, and very different mapping or sound types have been equally successful. The role of the mapping or sound quality must be further studied, and we particularly propose to focus on two important questions:

- *How does a particular mapping can favor agency in the interaction?* The auditory feedback should be clearly predictable by the user. This requires to study both the sound and movement perception in an interactive context (Effenberg, 2005).
- *When is the auditory feedback perceived as motivational?* Fundamental studies of sensori-motor learning with auditory feedback has generally avoided to take into account possible emotional response to the auditory feedback. Nevertheless, even the simplest ones using pure tone might trigger positive or negative effects, which can potentially affect the task perception, adaptation and learning (Lemaitre et al., 2012).

Third, our studies as well as many other published results point toward a large variability between participants. Such findings might be put in parallel with the large variability found in rhythmic ability, which motivated the establishment of a standard test called BAASTA (Farrugia et al., 2012). We believe that such a test would be highly useful for movement sonification in interactive systems. This would represent a first step toward more reproducible results and build understanding of possible causes of this variability. This point is also crucial to develop real-world applications. Moreover, sound design applications provide extremely fruitful cases to study sound perception as an active and contextual process. In that new framework, sound perception studies should be redesigned in relation to gesture and to user's objectives.

As already mentioned, movement sonification or more generally the use of auditory feedback have been already proposed for specific applications. Beyond artistic communities

which have already largely included movement-based interactive systems, the most prominent ones are rehabilitation (Robertson et al., 2009; Roby-Brami et al., 2014; Scholz et al., 2014; Katan et al., 2015), sport learning and training (Effenberg, 2004; Eriksson and Bresin, 2010; Boyer et al., 2013b) while human-computer interaction also shows a growing interests (Franinović and Serafin, 2013; Oh et al., 2013).

In stroke patients, the sound based therapies are specifically promising to target the impairment of the upper-limb. The contemporary guidelines for rehabilitation insist on the similarity between sensori-motor learning and recovery phenomenon. Thus, therapy should be improved both in quantity and quality: on the one hand it should be based on massive exercise repetitions, emphasizing on sensory-motor reciprocity and multisensory integration. On the other hand, the therapy should be adapted to the needs of each individual: the exercises should be shaped according to the precise capabilities of the person and should evolve according to his/her abilities and progress during learning. Sound feedback is frequently integrated into virtual and augmented reality rehabilitation training but its use is often limited to rhythmic auditory cues or reinforcement feedback signaling only the success to an exercise. We propose that sonification could be further developed to target specific impairments in stroke patients as a continuous feedback during movement execution. Sonification is particularly interesting to signal to the patients some impairment, which they might not be aware of, particularly if they have somatosensory impairments, for example error in direction, in coordination or lack of movement smoothness (Maulucci and Eckhouse, 2001), in coordination of reaching and grasping for prehension or in grasping to lift coordination (Hsu et al., 2012). Thanks to a braked elbow orthosis, we simulated the disrupted shoulder-elbow coordination observed in hemiparetic stroke patients and used this device to test sonification strategies that we developed to target shoulder elbow coordination. Further studies are needed in order to find a compromise between two possibly contradictory requirements: target the specific impairments of stroke patients and develop motivation linked to exploration of sophisticated auditory-motor coupling.

Beside the fundamental aspects we described about the understanding of the different auditory feedback mechanisms that can contribute to sensori-motor learning, the development of rigorous—and shared—sound design methodologies is crucial for grounding these applications. As a matter of fact, the use of sound in any technological applications could lead to user annoyance or discomfort, even if globally beneficial for movement training. We therefore advocate for more interdisciplinary research bringing together sound designer, musicians, engineers, cognitive scientists, to work toward efficient applications using movement-based sonification. On the one hand, the collaboration with sound artists and musicians is generally necessary to design pleasant and motivational interactive sound and music systems, on the other hand sound design research should further develop methods to assess naturalness

and pleasantness of sonic interactive system (Susini et al., 2012).

AUTHOR CONTRIBUTIONS

FB drafted the paper, and all the other authors (EB, JF, OH, PS, AR, SH) revised the article critically for important intellectual content. All authors contributed to the research project described in this Perspective article (i.e., Legos project) and participated in formalizing the research questions and perspectives. All authors gave their final approval for publication of this paper.

REFERENCES

- Bevilacqua, F., Schnell, N., Rasamimanana, N., Bloit, J., Fléty, E., Caramiaux, B., et al. (2013). "De-MO: designing action-sound relationships with the MO interfaces," in *CHI '13 Extended Abstracts on Human Factors in Computing Systems* (Paris). doi: 10.1145/2468356.2479571
- Bevilacqua, F., Schnell, N., Rasamimanana, N., Zamborlin, B., and Guédy, F. (2011). "Online gesture analysis and control of audio processing," in *Musical Robots and Interactive Multimodal Systems*, Vol. 74, eds J. Solis and K. Ng (Springer Berlin Heidelberg), 127–142. doi: 10.1007/978-3-642-22291-7_8
- Boyer, E. (2015). *Continuous Auditory Feedback for Sensorimotor Learning*. Ph.D. thesis, Université Pierre et Marie Curie - Paris VI.
- Boyer, E., Vandervoorde, L., Bevilacqua, F., and Hanneton, S. (2015). "Touching sounds: perception of the curvature of auditory virtual surfaces," in *Virtual Reality (VR), 2015 IEEE* (Arles), 153–154. doi: 10.1109/VR.2015.7223341
- Boyer, E. O., Babayan, B. M., Bevilacqua, F., Noisternig, M., Warusfel, O., Roby-Brami, A., et al. (2013a). From ear to hand: the role of the auditory-motor loop in pointing to an auditory source. *Front. Comput. Neurosci.* 7:26. doi: 10.3389/fncom.2013.00026
- Boyer, E. O., Bevilacqua, F., Phal, F., and Hanneton, S. (2013b). Low-cost motion sensing of table tennis players for real time feedback. *Int. J. Table Tennis Sci.* 8, 1–4.
- Boyer, E. O., Pyanet, Q., Hanneton, S., and Bevilacqua, F. (2014). "Learning movement kinematics with a targeted sound," in *Sound, Music & Motion, Lecture Notes in Computer Science*, Vol. 8905, eds M. Aramaki, O. Derrien, R. Kronland-Martinet, and S. Ystad (Springer International Publishing), 218–233. doi: 10.1007/978-3-319-12976-1_14. Available online at: http://link.springer.com/chapter/10.1007%2F978-3-319-12976-1_14
- Caramiaux, B., Bevilacqua, F., Bianco, T., Schnell, N., Houix, O., and Susini, P. (2014a). The role of sound source perception in gestural sound description. *ACM Trans. Appl. Percept.* 11, 1–19. doi: 10.1145/2536811
- Caramiaux, B., Schnell, N., Françoise, J., and Bevilacqua, F. (2014b). Mapping through listening. *Comput. Music J.* 38, 34–48. doi: 10.1162/COMJ_a_00255
- Castiello, U., Giordano, B. L., Begliomini, C., Ansuini, C., and Grassi, M. (2010). When ears drive hands: the influence of contact sound on reaching to grasp. *PLoS ONE* 5:e12240. doi: 10.1371/journal.pone.0012240
- Cirstea, C. M., Pfitto, A., and Levin, M. F. (2006). Feedback and cognition in arm motor skill reacquisition after stroke. *Stroke* 37, 1237–1242. doi: 10.1161/01.STR.0000217417.89347.63
- Dubus, G., and Bresin, R. (2013). A systematic review of mapping strategies for the sonification of physical quantities. *PLoS ONE* 8:e82491. doi: 10.1371/journal.pone.0082491
- Effenberg, A. (2004). "Using sonification to enhance perception and reproduction accuracy of human movement patterns," in *International Workshop on Interactive Sonification 2004* (Bielefeld), 1–5.
- Effenberg, A. O. (2005). Movement sonification: effects on perception and action. *IEEE Multimedia* 12, 53–59. doi: 10.1109/MMUL.2005.31
- Eriksson, M., and Bresin, R. (2010). "Improving running mechanics by use of interactive sonification," in *Proceedings of the Interaction Sonification Workshop (ISON) 2010* (Stockholm: KTH Royal Institute of Technology).

ACKNOWLEDGMENTS

This work was funded by the French National Research Agency (LEGOS project -ANR-11-BS02-012) and the Labex SMART (supported by French state funds managed by the ANR within the "Investissements d'Avenir" program under reference ANR-11-IDEX-0004-02). We thank all the participants of the Legos project for their important contributions, as well as our external collaborators Ana Tajadura-Jimenez, Nadia Berthouze, and Nathanaël Jarrassé, whose works are reported in this paper.

- Farrugia, N., Benoit, C.-E., Harding, E., Kotz, S. A., and Dalla Bella, S. (2012). "BAASTA : battery for the assessment of auditory sensorimotor and timing abilities," in *Proceedings of the 12th International Conference on Music Perception and Cognition and the 8th Triennial Conference of the European Society for the Cognitive Sciences of Music* (Thessaloniki).
- Fiebrink, R., and Cook, P. R. (2010). "The wekinator: a system for real-time, interactive machine learning in music," in *Proceedings of The Eleventh International Society for Music Information Retrieval Conference (ISMIR 2010)* (Utrecht).
- Françoise, J. (2015). *Motion-Sound Mapping By Demonstration*. Ph. D. thesis, Université Pierre et Marie Curie - Paris VI.
- Françoise, J., Chapuis, O., Hanneton, S., and Bevilacqua, F. (2016). "Soundguides: adapting continuous auditory feedback to users," in *Proceedings of the 34th International Conference on Human Factors in Computing Systems, Extended Abstracts, CHI EA '16* (San Jose, CA: ACM).
- Franinović, K., and Serafin, S. (2013). *Sonic Interaction Design*. Cambridge, MA; London, UK: MIT Press.
- 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
- Hanneton, S., Herquel, P., and Auvray, M. (2015). Intermodal recoding of a video game: learning to process signals for motion perception in a pure auditory environment. *Int. J. Adapt. Control Signal Process.* 29, 1475–1483. doi: 10.1002/acs.2549
- Houix, O., Bevilacqua, F., Misdariis, N., Susini, P., Fléty, E., Françoise, J., et al. (2015). "Objects with multiple sonic affordances to explore gestural interactions," in *xCoAx 2015. Computation, Communication, Aesthetics & X*, xCoAx 2015: Proceedings of the Third Conference on Computation, Communication, Aesthetics and X (Glasgow: Alison Clifford, Miguel Carvalhais and Mario Verdicchio), 296–303.
- Houix, O., Misdariis, N., Susini, P., Bevilacqua, F., and Gutierrez, F. (2014). "Sonically augmented artifacts: design methodology through participatory workshops," in *Lecture Notes in Computer Science*, Vol. 8905, eds M. Aramaki, O. Derrien, R. Kronland-Martinet, and S. Ystad (Springer International Publishing), 20–40. doi: 10.1007/978-3-319-12976-1_2. Available online at: http://link.springer.com/chapter/10.1007%2F978-3-319-12976-1_2
- Hsu, H.-Y., Lin, C.-F., Su, F.-C., Kuo, H.-T., Chiu, H.-Y., and Kuo, L.-C. (2012). Clinical application of computerized evaluation and re-education biofeedback prototype for sensorimotor control of the hand in stroke patients. *J. Neuroeng. Rehabil.* 9, 1–9. doi: 10.1186/1743-0003-9-26
- Hunt, A., Wanderley, M. M., and Paradis, M. (2003). The importance of parameter mapping in electronic instrument design. *J. New Music Res.* 32, 429–440. doi: 10.1076/jnmr.32.4.429.18853
- Katan, S., Grierson, M., and Fiebrink, R. (2015). "Using interactive machine learning to support interface development through workshops with disabled people," in *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, CHI '15* (New York, NY: ACM), 251–254.
- Kramer, G., Walker, B. N., Bonebright, T., Cook, P. R., Flowers, J., and Miner, N. (1999). *Sonification Report: Status of the Field and Research Agenda*. Technical Report, University of Nebraska - Lincoln.

- Lemaitre, G., Houix, O., Susini, P., Visell, Y., and Franinovic, K. (2012). Feelings elicited by auditory feedback from a computationally augmented artifact: the flops. *IEEE Trans. Affect. Comput.* 3, 335–348. doi: 10.1109/T-AFFC.2012.1
- Leman, M. (2008). *Embodied Music Cognition and Mediation Technology*. Cambridge, MA; London, UK: MIT Press.
- Maes, P.-J., Leman, M., Palmer, C., and Wanderley, M. M. (2014). Action-based effects on music perception. *Front. Psychol.* 4:1008. doi: 10.3389/fpsyg.2013.01008
- Maulucci, R. A., and Eckhouse, R. H. (2001). Retraining reaching in chronic stroke with real-time auditory feedback. *NeuroRehabilitation* 16, 171–182.
- Oh, U., Kane, S. K., and Findlater, L. (2013). “Follow that sound: using sonification and corrective verbal feedback to teach touchscreen gestures,” in *Proceedings of the 15th International ACM SIGACCESS Conference on Computers and Accessibility, ASSETS '13* (New York, NY: ACM), 13:1–13:8. doi: 10.1145/2513383.2513455
- Rath, M., and Rocchesso, D. (2005). Continuous sonic feedback from a rolling ball. *IEEE Multimedia* 12, 60–69. doi: 10.1109/MMUL.2005.24
- Rath, M., and Schleicher, R. (2008). On the relevance of auditory feedback for quality of control in a balancing task. *Acta Acust. United Acust.* 94, 12–20. doi: 10.3813/AAA.918003
- Ripollés, P., Rojo, N., Grau-Sánchez, J., Amengual, J., Càmarà, E., Marco-Pallarés, J., et al. (2015). Music supported therapy promotes motor plasticity in individuals with chronic stroke. *Brain Imaging Behav.* doi: 10.1007/s11682-015-9498-x. Available online at: <http://link.springer.com/article/10.1007/s11682-015-9498-x>. [Epub ahead of print].
- Robertson, J. V. G., Hoellinger, T., Lindberg, P., Bensmail, D., Hanneton, S., and Roby-Brami, A. (2009). Effect of auditory feedback differs according to side of hemiparesis: a comparative pilot study. *J. Neuroeng. Rehabil.* 6:45. doi: 10.1186/1743-0003-6-45
- Roby-Brami, A., Van Zandt-Escobar, A., Jarrassé, N., Robertson, J., Schnell, N., Boyer, E., et al. (2014). Toward the use of augmented auditory feedback for the rehabilitation of arm movements in stroke patients. *Ann. Phys. Rehabil. Med.* 57, e4–e5. doi: 10.1016/j.rehab.2014.03.015
- Rocchesso, D., Polotti, P., and Delle Monache, S. (2009). Designing continuous sonic interaction. *Int. J. Design* 3, 13–25.
- Rosati, G., Oscari, F., Spagnol, S., Avanzini, F., and Masiero, S. (2012). Effect of task-related continuous auditory feedback during learning of tracking motion exercises. *J. Neuroeng. Rehabil.* 9:79. doi: 10.1186/1743-0003-9-79
- Schiavio, A., and Altenmüller, E. (2015). Exploring music-based rehabilitation for parkinsonism through embodied cognitive science. *Front. Neurol.* 6:217. doi: 10.3389/fneur.2015.00217
- Schlaug, G. (2015). “Musicians and music making as a model for the study of brain plasticity,” in *Music, Neurology, and Neuroscience: Evolution, the Musical Brain, Medical Conditions, and Therapies*, Vol. 217 of Progress in Brain Research, Chapter 3, eds S. F. Eckart Altenmüller and F. Boller (Amsterdam: Elsevier), 37–55. Available online at: <http://store.elsevier.com/Music-Neurology-and-Neuroscience-Evolution-the-Musical-Brain-Medical-Conditions-and-Therapies/isbn-9780444635518/>
- Schmidt, R. A. (1988). *Motor Control and Learning: A Behavioral Emphasis, 2nd Edn*. Champaign, IL: Human Kinetics.
- Schneider, S., Schönle, P. W., Altenmüller, E., and Münte, T. F. (2007). Using musical instruments to improve motor skill recovery following a stroke. *J. Neurol.* 254, 1339–1346. doi: 10.1007/s00415-006-0523-2
- Schnell, N., Röbel, A., Schwarz, D., Peeters, G., and Borghesi, R. (2009). “MuBu and friends – assembling tools for content based real-time interactive audio processing in Max/MSP,” in *Proceedings of the International Computer Music Conference (ICMC 2009)* (Montreal, QC), 423–426.
- Scholz, D. S., Wu, L., Schneider, J., Pirzer, J., Grossbach, M., Rollnik, J. D., et al. (2014). Sonification as a possible stroke rehabilitation strategy. *Front. Neurosci.* 8:332. doi: 10.3389/fnins.2014.00332
- Sigrist, R., Rauter, G., Riener, R., and Wolf, P. (2013). Augmented visual, auditory, haptic, and multimodal feedback in motor learning: a review. *Psychon. Bull. Rev.* 20, 21–53. doi: 10.3758/s13423-012-0333-8
- Singh, A., Piana, S., Pollarolo, D., Volpe, G., Varni, G., Tajadura-Jimenez, A., et al. (2016). *Go-with-the-flow*: tracking, analysis and sonification of movement and breathing to build confidence in activity despite chronic pain. *Hum. Comput. Interact.* 31, 1–49. doi: 10.1080/07370024.2015.1085310
- Susini, P., Misdariis, N., Lemaitre, G., and Houix, O. (2012). Naturalness influences the perceived usability and pleasantness of an interface's sonic feedback. *J. Multimodal User Interfaces* 5, 175–186. doi: 10.1007/s12193-011-0086-0
- Tajadura-Jimenez, A., Bianchi-Berthouze, N., Furfaro, E., and Bevilacqua, F. (2015). Sonification of surface tapping changes behavior, surface perception, and emotion. *IEEE MultiMedia* 22, 48–57. doi: 10.1109/MMUL.2015.14
- Tajadura-Jiménez, A., Liu, B., Bianchi-Berthouze, N., and Bevilacqua, F. (2014). “Using sound in multi-touch interfaces to change materiality and touch behavior,” in *Proceedings of the Nordic Conference on Human-Computer Interaction (NordiCHI'14)* (Helsinki), 199–202.
- Thaut, M. H. (2015). The discovery of human auditory-motor entrainment and its role in the development of neurologic music therapy. *Prog. Brain Res.* 217, 253–266. doi: 10.1016/bs.pbr.2014.11.030
- Wanderley, M. M., and Depalle, P. (2004). Gestural control of sound synthesis. *Proc. IEEE* 92, 632–644. doi: 10.1109/JPROC.2004.825882
- Zampini, M., and Spence, C. (2004). The role of auditory cues in modulating the perceived crispness and staleness of potato chips. *J. Sens. Stud.* 19, 347–363. doi: 10.1111/j.1745-459x.2004.080403.x

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.

Copyright © 2016 Bevilacqua, Boyer, Françoise, Houix, Susini, Roby-Brami and Hanneton. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Biomusic: An Auditory Interface for Detecting Physiological Indicators of Anxiety in Children

Stephanie Cheung^{1,2*†}, Elizabeth Han^{1,2†}, Azadeh Kushki^{1,2}, Evdokia Anagnostou^{2,3} and Elaine Biddiss^{1,2}

¹ Institute of Biomaterials and Biomedical Engineering, University of Toronto, Toronto, ON, Canada, ² Bloorview Research Institute, Holland Bloorview Kids Rehabilitation Hospital, Toronto, ON, Canada, ³ Department of Paediatrics, University of Toronto, Toronto, ON, Canada

OPEN ACCESS

Edited by:

Diego Minicacchi,
University of Florence, Italy

Reviewed by:

Jose Luis Contreras-Vidal,
University of Houston, USA
Nadia Bianchi-Berthouze,
University College London, UK

*Correspondence:

Stephanie Cheung
scheung@hollandbloorview.ca

[†]These authors have contributed
equally to this work.

Specialty section:

This article was submitted to
Neuroprosthetics,
a section of the journal
Frontiers in Neuroscience

Received: 30 April 2016

Accepted: 15 August 2016

Published: 30 August 2016

Citation:

Cheung S, Han E, Kushki A,
Anagnostou E and Biddiss E (2016)
Biomusic: An Auditory Interface for
Detecting Physiological Indicators of
Anxiety in Children.
Front. Neurosci. 10:401.
doi: 10.3389/fnins.2016.00401

For children with profound disabilities affecting communication, it can be extremely challenging to identify salient emotions such as anxiety. If left unmanaged, anxiety can lead to hypertension, cardiovascular disease, and other psychological diagnoses. Physiological signals of the autonomic nervous system are indicative of anxiety, but can be difficult to interpret for non-specialist caregivers. This paper evaluates an auditory interface for intuitive detection of anxiety from physiological signals. The interface, called “Biomusic,” maps physiological signals to music (i.e., electrodermal activity to melody; skin temperature to musical key; heart rate to drum beat; respiration to a “whooshing” embellishment resembling the sound of an exhalation). The Biomusic interface was tested in two experiments. Biomusic samples were generated from physiological recordings of typically developing children ($n = 10$) and children with autism spectrum disorders ($n = 5$) during relaxing and anxiety-provoking conditions. Adult participants ($n = 16$) were then asked to identify “anxious” or “relaxed” states by listening to the samples. In a classification task with 30 Biomusic samples (1 relaxed state, 1 anxious state per child), classification accuracy, sensitivity, and specificity were 80.8% [standard error (SE) = 2.3], 84.9% (SE = 3.0), and 76.8% (SE = 3.9), respectively. Participants were able to form an early and accurate impression of the anxiety state within 12.1 (SE = 0.7) seconds of hearing the Biomusic with very little training (i.e., <10 min) and no contextual information. Biomusic holds promise for monitoring, communication, and biofeedback systems for anxiety management.

Keywords: sonification, disability, anxiety, augmentative and alternative communication (AAC), music

1. INTRODUCTION

Medical advancements have led to a growing number of people surviving previously fatal medical complications, and subsequently living with profound disabilities. For these individuals, survival depends on life-supporting technologies and teams of caregivers who can anticipate and respond to their complex continuing care needs. Even when cognitive function remains intact, many of these individuals are unable to communicate via traditional pathways (e.g., speech, gestures) and/or augmented and assisted communication (AAC) devices (Hogg et al., 2001) because of physical limitations. As such, it often falls to caregivers to decipher their preferences, intent, and

emotions from sparse behavioral and contextual cues (Adams and Oliver, 2011; Blain-Moraes et al., 2013) that may be difficult to detect, non-obvious to those unfamiliar with the individual, and inconsistent/unavailable across individuals (Adams and Oliver, 2011). In the absence of a reliable communication pathway, the needs, thoughts, and feelings of those with profound disabilities are at risk of being overlooked, which presents concerning challenges to health and well-being (Blain-Moraes et al., 2013). This motivates the urgent need to establish communication channels with individuals with profound disabilities and methods for discerning their emotional states.

Emotion-modulated physiological signals may provide additional cues when interpreting the affective states of those with profound disabilities. Vos et al. (2012) have demonstrated that changes in skin temperature and heart rate may be associated with positive and negative valence emotions in persons with profound disabilities. While several groups have developed algorithmic classifiers that use machine learning to identify and display affect from physiological signals of healthy adults (Wen et al., 2014), the application of this research to individuals with profound disabilities presents an additional difficulty. Methodologically, it is extremely challenging to develop classifiers with a population who are unable to verify their performance or communicate the “ground truth.” Ethically, we must be conscious of the potential challenges of assigning affective state labels to individuals who can neither confirm nor correct their accuracy.

An alternative is to consider physiological signals as an additional source of information that can be continuously streamed, rather than discretely classified. Caregivers may learn to interpret this information and integrate it alongside other contextual and behavioral cues. In contrast to a computerized classifier, this approach relies on human intelligence and aptitude for pattern recognition. It does not presume to assign labels, but rather hinges on the ability to effectively present multiple complex, streaming physiological signals to caregivers in a way that is easy and intuitive to understand. Reliance on human interpretation may introduce variability, but it may also offer more flexibility to accommodate a wider range of person- and condition-specific heterogeneity, to base decisions on multiple information sources (e.g., physiology, behavior, context), and to avoid ethical challenges associated with assigning definitive emotion labels.

Even for experts, it may be extremely challenging to make sense of complex raw physiological data (Sanderson, 2006). Intuitive auditory displays could make the valuable information contained in these signals more accessible. To this purpose, we created Biomusic, an auditory interface that converts physiological signals to sound. The system outputs sounds in MIDI format, which can be listened to in real-time or offline. Specifically, the music’s melody is driven by changes in electrodermal activity; the musical key, established by tonic chords, is linked to changes in skin temperature; a rhythmic, percussive beat is associated with heart rate (extracted from blood volume pulse); and respiration is projected via a “whooshing” embellishment resembling the sound of an exhalation (**Table 1**).

Mappings were based upon guidelines from Watson and Sanderson, who suggested that periodic signals such as heart rate may be naturally suited to tempo, while aperiodic signals may suit tonal representation (Watson and Sanderson, 2001). Thus, as per criteria outlined by Hermann (2008), Biomusic is objective (i.e., reflective of the nature of the inputs) and systematic in its mapping, is able to reproduce its output given the same input, and can be applied to physiological signals of multiple different users.

While visually-streamed physiological data demand continuous attention to interpret, Biomusic can provide caregivers with peripheral awareness of multiple physiological signals and trends while preoccupied with other aspects of care (Blain-Moraes et al., 2013). In a systematic review, Dubus and Bresin (2013) identified and categorized 7 sonification projects as having the primary purpose of “monitoring.” This suggests the utility of an auditory medium such as Biomusic to convey continuous, real-time information to an attentive listener. Previous studies have demonstrated the effectiveness of using sound to represent physiological signals. Yokoyama et al. (2002) mapped instantaneous heart rate to musical pitch and intervals for use in biofeedback for stress and blood pressure management. Wu et al. (2009) enabled participants to distinguish between rapid-eye movement and slow-wave sleep when presented with sonified electroencephalographs. As such, there is evidence to suggest that complex physiological signals can be interpreted by listening to their musical translations. Unlike these interfaces, Biomusic enables information from multiple physiological signals to be projected simultaneously through different but concurrent mappings, forming a holistic soundscape. To our knowledge, Biomusic is unique in this regard.

Biomusic has been successfully demonstrated in a clinical care context to augment communication between caregivers and children with profound disabilities by increasing a sense of reciprocity and co-presence, two fundamental qualities of human interaction (Blain-Moraes et al., 2013). Anecdotally, caregivers in this study noticed changes in the sound and quality of Biomusic that appeared to be associated with their interactions with the children. Caregivers in that study expressed a strong desire to decode or interpret Biomusic, which motivated the research question reported on in this paper: How effectively can affective states, specifically anxiety, be identified using Biomusic? The present study aimed to formally evaluate how well different physiological patterns associated with affective state can be detected via Biomusic. Specifically, we determined the accuracy and speed with which listeners with minimal training could distinguish anxious from relaxed physiological responses based on Biomusic alone. Anxiety was selected as the target emotion given its clinical importance and its suspected prevalence in individuals with profound disabilities. In typical populations, anxiety modulates multiple physiological signals through excitation of the sympathetic nervous system and inhibition of the parasympathetic nervous system (Kreibitz, 2010). This results in:

- Increased electrodermal activity due to increased perspiration (Vetrugno et al., 2003).

TABLE 1 | Physiological Signal-to-Music Mapping.

Physiological signal	Feature extracted	Musical representation
Electrodermal activity	Average skin conductance over a 0.25-s window in μS	Melodic pitch (“goblins” MIDI-sound): Three octaves of a major-key scale are mapped to 1 μS of electrodermal activity. The melody establishes the first windowed and averaged electrodermal activity value as the tonic center in a major scale. Subsequent EDA values are compared to the first, and differences result in a pitch shift relative to the magnitude of change. Extreme levels of electrodermal activity are transposed to keep the melody in a comfortable range. Thus, quickly-changing EDA results in active melodic runs. Melody may be suited to reflect the salience of continuous changes in electrodermal conductivity with anxiety state.
Skin temperature	Instantaneous skin temperature in degrees C	Key change (“choir” MIDI-sound chords): Major tonic chords are mapped to instantaneous skin temperature. This feature is updated every two bars. The first skin temperature value is mapped to the tonic chord of a C-major scale, and subsequent increases or decreases in temperature raise or lower the key of the music, and therefore the chord, by one semitone. Chords indicative of key change may be suited to represent gradually-continuous changes in skin temperature.
Blood volume pulse	Average interbeat interval over four periods of the signal	Tempo (“melodic tom” MIDI-sound): The pace of the music, marked by a percussive beat, is updated every musical bar. This tempo is determined from the average interbeat interval of the blood volume pulse signal over four periods. Therefore, increased heart rate increases the tempo of the music and its beat. A drum beat is suited to represent the rhythmic, cyclic pulse of the circulatory system.
Respiration	Duration of exhalation (peak-to-trough intervals of the signal)	“Whoosh”: Expiratory time is mapped to the duration of a whooshing, seashore embellishment resembling the sound of an exhalation. The “Whoosh embellishment is sustained as long as exhalation is detected. During inhalation, this sound is not played.

- Decreased fingertip skin temperature due to vasoconstriction (Rimm-Kaufman and Kagan, 1996).
- Increased heart rate due to increased cardiac contraction (Franchini and Cowley, 2004).
- Increased respiration rate resulting from faster, shallower breathing (Kreibig, 2010).

Through two experiments, we aimed to determine the feasibility of the Biomusic interface for conveying anxiety-related changes in physiological responses. Two experiments were conducted and in each, the following approach was used. First, physiological recordings were collected from children during relaxed and anxiety-provoking conditions. Next, anxious and relaxed state Biomusic samples were generated from these physiological recordings. Last, adult participants were asked to listen to the Biomusic samples and to identify the anxiety state associated with each, based on the Biomusic alone and no additional contextual cues. Anxious and relaxed state Biomusic samples are included in Supplementary Materials.

In the first experiment, we generated Biomusic samples from children who were considered to be typically-developing and represented the best case use scenario, having typical physiological signals and the ability to report emotions to corroborate their anxiety state. The second experiment followed a protocol identical to the first, but with the addition of Biomusic samples generated from children with autism spectrum disorders (ASD). Children with ASD have been reported to have atypical anxiety-modulated physiological responses and difficulty recognizing and expressing emotional state (Kushki et al., 2013). The ASD group represents a stepping-stone population as we work toward the eventual use case: a child with profound disabilities, with possibly typical or atypical signals, and no ability to report emotion.

2. EXPERIMENT 1: DETECTION OF ANXIETY STATE THROUGH BIOMUSIC IN TYPICALLY DEVELOPING CHILDREN

In the first experiment, we tested the feasibility of Biomusic technology to convey anxiety by having adult raters classify samples of Biomusic. We chose to use the Biomusic of typically-developing children, who represent the best case use scenario, having typical physiological signals and the ability to report emotions to corroborate their anxiety state. In this experiment, we aimed to (1) quantify the performance (sensitivity, specificity, accuracy) of physiological sonification as an anxiety screening tool, (2) determine the latency in seconds for detecting an anxious or relaxed state, (3) explore the confidence with which detection are made, and (4) identify the most influential musical elements in the decision. All protocols were reviewed and approved by the science and ethical review board at Holland Bloorview Kids Rehabilitation Hospital. Informed consent and assent were obtained from adults and children respectively.

2.1. Methods

2.1.1. Biomusic

Biomusic samples were created from physiological signals of typically-developing children (aged 8–13) acquired during anxiety-inciting and relaxing conditions. Signals were acquired in individual sessions. Physiological recordings were obtained using the ProComp Infiniti data acquisition system and Biocomp Infiniti software (Thought Technology, Montreal, Canada). Four physiological signals were collected using non-invasive sensors on the fingers of the non-dominant hand (electrodermal activity, skin temperature, blood volume pulse) and chest (respiration). Electrodermal activity was recorded using dry Ag/Ag-Cl sensors,

which were placed on intermediate phalanges on the second and third fingers of the non-dominant hand. Skin temperature was measured by a surface thermistor, fixed to the distal phalanx of the fifth finger, and blood volume pulse was monitored by an infrared photoplethysmography sensor, attached to the distal phalanx of the fourth finger. Respiration was recorded using a piezoelectric belt positioned around the thoracic cage. Children were instructed to limit unnecessary movements to avoid interfering with the sensors. All physiological signals were sampled at 256 Hz and passed through a 5th order Butterworth anti-aliasing filter built in to the Biocomp Infinity.

Data acquisition began with a 20-min relaxed state condition while children watched a low-intensity nature video. Jennings et al. (1992) have suggested that tasks that require some attention may be more representative of physiological “baseline” states than those that do not. Following this period, children self-reported their anxiety state using the gold standard, the child version of Spielberger’s State Trait Anxiety Inventory (STAI) (Spielberger et al., 1970). The STAI is composed of two sub-scales to measure state (i.e., situational) and trait anxiety. In this study, we used the state anxiety scale. The lowest and highest possible state scores on the raw scale are 20 points and 60 points, respectively. These scores are then normalized according to **Table 2** in the STAI manual (Spielberger et al., 1970). Anxiety state recordings were then collected in three 2-min trials which required children to solve anagrams of varying difficulty levels. Test-taking situations are a common method for provoking anxiety in children and have been shown to increase state anxiety on the STAI (Spielberger et al., 1970). Anagram tests have previously been used with children 8–12 years of age (Kaslow et al., 1983; Ward et al., 1987; Affrunti and Woodruff-Borden, 2014). Each anxious-state trial was followed immediately by a 5-min break during which the child engaged in the relaxed-state condition (watching a nature video). Following the anxious-state trials, children reported state anxiety on the STAI, focusing on how they felt during the anagram tasks. This protocol is outlined in **Figure 1**. **Table 4** presents the means of the physiological signal features of interest during the relaxed and anxiety-provoking conditions.

The physiological data were considered valid only if children’s self-reported anxiety state matched the experimental condition

(e.g., the child reported higher state anxiety while doing the anagram task). This required an increase in state score between relaxed and anxious conditions (Spielberger et al., 1970). Not all child participants may have experienced anxiety; several reported that they liked the anagram task and found it enjoyable. Of 12 child participants, 7 children [mean = 10 years, standard deviation (SD) = 2 years] met the validity criteria outlined. Of these participants, two had fidgeted excessively during data collection. Data from these two participants were not used. In the remaining five participants, the average difference in normalized state anxiety scores between relaxed and anxious states was 6.2 points (SE = 1.7) (Range: 31–52 for relaxed and 42–57 for anxious). By referring to children’s self report to verify that the experimental condition had been met, we established, as well as possible, the emotional state of the children.

Once the emotional state of the physiological data was verified through referral to self-report, one anxious state and one relaxed state data segment (80-s in length) were randomly extracted from each child’s physiological recordings. To avoid selecting transitional states (i.e., relaxed-to-anxious or anxious-to-relaxed), segments at the beginning and end of experimental conditions were not used. This resulted in 10 segments of physiological data (5 anxious-state, 5 relaxed-state) that were translated into clips of Biomusic as per **Table 1** and recorded to be used offline. Of note, physiological signals were not referred to in the generation of the Biomusic excerpts so as not to bias the selection of songs.

2.1.2. Participants

Sixteen university student participants (11 female, 5 male, 21.1 ± 2.6 years) were recruited through community postings to classify the 10 segments of Biomusic as either anxious or relaxed state. All participants had normal or corrected to normal hearing. This sample was highly representative of the volunteer population at Holland Bloorview Kids Rehabilitation Hospital—a number of whom interact routinely with children with profound communication disabilities in therapeutic recreation programs, school programs, and in their day-to-day care.

2.1.3. Evaluation Protocol

The feasibility of the Biomusic interface was evaluated for sensitivity, specificity, accuracy, and latency as follows. The classification session began with a brief 10-min training period during which one example of relaxed-state Biomusic and one example of anxious-state Biomusic were played. The experimenter explained the sounds that might be expected based on current understanding of the psychophysiology of anxiety (e.g., accelerated drum beat, shorter and faster whooshes, larger and more frequent runs, and changes in the melody). These 80-s training samples were generated using sections of a typically-developing child’s signal recordings, which exhibited the expected anxiety-modulated physiological trends. The training samples were not included in the test set.

Following the training period, participants were asked to listen to all 10 test samples of Biomusic, presented in random order, and to classify each one as either “anxious” or “relaxed” state. Samples were randomized with respect to the child and to the anxiety

TABLE 2 | Means and standard error of physiological features recorded in Experiment 1 from typically developing children ($n = 5$) during the relaxed state condition and the anxiety-provoking condition.

Typically-developing (anagram)		
Signal feature	Relaxed mean	Anxious mean
EDA (μ S) *	1.131, SE = 0.390	3.550, SE = 0.706
Skin temperature ($^{\circ}$ C)	32.086, SE = 1.826	30.412, SE = 2.312
Blood volume inter-pulse interval (s)	0.797, SE = 0.049	0.761, SE = 0.025
Respiration inter-breath interval (s) *	3.333, SE = 0.296	2.269, SE = 0.107

*Indicates features that were significantly different ($p < 0.05$) between relaxed and anxious state conditions.

condition. Participants had access only to the Biomusic and had no information regarding the child or the context during which the physiological signals were recorded. Participants listened to the songs in a quiet room with only the experimenter present. The songs were presented via a conventional set of computer speakers. Participants were instructed to classify the anxiety state of each Biomusic sample as soon as they felt confident in their decision. This first classification is referred to throughout this paper as their “initial impression.” The selection (i.e., anxious or relaxed state) was made on a computer screen and the time of response was recorded. After the Biomusic sample was heard in its entirety, participants indicated their final classification decision. Participants then rated their confidence in their final decision on a 5-point Likert scale ranging from 1 (low) to 5 (high). Participants also reported the most influential musical component (i.e., melody, chords, drum beat, or “whoosh”) on their classification. This protocol is described in **Figure 2**.

2.1.4. Data Analysis

The primary analysis was to determine the sensitivity, specificity, and overall classification accuracy of the Biomusic interface for classifying relaxed and anxious states from physiological signals of typically-developing children. Additionally, the latency (i.e., the average length of time required for the participants to form an initial impression of the Biomusic sample) was recorded. The median self-reported confidence was calculated to evaluate listeners' certainty during the classification task. Last, frequency counts were compiled to determine the most influential musical component in the interpretation of the Biomusic.

2.2. Results and Discussion

On average, participants made initial classifications of the Biomusic with high accuracy [83.9%, standard error (SE) = 2.9%], sensitivity (87.1%, SE = 3.4%), and specificity (80.6%, SE = 3.4%) (**Table 3**). This indicates that participants

could accurately classify relaxed or anxious states in typically-developing children with very little (i.e., <10 min) training and no contextual information. On average, participants listened to the Biomusic for 12.1 [standard error (SE) = 0.7 s] seconds before registering their initial impression of its associated emotional state. Similarly, classification accuracies for the final impression (registered after listening to the full 80 s sample) were comparable: accuracy (83.8%, SE = 2.4%), sensitivity (90.0%, SE = 3.2%), and specificity (77.5%, SE = 3.6%) were high. This suggests that 12 s of the Biomusic is enough to form an accurate impression of its associated emotional state.

Participants reported high confidence in their ability to classify emotional states based upon the auditory information presented through Biomusic. For both anxious and relaxed Biomusic, the median confidence rating was 4 [interquartile range (IQR) = 2] on a 5-point Likert scale ranging from 1 (low) to 5 (high). Melody was reported by participants as the most influential component for 56.9% of the Biomusic excerpts, followed by drum (25.0%), chords associated with key changes (9.4%), and the cyclic whoosh (8.8%). This trend was apparent for both Relaxed and Anxious state songs.

This first experiment demonstrates the feasibility for Biomusic to convey the information necessary for a listener to distinguish anxious from relaxed state in typical physiological signals.

3. EXPERIMENT 2: DETECTION OF ANXIETY STATE THROUGH BIOMUSIC IN TYPICALLY DEVELOPING CHILDREN AND CHILDREN WITH AUTISM SPECTRUM DISORDERS

Having demonstrated that Biomusic could allow discrimination between relaxed and anxious states of typically-developing children to quickly, with high performance, and high

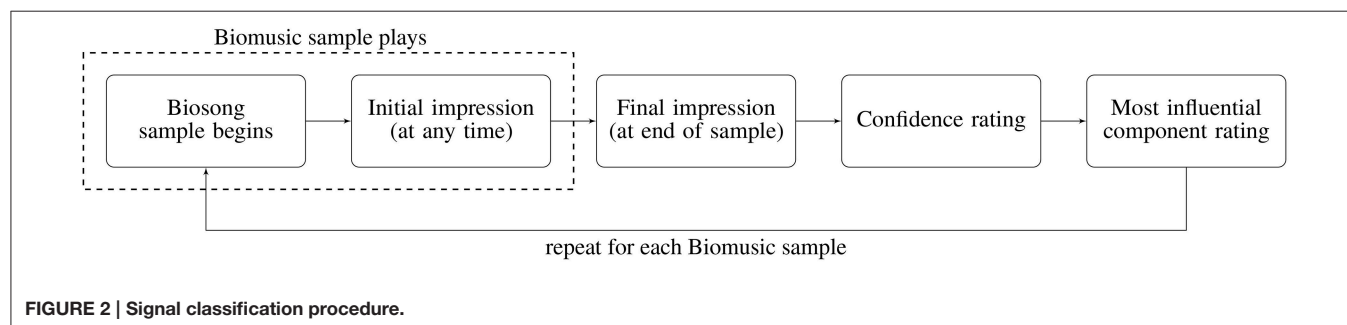
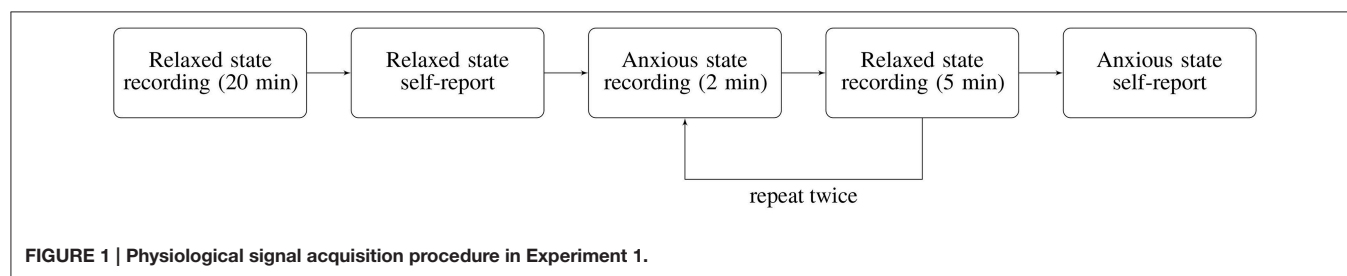


TABLE 3 | Biomusic performance (initial classification).

Experiment 1 Biomusic	Accuracy (%)	Sensitivity (%)	Specificity (%)
Typically-developing (anagram)	84.9, SE = 2.9	89.7, SE = 4.1	80.0, SE = 4.5
Experiment 2 Biomusic	Accuracy (%)	Sensitivity (%)	Specificity (%)
Typically-developing (Stroop)	82.4, SE = 3.9	91.7, SE = 3.9	73.3, SE = 5.1
ASD	83.4, SE = 3.9	83.2, SE = 6.5	83.6, SE = 3.6
Typically-developing (anagram)	84.8, SE = 3.8	85.0, SE = 4.4	85.0, SE = 4.4
All (Experiment 2)	83.9, SE = 2.9	87.1, SE = 3.4	80.6, SE = 3.4

TABLE 4 | Means and standard error of physiological features recorded in Experiment 2 from children with ASD ($n = 5$) and typically developing children ($n = 5$) during the relaxed state condition and the anxiety-provoking condition.

ASD		
Signal feature	Relaxed mean	Anxious mean
EDA (μ S) *	4.822, SE = 1.627	7.028, SE = 1.577
Skin temperature ($^{\circ}$ C)	28.424, SE = 1.512	29.251, SE = 1.918
Blood volume inter-pulse interval (s) *	0.744, SE = 0.024	0.686, SE = 0.022
Respiration inter-breath interval (s)	2.211, SE = 0.337	1.616, SE = 0.174
Typically-developing (Stroop)		
Signal feature	Relaxed mean	Anxious mean
EDA (μ S) *	4.545, SE = 0.924	8.976, SE = 1.902
Skin temperature ($^{\circ}$ C)	30.208, SE = 2.218	29.467, SE = 1.875
Blood volume pulse inter-pulse interval (s)	0.757, SE = 0.044	0.706, SE = 0.036
Respiration inter-breath interval (s)	2.987, SE = 0.733	1.546, SE = 0.093

*Indicates features that were significantly different ($p < 0.05$) between relaxed and anxious state conditions.

classification confidence, we sought to determine if these measures would be similar in Biomusic recorded from children with autism spectrum disorders (ASD). In this experiment, adult participants classified the Biomusic of both typically developing children and children with ASD. In comparison to typically-developing children, children with ASD may have atypical physiological responses during anxiety (Kushki et al., 2013). Children with ASD have been found to exhibit raised heart rate during non-anxious and anxiety-inciting situations, a narrower range of electrodermal activity differences between non-anxious and anxiety-inciting conditions, and atypically anxiety-modulated skin temperature (Kushki et al., 2013). This population may also have difficulty recognizing and expressing their feelings (Lang et al., 2010; Hallett et al., 2013). These children represent a stepping-stone population as we work toward the eventual use case: a child with profound disabilities, with possibly typical or atypical signals, and no ability to report emotion. All protocols were reviewed and approved by the science and ethical review board at Holland Bloorview Kids

Rehabilitation Hospital. Informed consent and assent were obtained from adults and children respectively.

3.1. Materials and Methods

3.1.1. Biomusic

We generated 20 Biomusic samples from a new set of children whose signals were not used to generate Biomusic in Experiment 1. These children had participated in a separate research study at our institution, through which physiological signals had been recorded in anxious and relaxed conditions. 10 Biomusic samples were created from physiological signals of typically-developing children (mean = 10 years, SD = 2 years), and 10 Biomusic samples were created from physiological signals of children with ASD (mean = 11 years, SD = 3 years). For this new group of children, physiological signals were recorded while the Stroop test was administered. Unlike the anagram test used in Experiment 1, this test required no hand movements from the children, who were instructed to limit their movements. The Stroop test has been used to elicit anxiety in previous research (Ozonoff and Jensen, 1999; Christ et al., 2007; Adams and Jarrold, 2009). Although two different tests were used to elicit anxiety in Experiments 1 and 2, this is a methodological strength, as it is important to know if the findings in Experiment 1 were specific to the anagram task, or could be observed in other anxiety-eliciting tasks. The relaxed condition, a low-intensity video remained the same as in Study 1. Means of the physiological signal features of interest are provided in Table 4. There was no significant difference between means of signals from typically developing children and children with ASD in either baseline or anxious conditions. Children self-reported their anxiety state on the State Trait Anxiety Scale—Child version (Spielberger et al., 1970) as previously described. Though this scale was administered to typically-developing children and children with ASD alike, it should be noted that children with ASD may have difficulty recognizing and expressing emotional state (Lang et al., 2010; Hallett et al., 2013). As in Experiment 1, validation of anxiety state required an increase in state score between baseline and anxious conditions. The average difference in normalized state anxiety score observed between relaxed and anxious states was 10.8 points (SE = 3.3) for typically developing children and 8.8 points (SE = 2.7) for children with ASD (Range: For typically-developing children, 44–54 for relaxed and 56–59 for anxious. For children with ASD, 35–56 for relaxed and 51–63 for anxious). This second set of 20 Biomusic samples was combined with the 10 Biomusic samples used in Study 1 to create a total test set of 30 songs. The Experiment 1 Biomusic samples were included in Experiment 2 to explore the potential impact of using two different anxiety-provoking tasks on the results.

3.1.2. Participants

Twelve of the original 16 participants recruited for Experiment 1 also took part in Experiment 2, which took place approximately 1 month after Experiment 1. The four participants who did not participate in Experiment 2 had completed their terms as students and were not available for follow-up.

3.1.3. Evaluation Protocol

Following a short training period (<5 min) to remind participants of the characteristics of anxious and relaxed Biomusic, participants completed the same task as in Experiment 1, but with all 30 test samples of Biomusic presented in random order.

3.1.4. Data Analysis

Data analyses were carried out as in Experiment 1.

4. RESULTS AND DISCUSSION

Table 3 presents the means of participants' initial classification accuracy (83.9%, SE = 2.9%), sensitivity (87.1%, SE = 3.4%), and specificity (80.6%, SE = 3.4%). Performance measures for the subset of songs recorded from typically-developing children ($n = 20$) and the subset associated with children with ASD ($n = 10$) are noted. Classification accuracy associated with initial and final impressions and for Biomusic recorded from children with ASD and from typically developing children were comparable (i.e., within 2.5%). On average, participants listened to the Biomusic for 11.3 s (SE = 0.5 s) before completing the classification task. Class confusion matrices (**Figures 3–5**) demonstrate the errors made by participants during the classification task. False positive errors occurred in 10% of classifications while false negative errors occurred in 7%.

Participants reported high confidence in their interpretation of Biomusic of typically developing children (median = 4, IQR = 1) and children with ASD (median = 4, IQR = 1). When considering which sounds influenced their classification of the Biomusic sample as anxious- or relaxed-state, participants reported melody (electrodermal activity) as the most influential component for 59.3% of the Biomusic excerpts, followed by drum (heart rate) for 24.8%, chords associated with key change (skin temperature) for 10.9%, and the “whoosh” (respiration rate) for 5.0%. As in Experiment 1, this trend was apparent for both relaxed and anxious state songs.

Thus, Experiment 2 demonstrated that in both typically developing children and children with ASD, anxiety can be conveyed through Biomusic. As in Experiment 1, adults' classifications were made quickly, confidently, and with high performance.

5. GENERAL DISCUSSION

This study aimed to evaluate the performance of the Biomusic interface for identifying distinct patterns of physiological signals associated with anxiety, a clinically significant emotion. This work responds to the need for novel AAC technologies to support communication between caregivers and individuals with profound disabilities. This represents an intermediary, but necessary step toward understanding how/if emotional states can be interpreted from Biomusic.

Classification accuracy, sensitivity, and specificity of the Biomusic interface were high and indicated that participants could reliably classify relaxed or anxious states in typically-developing children and in children with ASD. Participants

performed this task with very little (i.e., <10 min) training and no contextual information. On average, participants listened to the Biomusic for about 11–12 s in both experiments before identifying its affective state. The latency of detection is an important measure to assess the appropriateness of Biomusic for practical applications as it relates to how quickly a caregiver could potentially respond to an anxious child. Because Biomusic itself is not computationally intensive, it has been used in real-time, clinical settings (Blain-Moraes et al., 2013) without introducing computational delay.

Analyses were carried out to identify if certain songs were repeatedly misclassified. Of note, three songs were consistently misclassified above chance. Visual inspection of these excerpts revealed that the physiological signals did not show the expected trends associated with their test condition, suggesting that they were perhaps incorrectly labeled by the child, as opposed to incorrectly classified by the listener. Two of these Biomusic excerpts were associated with typically-developing children (baseline, normalized STAI = 31; baseline, normalized

		Classified State		
		anxious	relaxed	total
Actual State	anxious	40%	10%	50%
	relaxed	9%	41%	50%
total		50%	50%	

FIGURE 3 | Confusion matrix for classifications of biomusic from typically developing children (anagram) in Experiment 2.

		Classified State		
		anxious	relaxed	total
Actual State	anxious	45%	5%	50%
	relaxed	15%	35%	50%
total		60%	40%	

FIGURE 4 | Confusion matrix for classifications of biomusic from typically developing children (Stroop) in Experiment 2.

		Classified State		
		anxious	relaxed	total
Actual State	anxious	43%	7%	50%
	relaxed	11%	39%	50%
total		54%	46%	

FIGURE 5 | Confusion matrix for classifications of biomusic from children with ASD in Experiment 2.

STAI = 49) and one with a child with ASD (baseline, normalized STAI = 56).

Participants reported high confidence in their ability to classify emotional states based upon the auditory information presented through Biomusic. In both experiments, melody (electrodermal activity) was reported as the most influential component in the classification task. This was not entirely surprising as the mapping of skin conductance to melody (one of the most prominent and easily distinguishable musical components) was a design decision made to sonically reflect the salience of EDA patterns in physiological manifestations of anxiety as suggested in previous studies (Kreibig, 2010; Kushki et al., 2013). It is notable that we deliberately designed physiological-musical mappings that attempted to preserve the natural qualities and connotations of the biosignal in question (e.g., expiration length was matched to an embellishment that resembled the sound of an exhalation). Such considerations may improve the ability of users to remember and interpret the significance of musical changes. However, future research is required to determine if other physiological-musical mappings could further improve accuracy, particularly for individuals with atypical electrodermal activity responses. Interestingly, the electrodermal activity of children with ASD has been reported to respond atypically to anxious situations (Kushki et al., 2013). However, participants in Experiment 2 were able to classify signals of typically developing children and children with ASD with similar high performance.

Of note, the initial classification sensitivity of the Biomusic (89.7% in Experiment 1, 84.9% in Experiment 2) compares well with that of the established, observational scale, the modified Yale Preoperative Anxiety Scale (85%), which has been previously validated against the gold-standard State Trait Anxiety Inventory for Children (Kain et al., 1997). Accuracy is also similar to that achieved via algorithmic pattern analysis of physiological signals, though direct comparison is often difficult due to methodological differences between studies (Wen et al., 2014). For example, Rani et al. (2006) evaluated K Nearest Neighbor, Regression Tree, Bayesian Network, and Support Vector Machine algorithms

for the detection of anxiety in healthy adults, elicited using an anagram test. Inputs for these classifiers were chosen by hand on a case-by-case basis for each participant, due to physiological response variability across participants. Accuracy of these classifiers ranged between 80.38 and 88.86% further research with larger and more heterogeneous samples is warranted, our Biomusic system appeared to be comparable for anxiety detection in typically-developing participants and those with ASD. This potential for generalization is essential for the intended use scenario where there is a high potential for condition- and person-specific variability, but it may be impossible to train person-specific algorithms given the lack of self-report.

5.1. Study Limitations and Future Work

Note that generalizations of this study should be made with caution. While the target user for Biomusic is a child with profound disabilities, having possibly typical or atypical physiological signals and no ability to self-report emotional state, this presents a paradox to system validation: the children most in need of the Biomusic interface are also the population for which Biomusic development/evaluation is most challenging. As such, it is important to first establish the performance of the Biomusic interface where self-report is possible in order to maximize our understanding of the technology and potential success when translating to clinical populations with more severe communication challenges. Although gold standard measures were used to define anxious and relaxed states through self-reports, it is near impossible to ascertain these emotions with absolute certainty, even in the typically-developing group. While in the majority of cases, visual analysis of the physiological signals also supported the assigned classification labels, there were three notable exceptions. These songs were all associated with poor classification accuracies, which may indicate that the overall detection accuracies reported are quite conservative.

The two training samples were not used in the testing phase of either experiment. However, separate data segments from the same child were used to generate two distinct Biomusic samples that were included in the test set. Plots of the training and testing segments of this child's physiological signals are included in the Supplementary Materials. Three researchers listened to the Biomusic prior to conducting these experiments and were unable to discern a child-specific "signature" to the sonification that could contribute to a learning effect. If such a signature is present, it is unlikely that it could have been discerned with a single 80-s sample of an anxious and relaxed state as presented in participant orientation. The classification accuracy of this child's test samples was 82%. For comparison, the average classification accuracy of children's test samples was 82% in the typically developing group that completed anagrams. The average classification accuracy of all children's test samples was 81%.

In a follow-up questionnaire, we asked the 16 adult participants the question, "Are the songs pleasant to listen to?" 15 participants responded "Yes," indicating that Biomusic is not only informative, but aesthetically appealing to the listener. Further research will be needed to probe the aesthetics and usability of Biomusic in a care context.

Given the promising results of this study, future research is warranted with larger and more diverse samples of children and listeners (e.g., parents, nurses, therapists, physicians) in a real-time context. It would be interesting to explore if Biomusic has the specificity required to distinguish high intensity/negative valence states (e.g., anxiety) from high intensity/positive valence states (e.g., excitement), or if valence is more effectively identified through contextual cues. Our experiment was not designed to explore multiple emotional states or to evaluate how well transitions between emotional states can be detected. The complexity of affective responses (e.g., prevalence of mixed emotions and the fact that transitions between emotions in practice can rarely be discretely defined as the intensity and valence of emotions are often considered to lie on a spectrum) makes these experiments extremely difficult to design and carry out, particularly in populations where self-report is a challenge. The complexity and continuous nature of affective responses is in fact one of the motivators underlying the development of our Biomusic interface which can present multiple physiological signals continuously as a supplement to overt affective responses and contextual information. This study was designed to investigate the usability and validity of the interface in a well-controlled environment prior to deployment in more contextually-rich and clinically relevant environments.

Our anxiety-provoking test conditions generated Biomusic samples that were distinguishable from relaxed-condition samples, even though elicited anxiety was likely moderate. For reference, the children in our samples had an average normalized state anxiety score of 55.9 points during the anxiety eliciting tasks. For children in another study who were told to imagine a school test-taking scenario ($n = 913$), the average normalized state anxiety score on the STAI was 41.8 points for males and 43.8 points for females (Spielberger et al., 1970). For children pre-surgery ($n = 39$), average normalized state anxiety scores of 69.89 and 71.90 points in females have been reported (Alirezai et al., 2008). In comparing these different contexts, we hypothesize that the tasks used in this study were only able to elicit moderate increases in anxiety. Future work, likely in real-world contexts, are needed to explore the utility of the Biomusic interface across a wider range of intensities and emotions.

REFERENCES

- Adams, D., and Oliver, C. (2011). The expression and assessment of emotions and internal states in individuals with severe or profound intellectual disabilities. *Clin. Psychol. Rev.* 31, 293–306. doi: 10.1016/j.cpr.2011.01.003
- Adams, N. C., and Jarrold, C. (2009). Inhibition and the validity of the stroop task for children with autism. *J. Autism Dev. Disord.* 39, 1112–1121. doi: 10.1007/s10803-009-0721-8
- Affrunti, N. W., and Woodruff-Borden, J. (2014). Parental perfectionism and overcontrol: examining mechanisms in the development of child anxiety. *J. Abnorm. Child Psychol.* 43, 517–529. doi: 10.1007/s10802-014-9914-5
- Alirezai, N., Ashjaei, B., Shahrivar, Z., Moharari, F., Mahmoudi-gharaei, J., and Parizad, J. (2008). Effect of preoperative play interventions on post surgery anxiety. *Iran. J. Psychiatry* 3, 20–24. Available online at: <http://ijps.tums.ac.ir/index.php/ijps/article/view/488>
- Blain-Moraes, S., Chessier, S., Kingsnorth, S., McKeever, P., and Biddiss, E. (2013). Biomusic: a novel technology for revealing the personhood of people with

Further, it would also be important to evaluate the performance of the system during a real-time caregiving activity where contextual and behavioral cues are also integrated. Qualitative studies would be well-suited to probe these lines of investigation. Future studies could also assess whether sensitivity, specificity, and accuracy of detections remain high when individuals must engage in other tasks as is typical in real-life care-giving settings and if Biomusic is an acceptable interface for long-term use. Biomusic may also show promise as a tool for emotional self-regulation, particularly for children who struggle to identify and express their emotional state. This is an area worth further exploration.

AUTHOR CONTRIBUTIONS

SC, EH, and EB developed the study protocols. EA and AK contributed to study design for Experiment 2. SC, EH, and EB wrote the manuscript. SC and EH contributed equally to this paper.

FUNDING

The authors report no declarations of interest. This work was supported in part by a Discovery Grant from Natural Sciences and Engineering Council of Canada (#371828-11); the Ontario Ministry of Training, Colleges and Universities through an award to EH; the Holland Bloorview Kids Rehabilitation Hospital Foundation; and the Ward Family summer student research program through an award to SC.

ACKNOWLEDGMENTS

The authors would like to thank Stefanie Blain-Moraes and Pierre Duez who conceptualized and developed the original Biomusic software. We would also like to thank Hasmita Singh for her support during data collection.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fnins.2016.00401>

profound multiple disabilities. *Augment. Altern. Commun.* 29, 159–173. doi: 10.3109/07434618.2012.760648

- Christ, S. E., Holt, D. D., White, D. A., and Green, L. (2007). Inhibitory control in children with autism spectrum disorder. *J. Autism Dev. Disord.* 37, 1155–1165. doi: 10.1007/s10803-006-0259-y
- Dubus, G., and Bresin, R. (2013). A systematic review of mapping strategies for the sonification of physical quantities. *PLoS ONE* 8:e82491. doi: 10.1371/journal.pone.0082491
- Franchini, K. G., and Cowley, A. W. J. (eds.). (2004). “Autonomic control of cardiac function,” in *Primer on the Autonomic Nervous System* (London: Elsevier Academic Press), 134–138.
- Hallett, V., Lecavalier, L., Sukhodolsky, D. G., Cipriano, N., Aman, M. G., McCracken, J. T., et al. (2013). Exploring the manifestations of anxiety in children with autism spectrum disorders. *J. Autism Dev. Disord.* 43, 2341–2352. doi: 10.1007/s10803-013-1775-1
- Hermann, T. (2008). “Taxonomy and definitions for sonification and auditory display,” in *Proceedings of the 14th International Conference on*

- Auditory Display* (Paris). Available online at: <https://smartech.gatech.edu/handle/1853/49960>
- Hogg, J., Reeves, D., Roberts, J., and Mudford, O. C. (2001). Consistency, context and confidence in judgements of affective communication in adults with profound intellectual and multiple disabilities. *J. Intellect. Disabil. Res.* 45, 18–29. doi: 10.1046/j.1365-2788.2001.00289.x
- Jennings, J. R., Kamarck, T., Stewart, C., Eddy, M., and Johnson, P. (1992). Alternate cardiovascular baseline assessment techniques: vanilla or resting baseline. *Psychophysiology* 29, 742–750. doi: 10.1111/j.1469-8986.1992.tb02052.x
- Kain, Z. N., Mayes, L. C., Cicchetti, D. V., Bagnall, A. L., Finley, J. D., and Hofstadter, M. B. (1997). The yale preoperative anxiety scale: how does it compare with a “gold standard”? *Anesth. Analg.* 85, 783–788. doi: 10.1213/00000539-199710000-00012
- Kaslow, N. J., Tanenbaum, R. L., Abramson, L. Y., Peterson, C., and Seligman, M. E. P. (1983). Problem-solving deficits and depressive symptoms among children. *J. Abnorm. Child Psychol.* 11, 497–501. doi: 10.1007/BF00917078
- Kreibitz, S. D. (2010). Autonomic nervous system activity in emotion: a review. *Biol. Psychol.* 84, 394–421. doi: 10.1016/j.biopsycho.2010.03.010
- Kushki, A., Drumm, E., Pla Mobarak, M., Tanel, N., Dupuis, A., Chau, T., et al. (2013). Investigating the autonomic nervous system response to anxiety in children with autism spectrum disorders. *PLoS ONE* 8:e59730. doi: 10.1371/journal.pone.0059730
- Lang, R., Regeister, A., Lauderdale, S., Ashbaugh, K., and Haring, A. (2010). Treatment of anxiety in autism spectrum disorders using cognitive behaviour therapy: a systematic review. *Dev. Neurorehabil.* 13, 53–63. doi: 10.3109/17518420903236288
- Ozonoff, S., and Jensen, J. (1999). Brief report: specific executive function profiles in three neurodevelopmental disorders. *J. Autism Dev. Disord.* 29, 171–177. doi: 10.1023/A:1023052913110
- Rani, P., Liu, C., Sarkar, N., and Vanman, E. (2006). An empirical study of machine learning techniques for affect recognition in humanrobot interaction. *Pattern Anal. Appl.* 9, 58–69. doi: 10.1007/s10044-006-0025-y
- Rimm-Kaufman, S. E., and Kagan, J. (1996). The psychological significance of changes in skin temperature. *Motiv. Emot.* 20, 63–78. doi: 10.1007/BF02251007
- Sanderson, P. (2006). The multimodal world of medical monitoring displays. *Appl. Ergon.* 37, 501–512. doi: 10.1016/j.apergo.2006.04.022
- Spielberger, C. D., Gorsuch, R. L., and Lushene, R. E. (1970). *Manual for the State-Trait Anxiety Inventory*. Menlo Park, CA: Mind Garden, Inc.
- Vetrugno, R., Liguori, R., Cortelli, P., and Montagna, P. (2003). Sympathetic skin response. *Clin. Auton. Res.* 13, 256–270. doi: 10.1007/s10286-003-0107-5
- Vos, P., De Cock, P., Munde, V., Petry, K., Van Den Noortgate, W., and Maes, B. (2012). The tell-tale: What do heart rate; skin temperature and skin conductance reveal about emotions of people with severe and profound intellectual disabilities? *Res. Dev. Disabil.* 33, 1117–1127. doi: 10.1016/j.ridd.2012.02.006
- Ward, L. G., Friedlander, M. L., and Silverman, W. K. (1987). Children's depressive symptoms, negative self-statements, and causal attributions for success and failure. *Cogn. Ther. Res.* 11, 215–227. doi: 10.1007/BF01183266
- Watson, M., and Sanderson, P. M. (2001). “Intelligibility of sonifications for respiratory monitoring in anesthesia,” *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, Vol. 45 (Minneapolis, MN), 1293–1297. doi: 10.1177/154193120104501708
- Wen, W., Liu, G., Cheng, N., Wei, J., Shangguan, P., and Huang, W. (2014). Emotion recognition based on multi-variant correlation of physiological signals. *IEEE Trans. Affect. Comput.* 5, 126–140. doi: 10.1109/TAFFC.2014.2327617
- Wu, D., Li, C.-Y., and Yao, D.-Z. (2009). Scale-free music of the brain. *PLoS ONE* 4:e5915. doi: 10.1371/journal.pone.0005915
- Yokoyama, K., Ushida, J., Sugiura, Y., Mizuno, M., Mizuno, Y., and Takata, K. (2002). Heart rate indication using musical data. *IEEE Trans. Biomed. Eng.* 49, 729–733. doi: 10.1109/TBME.2002.1010857

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.

Copyright © 2016 Cheung, Han, Kushki, Anagnostou and Biddiss. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Interactive Sonification of Spontaneous Movement of Children—Cross-Modal Mapping and the Perception of Body Movement Qualities through Sound

Emma Frid^{1*}, Roberto Bresin¹, Paolo Albornò² and Ludvig Elblaus¹

¹ Sound and Music Computing, Media Technology and Interaction Design, School of Computer Science and Communication, KTH Royal Institute of Technology, Stockholm, Sweden, ² Casa Paganini - Infomus Research Centre, DIBRIS Dipartimento di Informatica, Bioingegneria, Robotica e Ingegneria, Università di Genova, Genova, Italy

OPEN ACCESS

Edited by:

Diego Minciocchi,
University of Florence, Italy

Reviewed by:

Agnes Roby-Brami,
French Institute of Health and Medical
Research (INSERM), France
Sankaranarayani Rajangam,
Duke University, USA

*Correspondence:

Emma Frid
emmafrid@kth.se

Specialty section:

This article was submitted to
Neuroprosthetics,
a section of the journal
Frontiers in Neuroscience

Received: 13 July 2016

Accepted: 27 October 2016

Published: 11 November 2016

Citation:

Frid E, Bresin R, Albornò P and
Elblaus L (2016) Interactive
Sonification of Spontaneous
Movement of Children—Cross-Modal
Mapping and the Perception of Body
Movement Qualities through Sound.
Front. Neurosci. 10:521.
doi: 10.3389/fnins.2016.00521

In this paper we present three studies focusing on the effect of different sound models in interactive sonification of bodily movement. We hypothesized that a sound model characterized by continuous smooth sounds would be associated with other movement characteristics than a model characterized by abrupt variation in amplitude and that these associations could be reflected in spontaneous movement characteristics. Three subsequent studies were conducted to investigate the relationship between properties of bodily movement and sound: (1) a motion capture experiment involving interactive sonification of a group of children spontaneously moving in a room, (2) an experiment involving perceptual ratings of sonified movement data and (3) an experiment involving matching between sonified movements and their visualizations in the form of abstract drawings. In (1) we used a system constituting of 17 IR cameras tracking passive reflective markers. The head positions in the horizontal plane of 3–4 children were simultaneously tracked and sonified, producing 3–4 sound sources spatially displayed through an 8-channel loudspeaker system. We analyzed children's spontaneous movement in terms of energy-, smoothness- and directness-index. Despite large inter-participant variability and group-specific effects caused by interaction among children when engaging in the spontaneous movement task, we found a small but significant effect of sound model. Results from (2) indicate that different sound models can be rated differently on a set of motion-related perceptual scales (e.g., expressivity and fluidity). Also, results imply that audio-only stimuli can evoke stronger perceived properties of movement (e.g., energetic, impulsive) than stimuli involving both audio and video representations. Findings in (3) suggest that sounds portraying bodily movement can be represented using abstract drawings in a meaningful way. We argue that the results from these studies support the existence of a cross-modal mapping of body motion qualities from bodily movement to sounds. Sound can be translated

and understood from bodily motion, conveyed through sound visualizations in the shape of drawings and translated back from sound visualizations to audio. The work underlines the potential of using interactive sonification to communicate high-level features of human movement data.

Keywords: interactive sonification, movement analysis, movement sonification, mapping, motion capture, perception

1. INTRODUCTION

Interactive sonification is the discipline of interactive representation of data and data relationships by means of sound. If properly designed, it serves as a powerful and effective information display. In order to successfully design sonification applications one has to consider how meaning is ascribed to certain sounds. Closely linked to this topic is the notion of *mapping*, i.e., how input parameters are mapped to auditory output parameters in order to convey properties of the data through perceptually relevant acoustic features. There is a large set of possible mappings that could be used within the context of sonification of human movement (see e.g., Dubus and Bresin, 2013 for an overview). However, only a small subset of these mappings will produce perceptually relevant results (Roddy and Furlong, 2014). Our work is motivated by the fact that if links between sound properties and movement can be found, design of auditory information displays and sonification applications could be improved through use of more perceptually relevant and intuitive mappings. The work presented in this paper serves as first investigation in a series of attempts aimed at finding perceptually relevant attributes of sound synthesis for sonification of human movement. The aim is to investigate if different sound models can evoke different associations to motion and thereby induce different spontaneous movement characteristics.

It is clear that musical sounds can induce human body movement, but can certain properties of a sound influence, and be associated to, specific properties of bodily movement? The notion of *embodied cognition* assumes that the body is involved in, and required for, cognitive processes (Lakoff and Johnson, 1980, 1999). Following an embodied cognition perspective, we can approach music by linking perception to our body movement (Leman, 2007). Bodily movements can thus be said to reflect, imitate or support understanding of the content and structure of music (Burger et al., 2013). In our study we aim to expand on this notion of a link between music and motion to non-musical sounds. Based on the notion that music carries the capacity to activate the embodied domain of sounds by inducing movement (see for example Zentner and Eerola, 2010), we assume that spontaneous bodily movement to interactive sonification may reflect and imitate aspects of sound produced by the sonification system.

Following the ecological approach to auditory perception (Gaver, 1993) and the notion that no sound is produced without movement, we formulate the hypothesis that a sonification mode characterized by a sustained and continuous amplitude envelope will be associated to properties related

to smooth, continuous movements. A non-continuous sound characterized by acoustically abrupt events should accordingly be associated to properties related to non-continuous movements. We designed two sound models based on these hypotheses: one continuous sound model and one sound model characterized by a high level of amplitude modulation and sudden amplitude irregularities. For comparative purposes, we also designed a model that was considered to be perceptually in between these two models in terms of irregularities in amplitude.

The three sound models were used in three different experiments in which we investigated the relationship between above mentioned properties of a sound and bodily movements. Study 1 focused on investigating if three sound models would evoke different movement characteristics among children when moving freely in a room. Assuming that spontaneous movement in an interactive sonification task can be understood as a means of exploring the presented sound, our hypothesis was that movement at a specific point of measurement could be influenced by the specific sound model used at that time point. Study 2 focused on investigating if sound models used in 1 were rated differently on a set of motion-related perceptual scales by another group of participants. Our hypothesis was that different sound models would be rated differently. Study 3 focused on investigating if drawings depicting sounds recorded in Study 1 could be easily identified and matched to respective sound model in a forced-choice experiment. We hypothesized that participants would be able to correctly match recordings of one sound model to an abstract visual representation, i.e., a sound visualization in the form of a drawing, of the same sound model.

2. BACKGROUND

2.1. Sound and Movement

The link between sound and movement has been investigated in numerous studies throughout the years. Following an ecological perception point of view, interpretation of sounds is founded on knowledge on gestural actions required to produced the sound in question (Gaver, 1993). Keller and Rieger (2009) found that simply listening to music can induce movement. Janata and Grafton (2003) showed that passive music listening can involve activation of brain regions concerned with movement. As stated by Godøy and Jensenius (2009), it is not far-fetched to suggest that listeners' music-related movements often match overall motion and emotional features of a musical sound: guidelines for traditional gestures for musical conductors state that *legato* elicits smooth, connected gestures, while *accented* and *rhythmic* music elicits shorter and more jerky movements (Blatter, 2007).

A couple of studies have focused on spontaneous movement to musical sounds and how such movement trajectories can be analyzed and classified (Casciato et al., 2005; Godøy et al., 2005; Haga, 2008). Up to this point, few studies have focused on motion analysis of children's spontaneous movement patterns to sound and music (see for example Zentner and Eerola, 2010), especially in the context of interactive sonification (e.g., Källblad et al., 2008). The topic of spontaneous movement to musical sounds is related to research on *music-induced movement*, where focus lies on how people react corporeally to music. Several factors, such as musical features and individual factors, can affect the characteristics of such music-induced movements (Burger et al., 2013). Leman (2007) defined three components that could influence corporeal articulations in music-induced movement: *synchronisation*, *embodied attuning* and *empathy*. *Synchronisation* is a fundamental component that deals with synchronisation to a beat; *Embodied attuning* concerns the linkage between body movements to musical features more complex than a basic beat (such as e.g., harmony, melody, rhythm, tonality and timbre); *Empathy* links musical features to emotions and expressivity. According to Leman (2007), spontaneous movements to music appear to be closely related to predictions of local bursts of energy in the audio stream, such as beat and rhythms. Although multiple studies have focused on the effect of synchronisation/beat (Toiviainen et al., 2009; Burger et al., 2014) and expressive features in music (Buhmann et al., 2016) in this context, rather few projects have focused on the effect of more complex musical features such as e.g., timbral properties and their effect on music-induced spontaneous movement (see for example Burger et al., 2013).

There are some examples of studies on the relationship between sound and spontaneous movement in which participants have been instructed to trace sounds that they hear, i.e., to trace the perceptual features of a sound (see for e.g., Godøy et al., 2006, 2010; Nymoen et al., 2010). This is usually referred to as “*sound-tracing*”, a concept first introduced by Godøy et al. (2005) which is defined as the process of rendering perceptual features of sound through body motion. In a study by Caramiaux et al. (2014) in which participants were instructed to synchronously perform a gesture to a sound, it was found that if the cause of a sound could be identified, participants would perform spontaneous gestures that attempted to mimic the action producing the sound. However, if the sound contained no perceivable causality, the spontaneous movement would trace contours related to acoustic features. Moreover, abstract sounds were found to result in less gesture variability.

In the study presented in this paper, we use drawings as a means of describing perceived sounds. Drawings can, similar to words and gestures, serve as a high-level approach to description. Drawings have previously been used in contemporary music to either compose or describe music (Thiebaut et al., 2008). The association between sound and shapes has been investigated in numerous studies throughout the years in research on what is usually referred to as *shape symbolism*. Shape symbolism is a family of multisensory phenomena in which shapes give rise to experiences in different sensory modalities, the most common example being the “*Bouba-Kiki effect*” (Kohler, 1929,

1947), in which the words “kiki/takete” and “bouba/maluma” are associated with angular vs. rounded shapes. The hypothesis in our study, as previously suggested by Merer et al. (2013), is that drawings are a relevant means of describing motion in an intuitive way, and that the use of abstract sounds, in which the physical sources can not be easily identified, provide relevant and unbiased keys for investigating the concepts of motion.

2.2. Movement Analysis

In order to analyze movements of the children participating in the motion capture experiment described in Section 3, we extracted motion features from motion capture recordings. We followed the multi-layered conceptual framework for the analysis of expressive gestures proposed by Camurri et al. (2016). This framework consists of four layers allowing for both a bottom-up (from Layer 1 to 4) and top-down (from Layer 4 to 1) analysis; Layer 1 - Physical signals (e.g., positional data captured by IR cameras), Layer 2 - Low-level features (e.g., velocity), Layer 3 - Mid-level features (e.g., smoothness), Layer 4 - Expressive qualities (e.g., emotion). Following this layered approach, we included low-level features (i.e., Energy Index, *EI*, and Smoothness Index, *SI*, of head movements) and one mid-level feature (i.e., Directness Index, *DI*). The above mentioned features have previously been used in different contexts and research purposes in order to describe the expression of human gestures (Camurri et al., 2002), for investigating the emotional mechanisms underlying expressiveness in music performances (Castellano et al., 2008) and as potential descriptors to infer the affective state of children with Autism Spectrum Condition (Piana et al., 2013). We decided to include the Energy Index (*EI*) in the investigation since we hypothesized that this feature could be highly correlated with properties of the movements elicited by different sound models. Moreover, we decided to include Smoothness- and Directness-Index (*SI* and *DI*), since we were interested in how continuous the movement trajectories of the children would be for different sound models. A description of each feature can be found below.

2.2.1. Energy Index

This feature concerns the overall energy spent by the user during a movement and is computed as the total amount of displacement in all of the tracked points. Given a two-dimensional tracking information, we can define velocity of the i -th tracked point at frame f as:

$$v_i(f) = \sqrt{(\dot{x}_i(f))^2 + (\dot{y}_i(f))^2} \quad (1)$$

Where \dot{x}_i and \dot{y}_i are the first derivatives of the position coordinates. The Energy Index *EI* can then be computed as:

$$EI(f) = \frac{1}{2} \sum_{i=1}^J m_i \cdot v_i^2(f), \quad (2)$$

where J is the number of tracked points (or joints) of the subject's body. In the context of our experimental setup, we computed the energy of each subject by tracking only their head movements, using a single rigid body (a combination of markers in a unique

pattern that could be identified by the tracking system). $EI(f)$ is therefore an approximation of the head's kinematic energy estimated as one single point's kinetic energy (max number of points $J = 1$). To simplify the calculation, the weight m_1 was also set to 1.

2.2.2. Smoothness Index

The mathematical concept of smoothness is associated to the rate of variation of a function waveform. A smooth function varies "slowly" over time; smooth functions belong to the C^∞ class, i.e., functions that can be derived an infinite number of times. The third derivative of the movement position has often been used as descriptor for the smoothness of a motion trajectory (Flash and Hogan, 1985). Our algorithm for computing smoothness is based on the studies made by Viviani and Terzuolo (1982) and Todorov and Jordan (1998) that show an existing correlation between trajectory curvature and velocity. The Smoothness Index SI can be computed from the trajectory curvature and velocity. The curvature k measures the rate at which a tangent vector to the trajectory curve changes as the trajectory bends. As an example, the trajectory of a rigid body following the contour of a geometric shape, such as a square, will bend sharply in some points. This trajectory will thus be characterized by high curvature and low smoothness. In contrast, a straight line trajectory will have zero curvature and infinite smoothness (Glowinski et al., 2011).

We can define a bi-dimensional trajectory consisting of collection of consecutive coordinates $x_i(f)$ and $y_i(f)$ of the i -th tracked joint at frame f (in our case max i corresponds to one tracked point) and its velocity $v_i(f)$. The trajectory curvature k can be computed as:

$$k(x_i(f), y_i(f)) = \frac{\dot{x}_i(f) \cdot \ddot{y}_i(f) - \dot{y}_i(f) \cdot \ddot{x}_i(f)}{\left(\sqrt{\dot{x}_i^2(f) + \dot{y}_i^2(f)}\right)^{3/2}}, \quad (3)$$

where $\dot{x}_i(f)$, $\dot{y}_i(f)$, $\ddot{x}_i(f)$ and $\ddot{y}_i(f)$ are the first- and second-order derivatives of the coordinates x and y . In our work we define the Smoothness Index SI as the Pearson correlation coefficient ρ computed on the quantities $\log(k)$ and $\log(v)$. This index gives a measure of the relationship between velocity and curvature and it is calculated as:

$$\rho(k, v) = \frac{\text{cov}[\log(k), \log(v)]}{\sigma_{\log(k)} \cdot \sigma_{\log(v)}} \quad (4)$$

In the calculus of the Smoothness Index SI , k and v are evaluated over short time windows (30 ms). Therefore, we could approximate the covariance $\text{cov}[\log(k), \log(v)]$ by 1, as the k and v variate (or not) approximately at the same rate. We can then simplify the definition of the Smoothness Index SI to:

$$SI = \rho(k, v) = \frac{1}{\sigma_{\log(k)} \cdot \sigma_{\log(v)}} \quad (5)$$

2.2.3. Directness Index

The Directness Index DI is a measure of how much a given trajectory, generated by a tracked joint (in our case, point), is

direct or flexible. DI has been detected as one of the main motion features in the process of recognizing emotions (De Meijer, 1989). A direct movement is characterized by almost rectilinear trajectories. The DI is computed as the ratio between the length of the straight line connecting the first and last point of a trajectory and the sum of the lengths of each segment constituting the trajectory itself. Therefore, the more the DI value is near to value 1, the more direct is the trajectory. In the case where we have a two-dimensional trajectory, the Directness Index DI can be computed as:

$$DI = \frac{\sqrt{(x_{\text{end}} - x_{\text{start}})^2 + (y_{\text{end}} - y_{\text{start}})^2}}{\sum_{i=k}^N \sqrt{(x_{k+1} - x_k)^2 + (y_{k+1} - y_k)^2}}, \quad (6)$$

where x_{start} , y_{start} and x_{end} , y_{end} are the coordinates of the trajectory's start- and end-points in the 2D space and N represents the length of the trajectory.

3. STUDY 1: MOTION CAPTURE EXPERIMENT

3.1. Method

The first study focused on investigating if three sound models would evoke different movement characteristics among children when moving freely in a room. Our hypothesis was that the specific sound model used at a particular time point could influence spontaneous movement of the children at a specific point of measurement. To investigate this hypothesis, we carried out a repeated measures experiment in which longitudinal data of participants' movements was collected in a motion capture room fitted with an 8-channel loudspeaker system. For each participant, x- and y-position and velocity of rigid body markers (placed on the head) were tracked. The data was fed to a sonification software providing real-time feedback of the performed movements.

3.1.1. Participants

Two pre-school classes (4–5 vs. 5–6 years) from a kindergarten in Stockholm participated in the experiment. However, children in the age group 4–5 years failed to follow instructions in the experiment and were therefore excluded from the analysis, giving a total of $n = 11$ participants (2 boys and 9 girls, age 5–6 years, mean = 5.36, SD = 0.5). The participants were divided into groups of 3–4 participants, with a total of 3 groups. Each group participated in two sessions of the experiment: one recording session in the morning and one in the afternoon. A teacher from the kindergarten was always present during each session. We decided to work with children based on the assumption that younger participants would act more spontaneous than adults in a task involving free movement (since spontaneous movement is an integral aspect of active play). Moreover, it has been found in a study by Temmerman (2000) that children tend to have positive attitudes toward activities that provide opportunity to move freely to music.

There was no need for ethics approval since neither of the experiments presented in this paper involved deception or

stressful procedures¹. The research presented no risk to harm participants. Parents were required to return signed consent forms in which they agreed to their child's participation in the study. The informed consent included information about the study and the task; the form was distributed to make an effort to enable children to understand, to the degree they are capable, what their participation in the research would involve. All parents consented to both participation and possible future publishing of photos taken during the experimental session.

3.1.2. Equipment

The experiment was run at the Multimodal Interaction and Performance Laboratory (PMIL), dedicated to experiments involving motion capture and spatial audio, at KTH Royal Institute of Technology, Stockholm, Sweden. The experimental setup consisted of several different software and hardware systems that together formed a chain, starting with the motion capture system and ending with the generation and spatialization of the sound. The motion capture system used was an *Optitrack Prime 41*² setup using 17 IR cameras tracking passive reflective markers. The frequency of acquisition was 180 frames per seconds (resolution 4.1 MP, latency 5.5 ms). The cameras were placed on a circle, at a height of 2.44 m, following the perimeter of the room (the room measured 5.30 × 6.20 m). The trackable area in which the children were instructed to move was a rectangular area measuring 4.66 × 5.40 m, marked using tape on the floor. The system was controlled by the *Optitrack Motive*³ software. While tracking and recording, the *Motive* software also streamed data over a local network to a second computer, using the *NatNet*⁴ streaming protocol. A custom piece of software, written in C++, was running on the second computer that received the incoming NatNet data stream. Data was visualized and some additional calculations were performed on this second computer, whereafter original data was packaged with the calculated secondary data and send forward using the *Open Sound Control* (OSC) format⁵.

The final part of the chain was a third computer that took care of logging, sound generation and spatialization. The logging application, also a custom C++ solution, took every incoming OSC-message, added a local time stamp and wrote it to disk. The rationale for the double logging was to ensure that any issues caused by network transmission problems could be identified by comparing the recorded motion capture data in the head of the processing chain with the resulting data that actually arrived at the sound-producing computer. For the audio, a *Max/MSP*⁶ patch was used to both generate and spatialize the audio, as well as automatically run through the set of sound models for each session in the experiment. The *Max/MSP* patch also reported every change of state in the experiment with an OSC message to the logging application, meaning that the switching between

sound models was recorded together with the movement data. A regular digital video camera was used to record all experiment sessions. The camera was mounted on a tripod in a corner of the lab and was kept recording for the entire duration of the sessions.

After the experiment, motion capture data was pre-processed and segmented whereafter each segmented file was streamed via OSC to *EyesWeb* for feature extraction. The *EyesWeb XMI platform*⁷ is a development and prototyping software environment for both research purposes and interactive applications which provides a set of software modules for analysis of human movements and behavior (Camurri et al., 2003). In the present study we used *EyesWeb* libraries for analysis of 2D movement trajectories to extract expressive motion feature describing human movements both at a local temporal granularity (Energy Index), and at the level of entire movement unit (Smoothness- and Directness Index). A movement unit can be for example a single movement or a whole phrase. In this particular study, a movement unit is defined as a time window with a specific duration.

3.1.3. Stimuli

Each participant group was presented with five different auditory conditions. These conditions consisted of the three different sound models S1–S3 (sonification models) and excerpts from two pieces of music M1 and M2 (M1: “Piano Trio No. 1 in D Minor, Op. 49: II. Andante con moto tranquillo” by Felix Mendelssohn, and M2: “Le Carneval des Animaux: Final” by Camille Saint-Saëns). The conditions S1, S2, and S3 were interactive in the sense that the children's movement affected the generated sound. The musical conditions M1 and M2 were not interactive; children's movement was not mapped to the sound.

The musical pieces for conditions M1 and M2 were chosen since they in previous studies had been found to elicit certain emotions (Västfjäll, 2002; Camurri et al., 2006). M1 has been found to communicate tenderness and we therefore decided that this piece would be appropriate for the introductory part of the experiment. The purpose of including the M1 condition was to let the children get acquainted with the task and start moving to sound. M1 could however also have been used as a control condition, for comparative purposes. M2 has been found to elicit happiness and was therefore included in order to reward the children after successfully completing the experimental task.

For the sonification conditions S1–S3, we opted for sound models based on filtered noise. This decision was based on previous studies indicating that sounds with rich spectral content have been found to be more appealing to children with disabilities than other sounds (Hansen et al., 2012) and that the sound of speed and acceleration can be ecologically represented using simplified sound models reminding of the sound of wind, as for example in the sonification of rowing actions (Dubus and Bresin, 2015). Three sound models based on filtered white noise were defined: one producing smooth, wind-like sounds (S1); one model producing somewhat less smooth sounds characterized by more abruptly interrupted amplitude envelopes (S2); and one

¹For the management of participants' personal data, we followed rules according the KTH Royal Institute of Technology's Ethics Officer (Personuppgiftsombud).

²Optitrack Prime 41: <https://www.optitrack.com/products/prime-41/>

³Optitrack Motive: <https://www.optitrack.com/products/motive/>

⁴NatNet SDK: <https://www.optitrack.com/products/natnet-sdk/>

⁵Open Sound Control: <http://opensoundcontrol.org>

⁶Max/MSP: <https://cycling74.com/products/max/>

⁷EyesWeb XMI platform http://www.infomus.org/eyesweb_ita.php

producing very choppy and clicking sounds due to a high-level of interruptions in the amplitude envelope (S3).

For each sound model S1–S3, low-level movement parameters (velocity, and x- and y-position of the participant in the horizontal plane) were mapped to acoustic parameters. Mappings were chosen among the most frequently used ones in previous research; for example location to spatialization, velocity to pitch and energy to loudness (see Dubus and Bresin, 2013, for a complete review of mappings). During the experiment, each child represented a sound source. Spatialization was done in such a manner that each child could hear the sound source follow his or her movement in the room. Therefore, there were up to four sounds generated simultaneously, representing the movements of four children. This was achieved through the use of a VBap 1.0.3 object (Pulkki, 1997) by mapping distance from the center point in the room to spread of the virtual sound source and by mapping the participant's angle from the center point to the azimuth angle.

Sound model 1 (S1) was achieved by filtering white noise using the MaxMSP resonance filter `biquad~` object with mode “resonant”. Velocity magnitude of participant's movement in the 2D-plane was mapped to center frequency of the filter (50 to 1100 Hz) and to Q-factor (1.8 to 4.0). Amplitude modulation of the filtered signal was carried out using the `rand~` object, with input parameter 3 Hz. Finally, velocity magnitude was logarithmically scaled to amplitude of the signal, so that no sound was heard when the participant did not move. Sound model 2 (S2) was implemented in a similar manner as S1, with the difference that the resonance filter's center frequency was set to 100–900 Hz and Q-factor range was set to 0.1–0.3. Amplitude modulation was also increased to 18 Hz. The final sound model (S3) was also based on filtering white noise, but was implemented using a band-pass filter (object `biquad~`). Just like for the other two sound models, velocity magnitude was mapped to the center frequency (100–3000 Hz) and Q-factor (0.01–0.6). Amplitude modulation was achieved by triggering peaks using the `curve~` object⁸ which produced a non-linear ramp of length 250 ms, triggered every 50 to 800 ms, depending on velocity. See Figure 1 for the spectral content of 2 min of sound models S1, S2, and S3.

3.1.4. Experimental Procedure

Groups of 3–4 children were studied in each recording session. The participants were wearing hats with attached rigid body markers; trajectories could thus be defined as collections of consecutive points corresponding to the positions of the tracked head while performing a locomotor movement. We assume that head movements carry enough information about the children's expressiveness based on previous findings by Dahl and Friberg (2007) suggesting that expressive movements produced by musicians' head movements are as informative as whole body movements. Each experiment began with a brief introduction by the test leader, explaining to the participants that they were allowed to move freely in the motion capture area of the room (something that the younger participant group failed to do, thereby causing irregular data and problems with loss of

⁸connected to the following message box: “0.85, 0.0 50 0.5 0.0 200 –0.5”

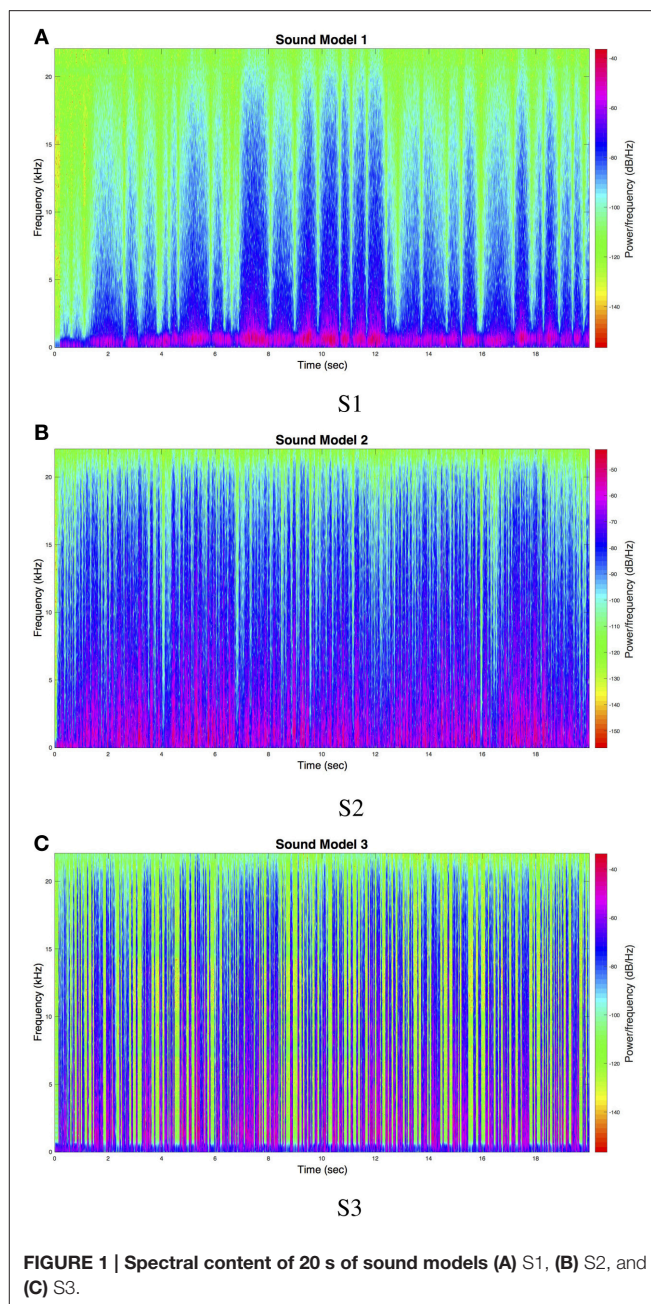


FIGURE 1 | Spectral content of 20 s of sound models (A) S1, (B) S2, and (C) S3.

tracking), and that their movements would produce sounds. Instructions were read from a pre-written manuscript. The instructions were followed by the music condition M1, in which the participants were allowed to move freely to music, but did not trigger any sounds themselves. After M1, a counterbalanced order of the sound models S1–S3 was presented to the participants. For these sound model conditions, participant's rigid body markers were mapped so that movement triggered sounds. Each sound model was presented six times. The entire experiment ended with another music model, namely M2, which was not either mapped to movement of the participants. The music conditions were 60 s long; the sound model

conditions S1–S3 were 36 s long. Each experimental session lasted approximately 13.5 min. Since each group participated in two sessions of the experiment, each sound model was presented 12 times in total, resulting in 12 observations per sound model and participant, respectively.

3.1.5. Analysis of Movement Features

The longitudinal data collected for the three level repeated measures experiment resulted in a data set in which repeated measurements (level 1) were nested within in our unit of analysis, i.e., participants (level 2), which were in turn nested within experiment groups (level 3). We used R (R Core Team, 2014) and the `lme4` package (Bates et al., 2015) to perform a linear mixed model (LMM) analysis of the relationship between movement features and sound model. Linear mixed-effects models are an extension of linear regression models for data that is collected in groups. Numerous studies have demonstrated the advantages of mixed effect models over traditional random-effects ANOVAs (e.g., Baayen et al., 2008; Quené and van den Bergh, 2008). A mixed-effect model consists of fixed (FE) and random effects (RE), where FE are the predictors and RE are associated with experimental units on an individual level, drawn at random from a population.

The standard form of a linear mixed-effect model is defined in Equation (7), where y is the known response variable, X is a fixed-effects design matrix, β is an unknown fixed-effect vector containing the regression coefficients, Z is a random-effects design matrix, b is an unknown random-effects vector and ϵ is the unknown observation error vector.

$$y = X\beta + Zb + \epsilon \quad (7)$$

Our main goal was to determine which predictors that were statistically significant and how changes in the predictors relate to changes in the response variable, not to build a model that could exactly emulate the effect of sonification on participant behavior. Since our research interest is centered around understanding why mean values of the dependent variable vary, we focused mainly on defining random intercept models. We defined a random intercept model for each feature index (FI) according to Equation (9) in which feature magnitude was as function of the fixed effect of sound model. A time variable (observation number 1–12) and session factor (recording before or after lunch) was also added as fixed effects when these were found to be significant, see Equation (9). The model resolved for non-independence by assuming different random intercepts for each participant and group, respectively. More complicated designs were also investigated, however, no random slope models converged. We described model fit by using the marginal and conditional R^2 for mixed-effects models, obtained using the `r.squaredGLMM` function in version 1.10.0 of the `MuMIn` package in R (Nakagawa and Schielzeth, 2013; Johnson, 2014; Barto, 2016). The marginal R^2 describes the proportion of variation explained by fixed effects and the conditional R^2 describes the proportion of variation in the data explained by both fixed and random effects (Nakagawa

and Schielzeth, 2013).

$$FI \sim \text{sound model} + (1|\text{group}) + (1|\text{participant}) + \epsilon \quad (8)$$

$$FI \sim \text{sound model} + \text{session} + \text{observation} + (1|\text{group}) + (1|\text{participant}) + \epsilon \quad (9)$$

3.2. Results

Examples of trajectories performed by a group of children for the three sound models are seen in **Figure 2**. Initial inspection of the data indicated considerable inter-participant variability. Issues with crossover effects (i.e., that one marker was accidentally mistaken for another marker number so that a swap of trajectories occurred for specific participants) and occlusion effects (resulting in gaps in the data) were identified in the initial stage of the data analysis. Since our analysis approach was based on computing higher order derivatives, we decided to remove all observations where crossover effects occurred, so as to reduce the risk of undesired peaks in the computed features. Observations in which tracking was insufficient due to occlusion or contained too few data points (due to the fact that participants moved outside of the trackable area) were also removed.

The recorded movement data was trimmed to 25 s long excerpts per observation, removing the first and last 6 s (original observations were 36 s long + 1 s fade between sound models). Trimming was done in order to include only the middle part of each observation. This was done to ensure that the transitions that contained fading between sound models were not included in the analysis. Moreover, trimming was done to ensure that children had stabilized their movement pattern for the sound model that was currently presented and had been active for at least a time interval of 6 s. One observation was thus defined as a recording segment of 25 s, for one specific participant and sound model. Two-dimensional tracked movement data was thereafter used to calculate the following features for all 25-s excerpts: Energy Index (EI), Smoothness Index (SI) and Directness Index (DI).

Mean values were computed for all recording segments, resulting in 12 observations per participant and sound model (six observations from the experimental session taking place before lunch and six observations from the experimental session after lunch). Data was then normalized to the range of 0 to 1. After removal of observations with erroneous tracking, we obtained a total of 302 observations (total number before removal was 396). A summary of the computed metrics per sound model and participant can be seen in **Figure 3**.

3.2.1. Energy Index

Descriptive statistics per sound model is seen in **Table 1** (results obtained when collapsing all observations). A LMM analysis of the relationship between sound model and Energy Index EI was carried out according to the formula specified in Equation (9). Visual inspection of residual plots did not reveal any obvious deviations from normality. P -values were obtained by likelihood ratio tests of the full model with the effect in question against the null model without the effect in question. Sound model significantly affected energy, $\chi^2_{(1)} = 7.593, p = 0.022$. Recording

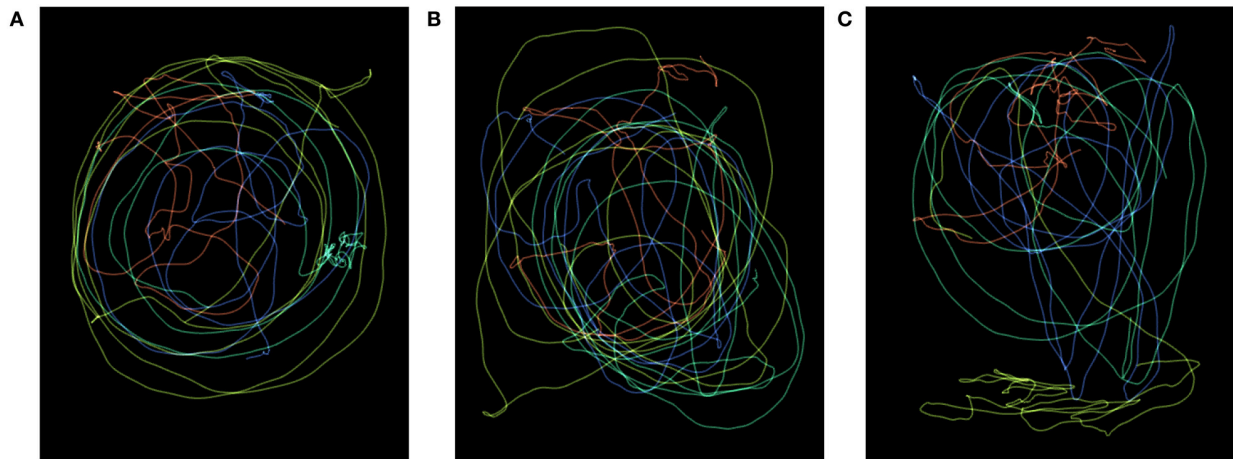


FIGURE 2 | Heat maps. Example of the trajectories performed by the four children in a group, for the three different sound models S1–S3. Each color represents one child, brighter colors corresponds to higher velocity. **(A)** Sound model S1. **(B)** Sound model S2. **(C)** Sound model S3.

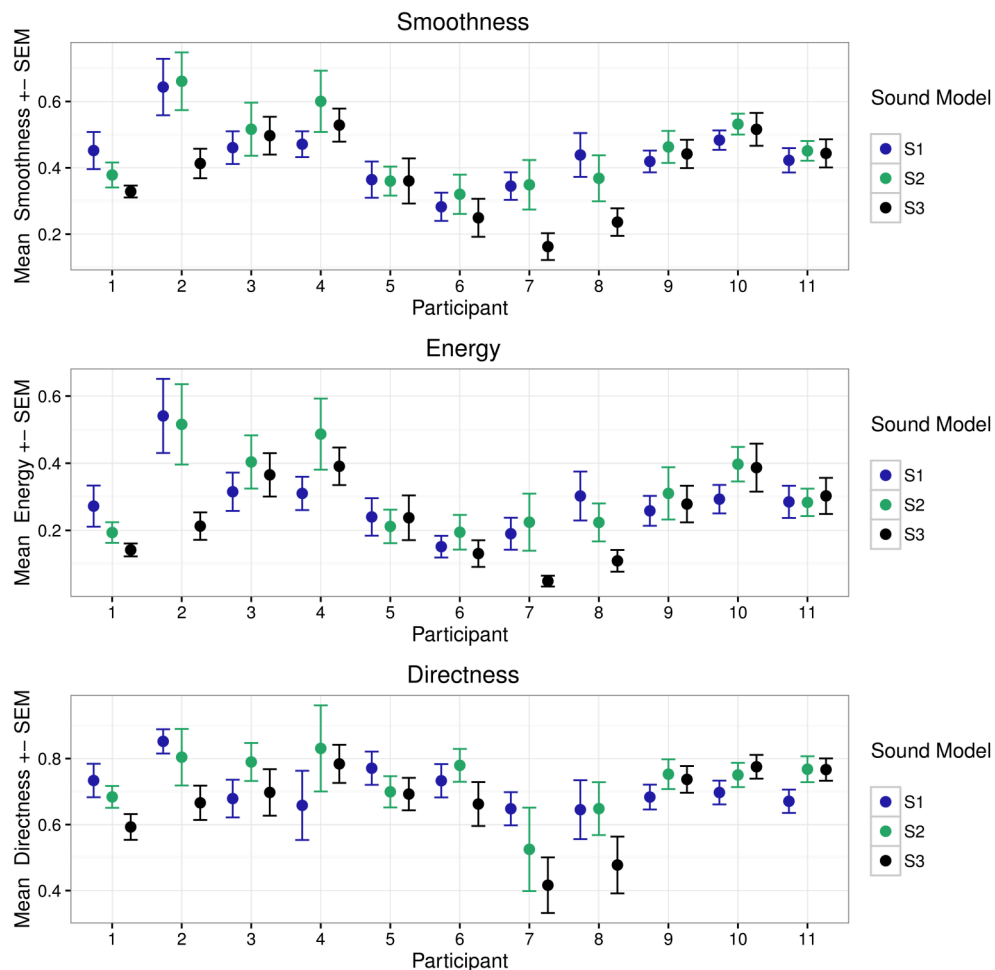


FIGURE 3 | Mean metrics per participant for the three sound models ($n = 12$ observations for each model). P1–P4 belong to group 1, P5–P8 belong to group 2, and P9–P11 belong to group 3.

session and observation number also significantly affected *EI*; $\chi^2_{(2)} = 3.855, p = 0.050$, lowering *EI* by 0.043 ± 0.023 by session, and $\chi^2_{(3)} = 22.989, p = 1.630e - 06$, lowering *EI* by 0.009 ± 0.002 by observation number. Afternoon sessions had generally lower *EI* values. *EI* also decreased for an increasing number of observations. This could possibly be explained by fatigue. Tukey's method for multiple comparisons of means indicated a significant difference between S2 and S3 ($p = 0.033$). Estimate for difference between S2 and S3 was -0.059 ± 0.024 . Although not significant ($p = 0.163$), estimate for difference between S1 and S3 were -0.043 ± 0.023 . Standard deviation described by random effects for participants, groups and residuals were 0.052, 0.054, and 0.168, respectively. The high value for residual standard deviation could well indicate that there might be effects that the model does not account for. Using a simple intercept model as the one defined in Equation (7) and computing pseudo-R-square including only

sound model as fixed factor, sonification could be said to explain about 2.069% of the variability in *EI*. The entire model defined in Equation (9) accounted for a total of 24.284% in *EI* variability. A summary of predicted energy values, involving both fixed and random effects for sound model, is seen in Figure 4A.

3.2.2. Smoothness Index

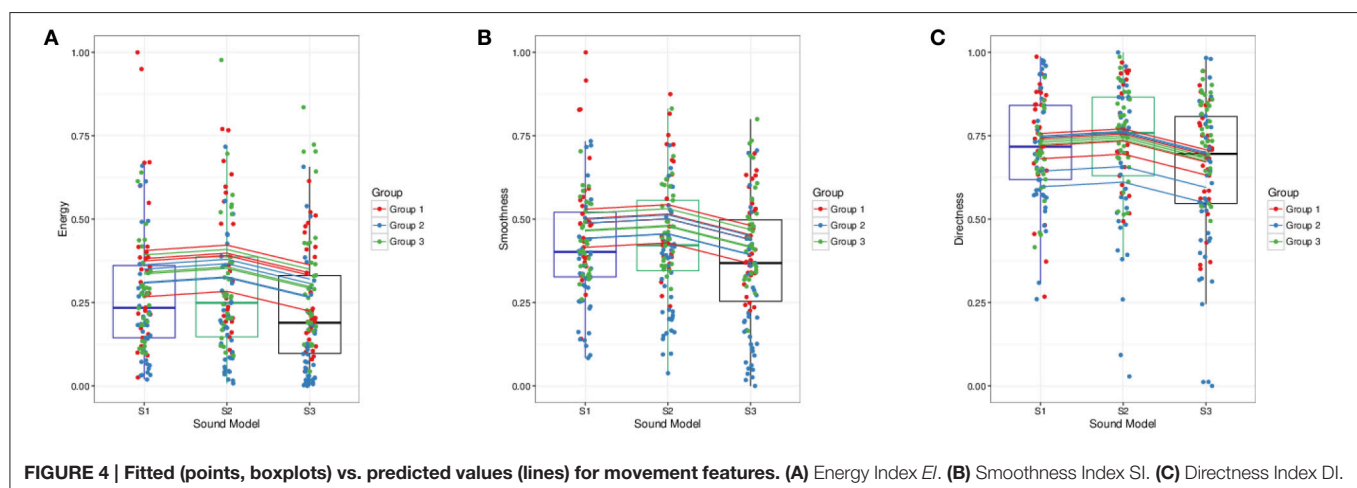
Analysis of the relationship between sound model and Smoothness Index *SI* was carried out according to the method described for Energy Index *EI*. Sound model significantly affected smoothness *SI*, $\chi^2_{(1)} = 10.714, p = 0.005$. There was also a significant effect of recording session, $\chi^2_{(2)} = 4.424, p = 0.035$, and for observation number, $\chi^2_{(3)} = 16.819, p = 4.113e - 05$. Afternoon session had generally lower *SI* values; *SI* was predicted to be lowered by 0.037 ± 0.017 between recording session 1 and 2. An increase in observation number decreased smoothness by 0.007 ± 0.002 . Tukey's method for multiple comparisons of means indicated a significant difference between S2 and S3 ($p = 0.008$). Estimate for difference between S2 and S3 was -0.063 ± 0.021 , i.e., lower smoothness for sound model S3. Although not significant $p = 0.050$, estimate for difference between S1 and S3 was -0.049 ± 0.021 . Standard deviations for the random effect of participant, group and residuals were 0.054, 0.068, and 0.149, respectively. Results obtained from computation of pseudo-R-square indicated that 2.699% of the total variability in smoothness could be described by the fixed factor. If including both session and observation number as fixed factors, as in Equation (9), the model accounts for 29.050% of the total variability. A plot of the predicted values for each participant and group, i.e., the sum of random and fixed effects coefficients for the main explanatory variable “(sound model),” can be seen in Figure 4B. Attempts to model higher level models involving random slopes converged but were not significantly different from the random intercept model defined in Equation (7).

3.2.3. Directness Index

Analysis of the relationship between sound model and Directness Index *DI* was carried out according to the method described

TABLE 1 | Descriptive statistics for Energy Index *EI*, Smoothness Index *SI* and Directness Index *DI*.

	Mean	Median	SD	N	SEM
Energy index					
S1	0.279	0.235	0.191	99	0.019
S2	0.291	0.249	0.199	97	0.020
S3	0.231	0.189	0.188	106	0.018
	Mean	Median	SD	N	SEM
Smoothness index					
S1	0.426	0.402	0.171	99	0.017
S2	0.434	0.422	0.175	97	0.018
S3	0.374	0.368	0.185	106	0.018
	Mean	Median	SD	N	SEM
Directness index					
S1	0.711	0.717	0.161	99	0.0162
S2	0.724	0.759	0.184	97	0.0187
S3	0.662	0.696	0.201	106	0.0196



for the other two indexes. Sound model significantly affected directness, $\chi^2_{(1)} = 7.418, p = 0.025$. No significant effect of observation number or session could be found, so these variables were not included in the model. Tukey's method for multiple comparisons of means indicated a significant difference between S2 and S3 ($p = 0.026$). Estimate for difference between S2 and S3 was -0.0628 ± 0.024 . Although not significant ($p = 0.1061$), estimate for difference between S1 and S3 was -0.0489 ± 0.024 . Standard deviation described by random effects for participants, groups and residuals were 0.0583, 0.026, and 0.172, respectively. Sonification could be said to explain about 2.153% of the variability in *DI*. The model in Equation (9) accounted for a total of 13.983% in directness variability. A summarizing figure of predicted directness values involving both fixed and random effects for sound model can be seen in **Figure 4C**.

3.3. Discussion

Analysis of movement features indicate some significant differences between sound models. However, due to large inter-participant variability, the effect of sound model appears to be rather small. Nevertheless, we can see tendencies toward greater mean and median values for Smoothness Index *SI* for model S1 and S2 than for model S3. The same tendency can be found for both Energy Index *EI* and the Directness Index *DI*; mean and median values are greater for S1 and S2, compared to mean and median values for S3. Assuming that participants were moving in a more continuous manner for S1 and S2, the relatively low mean and median values for S3 might be explained by many sudden interruptions in the trajectory path. As seen in **Figure 3**, participants appear to show similar behavior within groups (P1–P4 belong to group 1, P5–P8 belong to group 2, P9–P11 belong to group 3).

Previous studies have proven evidence that interacting individuals can coordinate their movements through detection of visual movement information (Schmidt et al., 1990) and that visually mediated interpersonal coordination is governed by an entrainment process (Richardson et al., 2007). It is reasonable to expect the movement behaviors of the children to spread within the group, this being a result of either entrainment or conscious and unconscious social interaction. Furthermore, non-spontaneous movements were also introduced via rule-based games or free play, especially during musical condition M2 that was a piece known by the children. None of the above mentioned effects were explicitly measured in this experiment; laying bare the layered subtleties of the children's group play and interpersonal coordination patterns was well beyond the scope of this study.

4. STUDY 2: PERCEPTUAL RATING OF AUDIO AND VIDEO

4.1. Method

The second study focused on investigating if sound models used in Study 1 could communicate certain hypothesized movement qualities. We therefore ran a perceptual test in which sound generated by children in the previous experiment were rated by

listeners along six different perceptual scales. The test was run during the *Festival della Scienza in Genova*, (October, 27th 2015).

4.1.1. Participants

Eight participants took part in the experiment, but only seven of them (5 women) completed the experiment and could therefore be included in the final analysis. The average age of these seven participants was 27.6 years (SD 11.8).

The research presented no risk to harm subjects and involved no procedures for which written consent is normally required outside of the research context. Each subject voluntarily decided to participate in the experiment and the collected data could not be coupled to the specific participant; there was no risk for potential harm resulting from a breach of confidentiality.

4.1.2. Stimuli

Recorded movements and sounds from Study 1 were used to produce the stimuli. Stimuli were presented in random order to the participants and were of three conditions: videos with audio (audio-video), videos without audio (video-only), and audio only (audio-only). The sounds used corresponded to excerpts of sounds generated using S1, S2, S3, and M2. The audio-video stimuli⁹ presented movements in the horizontal plane generated by children when moving to a specific sound model and the corresponding generated sound; each participant produced a trajectory corresponding to the changing position of the head as seen in a two-dimensional plane, parallel to the floor. Videos showed dots of different colors moving on a black background, each dot representing the movements of the head of each child in a group. For the video-only stimuli the audio track was muted, and for the audio-only stimuli the video was removed. In order to provide an idea of the movements showed by the dots, heat maps of the trajectories for each of the three different sound models S1–S3 as performed by the four children in a group are presented in **Figure 2**.

For each of the three conditions (audio-video, video-only, audio-only) there were 12 stimuli, corresponding to 4 sound models \times 3 variations. Participants were thus presented with a total of 36 stimuli. Each stimulus was 20 s long. All excerpts were taken from the first group of participants (4 children) from the morning session in Study 1, in which each sound model was presented six times. We chose to include the recordings corresponding to the first three of these variations in the current Study.

4.1.3. Equipment

Stimuli were presented using an online platform¹⁰ and evaluated using portable tablets. All participants wore headphones¹¹.

4.1.4. Experimental Procedure

The participants were presented with the following instructions on the screen of the tablet:

⁹Examples of stimuli can be found here: <https://kth.box.com/s/818vbk6m6nlkgk4vb4wvt0zt52y5uqy>

¹⁰SurveyGizmo: <http://surveygizmo.com>

¹¹The test can be found at: <http://www.surveygizmo.com/s3/2396910/Evaluation-of-sound-and-video-qualities>

In this test you will watch videos and listen to sounds. You will be asked to rate different properties of each of them by using sliders on the screen. Take all the necessary time, but try to answer as quickly as possible and to use the entire scale of the sliders. There is not right or wrong answer. You can repeat the playback of videos and sounds as many times as you need.

Participants were asked to rate the stimuli along six continuous semantic differential scales describing movement quality (Fluid, Energetic, Impulsive, Fast, Expressive, and Rigid) ranging from *Not at all* to *Very much* as minimum and maximum values, respectively (e.g., not at all fast, very fast)¹². The slider's start position was always placed in the middle of the scale, corresponding to value 50. Numerical values of the sliders were not visible to participants.

The six semantic scales were identified in previous research in which they were used in body motion analysis (for Fluid, Energetic, Impulsive, Fast, and Rigid; see Camurri et al., 2016) and for rating expressiveness in music performance resembling biological motion (for Expressive; see Juslin et al., 2002).

4.2. Results

The duration of the experiment was 44 min, on average (SD = 10). The participants' mean ratings were analyzed using a three-way repeated measures ANOVA, with the factors sound model (4 levels), sound model variation (3 levels), and condition (3 levels). The analysis was done separately for each of the six semantic differential scales. Before running the three-way ANOVA, a Mauchly test was run to verify if the assumption of sphericity had been met for the factors sound model, sound model variation, and condition. When needed, we report corrected degrees of freedom (using Greenhouse-Geisser estimates of sphericity). The analysis for the sound model factor is summarized below and in **Table 2**:

¹²Depending if the stimulus was a video or a sound only, the question to experiment participants was "The sound reminds of a movement that is:" or "The movements of the dots are:", respectively. Each scale was varying between 1 and 100.

Energetic: There was a significant main effect of condition, $F_{(2, 12)} = 13.609, p = 0.001$. Stimuli presented in the audio-only condition were in general rated as more energetic than stimuli with video, i.e., stimuli in both audio-video and video-only conditions. A Bonferroni *post hoc* comparison showed that the mean Energetic rating for the audio-only condition was significantly different (higher) from that of the two other stimuli categories ($p < 0.037$). It can also be observed that stimuli with sounds S1 and M2 were rated as more energetic than the other stimuli.

Expressive: A significant main effect of sound model was observed [$F_{(3, 18)} = 11.913, p < 0.0001$]; stimuli produced using sound model S1 and M2 were rated as the most expressive ones. M2 was rated as more expressive than S1. Sound models S2 and S3 received a mean rating below 50. Sound S3 was rated significantly different from sounds S1 and M2 (Bonferroni *post hoc* comparison, $p < 0.025$). A significant interaction between condition and sound was also observed [$F_{(6, 36)} = 5.941, p < 0.0001$]; stimuli using sound model S1 were rated as more expressive than stimuli produced with S2 and S3 for all conditions; stimuli produced using S3 were rated as the least expressive ones for all conditions. Stimuli generated using M2 were rated as the most expressive ones for the audio-only and audio-video condition.

Fast: No statistically significant effects were found. Nevertheless, stimuli in the condition audio-only were on average rated as more than 60% faster than stimuli in other conditions. Stimuli corresponding to sound S1 were perceived as the slowest, while those corresponding to sound model S2 were rated as the fastest ones (about 20% faster than the other sounds).

Fluid: For this scale a significant effect of factor sound model was found [$F_{(1.606, 9.638)} = 9.277, p = 0.007$, with corrected degrees of freedom and p -value]. Stimuli corresponding to S1 and M2 were rated as about 30% more fluid than stimuli from S2 and S3. Stimuli from all conditions corresponding to sound model S3 were rated as significantly less fluid than other stimuli. Pairwise comparisons between all sounds showed that there was a significant mean difference between ratings for sound model S1 compared to sound models S2 and S3, as well as between

TABLE 2 | Mean ratings and effect of sound model for the six different semantical scales used in Study 2.

	S1		S2		S3		M2	
	Mean	SE	Mean	SE	Mean	SE	Mean	SE
Energetic	55.063	6.177	46.810	6.599	36.841	5.873	63.952	3.630
Expressive	55.063 ^{a**}	6.717	46.810	6.599	36.841 ^{a**,b*}	5.873	63.952 ^{b*}	3.630
Fast	46.905	4.209	59.952	5.325	52.175	6.258	48.921	4.836
Fluid	56.302 ^{c*,d*}	6.197	47.937 ^{c*}	4.262	31.222 ^{d*,e**}	3.981	55.222 ^{e**}	2.217
Impulsive	56.683	3.210	52.254	3.095	56.508	5.791	44.905	4.879
Rigid	33.048	3.435	35.413	4.851	55.952	7.438	33.905	4.028

Significance levels: * $p \leq 0.05$, ** $p \leq 0.01$.

^asignificant difference in means between S1 and S3.

^bsignificant difference in means between S3 and M2.

^csignificant difference in means between S1 and S2.

^dsignificant difference in means between S1 and S3.

^esignificant difference in means between S3 and M2.

sound models S3 and M2 (Bonferroni *post hoc* comparison, $p < 0.05$). A significant interaction between factors condition and sound model was also found [$F_{(1,606,9,638)} = 9.277, p = 0.007$, with corrected degrees of freedom and p -value]. Audio-only stimuli corresponding to S2 and M2 were rated about 3 times more fluid than the other stimuli; for the video condition this difference was not present for model S2, while stimuli corresponding to sound S3 were rated as the least fluid ones in all conditions.

Impulsive: A significant effect of condition was found [$F_{(2, 12)} = 6.152, p = 0.014$]. Stimuli in the audio-only condition were in general rated as more impulsive than stimuli in the audio-video condition (Bonferroni *post hoc* comparison, $p < 0.028$). A significant interaction between factors sound and sound variation was also found [$F_{(3,36)} = 3.805, p = 0.005$]. Sounds S2 and S3 were rated as the more impulsive ones in the audio-only condition. Stimuli of sound M2 were rated as the least impulsive ones.

Rigid: No statistically significant effect of main factors was found. However, it was observed that sound S3 was rated as the most rigid one, about 60% more rigid than the other sounds. A significant interaction between condition and sound model was also observed [$F_{(6,36)} = 10.42, p < 0.0001$]. S3 was rated as the most rigid in all categories. Stimuli presented with sound S2 were rated as more rigid when presented without video, and less rigid when presented in other conditions. Stimuli including S1 and M2 received low mean ratings for all conditions and could thereby be considered to be perceived as non-rigid.

4.3. Discussion

To summarize, a significant main effect of sound model was observed for the scales Expressive and Fluid, in which sound models S1 and M2 were rated as more expressive and fluid than the other sound models. Sound model S3 was rated as more rigid and fast than other sound models, although this difference was not significant. A significant effect of condition was observed for scales Energetic and Impulsive. The interaction effect between condition and sound model was also observed to be significant for scales Expressive, Fluid and Rigid. These results confirm our initial hypothesis that sound model S1 would communicate the sensation of being more fluid, smoother (and possibly also slower and less rigid) while sound model S3 would be perceived as less fluid (and possibly also faster and more rigid).

5. STUDY 3: PERCEPTUAL RATING OF SOUND VISUALIZATIONS

5.1. Method

We hypothesize that the properties of the body motion used by the children for generating sounds S1–S3 can be found also in abstract representations of sound, i.e., sound visualizations in the form of drawings. More specifically, our hypothesis is that there is a consistent mapping of body motion qualities from one modality (sound) to another one (sound visualizations). To investigate this, we ran a three alternative forced-choice experiment (3AFC)

designed to see if participants could correctly match recordings of one sound model to an abstract visual representation (i.e., a drawing) of the same sound model.

5.1.1. Participants

146 students (68 women) from the Media Technology programme at KTH took part in the experiment. Their average age was 22.4 years (SD = 2.7).

As for the previous online experiment (Study 2), the research presented no risk to harm subjects and involved no procedures for which written consent is normally required outside of the research context. Each subject voluntarily decided to participate in the online study and there was no risk for potential harm resulting from a breach of confidentiality.

5.1.2. Stimuli

The 11 children who had participated in Study 1 (see Section 3 and 3.1.1 for ethics considerations) took part in a follow-up study that was set up as a drawing exercise. The children listened to excerpts of the two classical music stimuli (M1–M2) and the sonification sounds (S1–S3) that they had produced in the motion capture experiment. The excerpts were 2 min long. The children were asked to freely draw whatever they wanted while listening to each of the 2-min long five audio stimuli (S1–S3 and M1–M2). We consider these drawings to be abstract representations of the presented sounds. The idea of using drawings to depict sounds was inspired by previous work by Merer and colleagues (2008; 2013).

A selection of sound visualizations in the form of drawings from the drawing exercise described above was used as stimuli in the 3AFC experiment. We selected drawings that included abstract representations of the sounds from 4 children. This selection was done in order to avoid symbolic representations of the sounds (such as e.g., plants, birds or people), which could bias the perceptual ratings. Three drawings per child were used as stimuli in the 3AFC experiment: each drawing corresponded to each of sound models S1, S2, and S3¹³. Each drawing was presented with the same recorded sounds that had been presented in the drawing exercise. The drawings were processed to be black and white to enable the participants to focus simply on the patterns and trajectories in the drawings, not on color properties (see Figure 5).

5.1.3. Equipment

Stimuli were presented using the same online platform as in Study 2¹⁴. A link to the experiment¹⁵ was sent via email to the participants, who could use their preferred device (computer or portable device) to participate in the experiment.

¹³Drawings of the classical sounds were not included as stimuli in the experiment, since the research hypothesis of Study 3 only addressed aspects of the sonification models, and not music stimuli in general.

¹⁴SurveyGizmo: <http://surveygizmo.com>

¹⁵The test can be found and performed at the following link: <http://www.surveygizmo.com/s3/2607938/e9087ebd3a82>

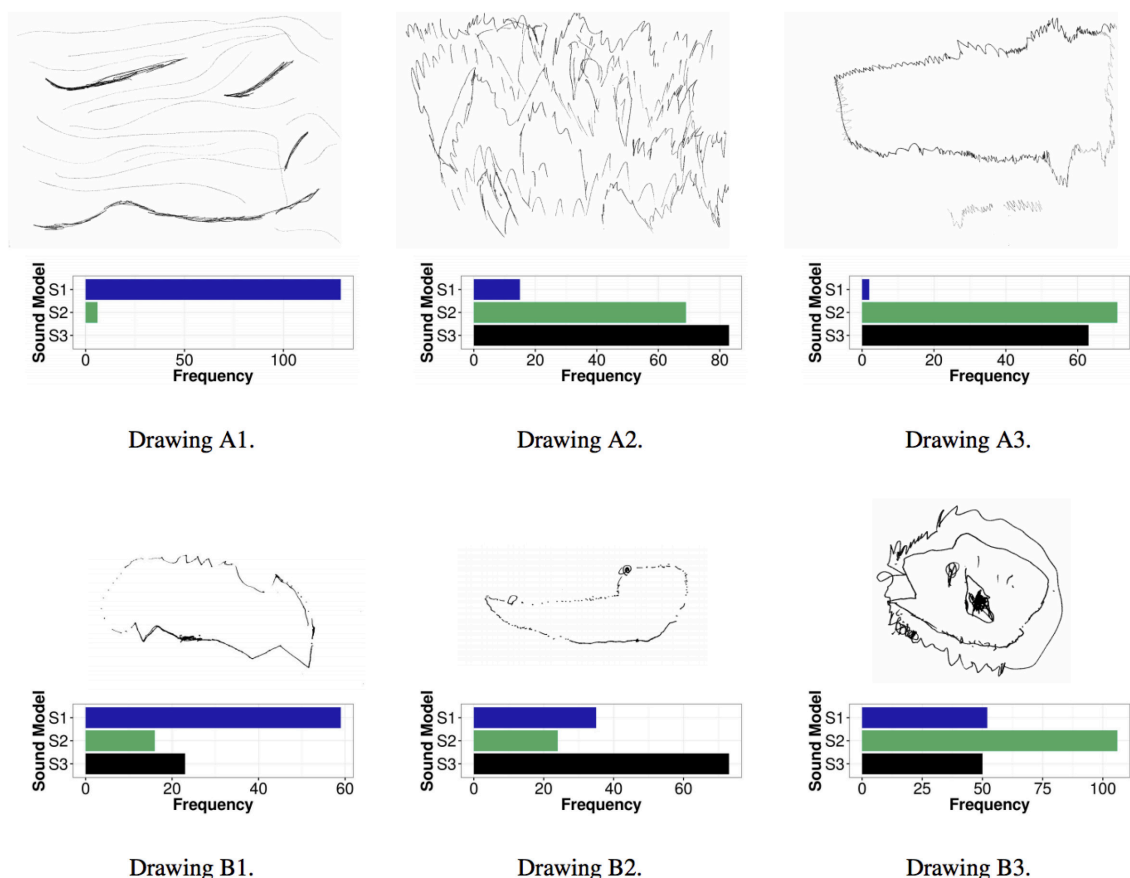


FIGURE 5 | Drawings by one 6-year girl (child A, top row) and by a 6-year boy (child B, bottom row) used in the experiment in Study 3; left to right drawings made while listening to sounds produced using model S1, S2, and S3, respectively.

5.2. Experimental Procedure

The following instructions were given to the participants:

In this test you will be asked to connect sounds to drawings. Take all the necessary time, but try to answer as quickly as possible. There is not right or wrong answer. You can repeat the playback of sounds as many times as you need. You are allowed to answer to the questions while the sound is still playing.

The total number of stimuli consisted 4 sets of drawings (from 4 different participants) \times 3 sound models, giving a total of 12 stimuli. Stimuli were presented in a randomized order per set of drawings. Participants were asked to make a three-alternative forced choice (3AFC) between three drawings and the presented sound.

5.3. Results

Based on findings from Study 1 in which it was concluded that results from the youngest children should be excluded from the analysis since these participants did not follow instructions correctly and also not fully understood the experimental task (see Section 3.1.1), drawings produced by the youngest children were excluded from the analysis. This decision was done in order to follow the same methodology as the one used in Study 1. Analysis of the obtained results were thus done on the answers obtained for the 2 drawings that had been produced by the oldest children

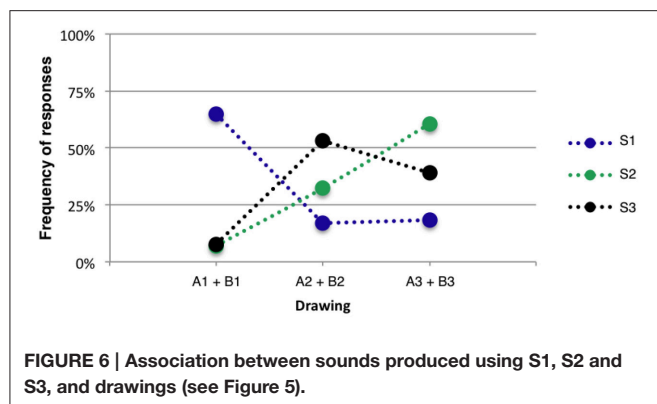
(one girl and one boy; referred to as child A and child B in Figure 5).

We ran a chi-square test to analyze the association between the two variables sound model (S1–S3) and drawing [$\chi^2_{(df=10, N=876)} = 436.514, p = 0.000$]. The results indicated a significant association between the two variables (expected counts were greater than 5), thus implying that certain sound models were associated to certain drawings. In particular, sound model S1 was clearly associated to drawings of S1, while S2 and S3 were mostly associated to drawings of either one of these two sound models (see Figure 5 for more details).

Analysis of response frequency when collapsing all results per drawing class (i.e., which sound model the drawing was actually depicting) showed that 64% of the participants associated sound model S1 to the corresponding visual representation of sound model S1. Only 8% of the participants associated sound model S1 to drawings depicting sound model S2 or S3 (see Figure 6).

5.4. Discussion

As mentioned by Glette et al. (2010), participants associations to sounds are very subjective and tracings of sound can therefore vary a lot between participants. Nevertheless, results from the sound visualization experiment indicated that participants rather



easily identified drawings portraying sound model S1. Drawings of sound model S2 and S3 did also match to the correct sound model for child A. For child B we can see an opposite behavior in which drawing of sound model S2 was matched to sound model S3, and vice versa. These results confirmed our hypothesis that qualities of movement present in the sound recordings could be transferred into sound visualizations produced by children and that listeners could subsequently recognize these qualities.

6. GENERAL DISCUSSION

Analysis of movement features in Study 1 indicated significant effect of sound model. However, due to large inter-participant variability, the effect of sound model appeared to be rather small. In general, the effect of group belonging appears to also be important in this context, as well as aspects of fatigue (observed in terms of significant effect of observation number and session number). The children showed different behavior throughout the experimental sessions and moved in a manner that was not very consistent; their movement patterns appeared to be more guided by the social interaction with other children than the overall features of the sounds. Nevertheless, we can see some tendencies toward greater mean and median values of smoothness and directness for the sound models S1 and S2 than for sound model S3. This might indicate that there are aspects related to sound model which would be interesting to explore further in the context of spontaneous movement induced by interactive sonification. Considering the open structure of the experiment (the children were allowed to move freely and interact with each other in groups), it is likely that a more controlled experiment would provide clearer results with higher statistical power. The fact that the children were very young and behaved accordingly was of course an aspect that affected the results (as previously mentioned, some of the data had to be excluded from the analysis). We propose follow-up studies in which the same sonification models are evaluated in a more controlled setting to fully be able to evaluate the effects of sonification model on induced movement.

Findings from the perceptual rating experiment (Study 2) indicate a significant effect of sound model on the perception of expressiveness and fluidity. More precisely, sound model S1 was found to communicate the sensation of being more fluid when compared to sound model S3. Although not significant, S3

was rated as 60% more rigid and fast than other sound models. One could suggest that certain properties of sound model S1 results in the fact that sounds produced using this model are perceived as more fluid and slow than sounds produced using sound model S3. Interestingly, we could also detect significant interactions between sound model and condition (audio-only, video-only or audio-video) for the expressiveness-, fluidity- and rigidity scales. These results support the hypothesis that different sound models can, by themselves, be perceived differently, but also that perception of movement qualities is indeed a multimodal phenomenon. Interestingly, the effect of condition was significant for energy- and impulsivity scales: when stimuli were only auditory it was perceived as more energetic and more impulsive than when stimuli also included a video visualization counterpart. This confirms the ability of sound to communicate high-level qualities of movement.

Although the experimental methodology of Study 3 could have been simplified, for example by using simple sound visualizations containing caricatures similar to the ones in the *Bouba-Kiki* experiments by Kohler (1929, 1947) instead of drawings, the ability to communicate high-level qualitative features of movement using only sound as a medium could be confirmed in the study. The qualities of a movement present in audio recordings were recognized in sound visualizations produced by children. Drawings portraying sound model S1 were rather easily identified as being a portrayal of the actual sound model S1. This supports our hypothesis that certain qualities of movement present in sound recordings can actually be translated into sound visualizations (and that these sound visualizations subsequently can be recognized by another independent group of listeners). Similarly to what Merer et al. (2013) suggested, i.e., that drawings are a relevant means of describing motion in an intuitive way, we can conclude that drawings can be successfully used as a tool for describing movement features which are present in a sound through meaningful sonification of movement properties.

To conclude, the three studies presented in this paper suggest that sound models can be designed and controlled so that: (1) sound might have an effect on bodily movement characteristics; (2) different sounds can be associated with different levels of motion qualities (e.g., fluid and expressive); (3) sound-only stimuli can evoke stronger perceived properties of movement (e.g., energetic, impulsive) compared to video stimuli; (4) sounds generated by body motion can be represented and associated with sound visualizations (drawings) in a meaningful way. The results obtained support the existence of a cross-modal mapping of body motion qualities from bodily movement to sounds and the potential of using interactive sonification to communicate high-level features of human movement data. Sound can be translated and understood from bodily motion, conveyed through sound visualizations in the form of drawings, and translated back from sound visualizations to sound.

AUTHOR CONTRIBUTIONS

RB: supervised the project; EF, RB, and LE: designed and performed the experiments; LE developed the software used for communication; EF developed sound models and analyzed

collected data from study 1 as well as edited the paper; RB analyzed data from study 2 and 3; PA developed analytical tools using the *EyesWeb* software.

FUNDING

This research has received funding from the European Unions Horizon 2020 research and innovation programme under grant

REFERENCES

- Baayen, R. H., Davidson, D. J., and Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *J. Mem. Lang.* 59, 390–412. doi: 10.1016/j.jml.2007.12.005
- Barto, K. (2016). *MuMIn: Multi-Model Inference*. R package version 1.15.6. Available online at: <https://cran.r-project.org/web/packages/MuMIn/index.html>
- Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* 67, 1–48. doi: 10.18637/jss.v067.i01
- Blatter, A. (2007). *Revisiting Music Theory: A Guide to the Practice*. New York, NY; London: Taylor & Francis.
- Buhmann, J., Desmet, F., Moens, B., Van Dyck, E., and Leman, M. (2016). Spontaneous velocity effect of musical expression on self-paced walking. *PLoS ONE* 11:e0154414. doi: 10.1371/journal.pone.0154414
- Burger, B., Thompson, M. R., Luck, G., Saarikallio, S., and Toiviainen, P. (2013). Influences of rhythm- and timbre-related musical features on characteristics of music-induced movement. *Front. Psychol.* 4:183. doi: 10.3389/fpsyg.2013.00183
- Burger, B., Thompson, M. R., Luck, G., Saarikallio, S. H., and Toiviainen, P. (2014). Hunting for the beat in the body: on period and phase locking in music-induced movement. *Front. Hum. Neurosci.* 8:903. doi: 10.3389/fnhum.2014.00903
- Camurri, A., Castellano, G., Ricchetti, M., and Volpe, G. (2006). *Subject Interfaces: Measuring Bodily Activation During an Emotional Experience of Music*. Berlin; Heidelberg: Springer.
- Camurri, A., Mazzarino, B., Ricchetti, M., Timmers, R., and Volpe, G. (2003). "Multimodal analysis of expressive gesture in music and dance performances," in *Gesture-based Communication in Human-computer Interaction*, eds A. Camurri and G. Volpe (Genova: Springer), 20–39.
- Camurri, A., Trocça, R., and Volpe, G. (2002). "Interactive systems design: a kansei-based approach," in *Proceedings of the 2002 Conference on New Interfaces for Musical Expression, Proceedings of the International Conference on New Interfaces for Musical Expression (NIME02)* (Dublin: National University), 30–37.
- Camurri, A., Volpe, G., Piana, S., Mancini, M., Niewiadomski, R., Ferrari, N., et al. (2016). "The dancer in the eye: towards a multi-layered computational framework of qualities in movement," in *Proceedings of the 3rd International Symposium on Movement and Computing* (Thessaloniki: ACM), 6. doi: 10.1145/2948910.2948927
- Caramiaux, B., Bevilacqua, F., Bianco, T., Schnell, N., Houix, O., and Susini, P. (2014). The role of sound source perception in gestural sound description. *ACM Trans. Appl. Percept.* 11, 1:1–1:19. doi: 10.1145/2422105.2422106
- Casciato, C., Jensenius, A. R., and Wanderley, M. M. (2005). "Studying free dance movement to music," in *Proceedings of ESCOM 2005 Performance Matters! Conference* (Porto).
- 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
- 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
- De Meijer, M. (1989). The contribution of general features of body movement to the attribution of emotions. *J. Nonverb. Behav.* 13, 247–268. doi: 10.1007/BF00990296
- Dubus, G., and Bresin, R. (2013). A systematic review of mapping strategies for the sonification of physical quantities. *PLoS ONE* 8:e82491. doi: 10.1371/journal.pone.0082491
- Dubus, G., and Bresin, R. (2015). Exploration and evaluation of a system for interactive sonification of elite rowing. *Sports Eng.* 18, 29–41. doi: 10.1007/s12283-014-0164-0
- Flash, T., and Hogan, N. (1985). The coordination of arm movements: an experimentally confirmed mathematical model. *J. Neurosci.* 5, 1688–1703.
- Gaver, W. W. (1993). What in the world do we hear? an ecological approach to auditory event perception. *Ecol. Psychol.* 5, 1–29. doi: 10.1207/s15326969eco0501_1
- Glette, K., Jensenius, A. R., and Godøy, R. I. (2010). "Extracting action-sound features from a sound-tracing study," in *Proceedings of the Second Norwegian Artificial Intelligence Symposium (NAIS)*, eds S. Yildirim and A. Kofod-Petersen (Gjøvik: Norsk forening for kunstig intelligens og Tapir Akademisk Forlag), 63–66.
- Glowinski, D., Dael, N., Camurri, A., Volpe, G., Mortillaro, M., and Scherer, K. (2011). Toward a minimal representation of affective gestures. *IEEE Trans. Affect. Comput.* 2, 106–118. doi: 10.1109/T-AFCC.2011.7
- Godøy, R. I., Haga, E., and Jensenius, A. R. (2005). "Playing 'air instruments': mimicry of sound-producing gestures by novices and experts," in *Gesture in Human-computer Interaction and Simulation*, eds S. Gibet, N. Courty, and J. Kamp (Heidelberg: Springer), 256–267.
- Godøy, R. I., Haga, E., and Jensenius, A. R. (2006). "Gesture in human-computer interaction and simulation: 6th international gesture workshop," in *Proceedings of the COST287 - ConGAS 2nd International Symposium on Gesture Interfaces for Multimedia Systems*, ed K. Ng (Île de Berder), 27–33.
- Godøy, R. I., and Jensenius, A. R. (2009). "Body movement in music information retrieval," in *Proceedings of the 10th International Society for Music Information Retrieval Conference (ISMIR)* (Kobe), 45–50.
- Godøy, R. I., Jensenius, A. R., and Nymoen, K. (2010). Chunking in music by coarticulation. *Acta Acust. United Acust.* 96, 690–700. doi: 10.3813/AAA.918323
- Haga, E. (2008). *Correspondences between Music and Body Movement*. Ph.D. thesis, Department of Musicology, University of Oslo.
- Hansen, K. F., Dravins, C., and Bresin, R. (2012). Active listening and expressive communication for children with hearing loss using getatable environments for creativity. *J. New Music Res.* 41, 365–375. doi: 10.1080/09298215.2012.739626
- Janata, P., and Grafton, S. T. (2003). Swinging in the brain: shared neural substrates for behaviors related to sequencing and music. *Nat. Neurosci.* 6, 682–687. doi: 10.1038/nn1081
- Johnson, P. C. (2014). Extension of nakagawa and schielzeth's r2glmm to random slopes models. *Methods Ecol. Evol.* 5, 944–946. doi: 10.1111/2041-210X.12225
- Juslin, P. N., Friberg, A., and Bresin, R. (2002). Toward a computational model of expression in music performance: the germ model. *Music. Sci.* 5(1 Suppl):63–122. doi: 10.1177/10298649020050S104
- Källblad, A., Friberg, A., Svensson, K., and Edelman, E. S. (2008). "Hoppsa universum-an interactive dance installation for children," in *Proceedings of the International Conference on New Interfaces for Musical Expression (NIME08)* (Genova), 128–133.
- Keller, P. E., and Rieger, M. (2009). Special issue – musical movement and synchronization. *Music Percept.* 26, 397–400. doi: 10.1525/mp.2009.26.5.397
- Kohler, W. (1929). *Gestalt Psychology*. New York, NY: Liveright Publishing.
- Kohler, W. (1947). *Gestalt Psychology*. New York, NY: Liveright Publishing.

ACKNOWLEDGMENTS

The authors would like to thank the children and teachers from kindergarten Sture who participated in Study 1 and 3.

- Lakoff, G., and Johnson, M. (1980). *Metaphors We Live By*. Chicago, IL: University of Chicago Press.
- Lakoff, G., and Johnson, M. (1999). *Philosophy in the Flesh: The Embodied Mind and Its Challenge to Western Thought*. New York, NY: Basic Books.
- Leman, M. (2007). *Embodied Music Cognition and Mediation Technology*. Cambridge, MA: MIT Press.
- Merer, A., Aramaki, M., Ystad, S., and Kronland-Martinet, R. (2013). Perceptual characterization of motion evoked by sounds for synthesis control purposes. *ACM Trans. Appl. Percept.* 10, 1:1–1:24. doi: 10.1145/2536811
- Merer, A., Ystad, S., Kronland-Martinet, R., and Aramaki, M. (2008). “Semiotics of sounds evoking motions: Categorization and acoustic features,” in *Computer Music Modeling and Retrieval. Sense of Sounds: 4th International Symposium, CMMR 2007, Copenhagen, Denmark, August 27–31, 2007. Revised Papers* (Berlin, Heidelberg: Springer), 139–158. doi: 10.1007/978-3-540-85035-9_9
- Nakagawa, S., and Schielzeth, H. (2013). A general and simple method for obtaining R² from generalized linear mixed-effects models. *Methods Ecol. Evol.* 4, 133–142. doi: 10.1111/j.2041-210x.2012.00261.x
- Nymoen, K., Glette, K., Skogstad, S. I. A., Torresen, J., and Jensenius, A. R. (2010). “Searching for cross-individual relationships between sound and movement features using an SVM classifier,” in *Proceedings of the International Conference on New Interfaces for Musical Expression (NIME10)* (Sydney, NSW), 259–262.
- Piana, S., Staglianò, A., Camurri, A., and Odone, F. (2013). “International workshop on intelligent digital games for empowerment and inclusion (IDGEI 2013),” in *IDGEI International Workshop* (Chania).
- Pulkki, V. (1997). Virtual sound source positioning using vector base amplitude panning. *J. Audio Eng. Soc.* 45, 456–466.
- Quené, H., and van den Bergh, H. (2008). Examples of mixed-effects modeling with crossed random effects and with binomial data. *J. Mem. Lang.* 59, 413–425. doi: 10.1016/j.jml.2008.02.002
- R Core Team (2014). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.
- Richardson, M. J., Marsh, K. L., Isenhower, R. W., Goodman, J. R., and Schmidt, R. C. (2007). Rocking together: dynamics of intentional and unintentional interpersonal coordination. *Hum. Move. Sci.* 26, 867–891. doi: 10.1016/j.humov.2007.07.002
- Roddy, S., and Furlong, D. (2014). Embodied aesthetics in auditory display. *Organ. Sound* 19, 70–77. doi: 10.1017/S1355771813000423
- Schmidt, R. C., Carello, C., and Turvey, M. T. (1990). Phase transitions and critical fluctuations in the visual coordination of rhythmic movements between people. *J. Exp. Psychol. Hum. Percept. Perform.* 16:227. doi: 10.1037/0096-1523.16.2.227
- Temmerman, N. (2000). An investigation of the music activity preferences of pre-school children. *Br. J. Music Educ.* 17, 51–60. doi: 10.1017/S0265051700000140
- Thiebaut, J.-B., Healey, P. G., Kinns, N. B., and Mary, Q. (2008). “Drawing electroacoustic music,” in *Proceedings of International Computer Music Conference (ICMC2008)* (Belfast), Vol. 8.
- Todorov, E., and Jordan, M. I. (1998). Smoothness maximization along a predefined path accurately predicts the speed profiles of complex arm movements. *J. Neurophysiol.* 80, 696–714.
- Toivainen, P., Luck, G., and Thompson, M. (2009). Embodied metre: hierarchical eigenmodes in spontaneous movement to music. *Cogn. Process.* 10, 325–327. doi: 10.1007/s10339-009-0304-9
- Västfjäll, D. (2002). Emotion induction through music: a review of the musical mood induction procedure. *Music. Sci.* 5, 173–211. doi: 10.1177/10298649020050S107
- Viviani, P., and Terzuolo, C. (1982). Trajectory determines movement dynamics. *Neuroscience* 7, 431–437. doi: 10.1016/0306-4522(82)90277-9
- Zentner, M., and Eerola, T. (2010). Rhythmic engagement with music in infancy. *Proc. Natl. Acad. Sci. U.S.A.* 107, 5768–5773. doi: 10.1073/pnas.1000121107

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.

Copyright © 2016 Frid, Bresin, Alborno and Elblaus. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



3Mo: A Model for Music-Based Biofeedback

Pieter-Jan Maes*, Jeska Buhmann and Marc Leman

Department of Art, Music and Theatre Sciences, Institute for Psychoacoustics and Electronic Music, Ghent University, Ghent, Belgium

OPEN ACCESS

Edited by:

David Rosenboom,
California Institute of the Arts, USA

Reviewed by:

Michael Thaut,
Colorado State University, USA
Peter Brunner,
Albany Medical College, USA

*Correspondence:

Pieter-Jan Maes
pieterjan.maes@ugent.be

Specialty section:

This article was submitted to
Neuroprosthetics,
a section of the journal
Frontiers in Neuroscience

Received: 24 March 2016

Accepted: 15 November 2016

Published: 02 December 2016

Citation:

Maes P-J, Buhmann J and Leman M
(2016) 3Mo: A Model for Music-Based
Biofeedback. *Front. Neurosci.* 10:548.
doi: 10.3389/fnins.2016.00548

In the domain of sports and motor rehabilitation, it is of major importance to regulate and control physiological processes and physical motion in most optimal ways. For that purpose, real-time auditory feedback of physiological and physical information based on sound signals, often termed “sonification,” has been proven particularly useful. However, the use of music in biofeedback systems has been much less explored. In the current article, we assert that the use of music, and musical principles, can have a major added value, on top of mere sound signals, to the benefit of psychological and physical optimization of sports and motor rehabilitation tasks. In this article, we present the 3Mo model to describe three main functions of music that contribute to these benefits. These functions relate the power of music to Motivate, and to Monitor and Modify physiological and physical processes. The model brings together concepts and theories related to human sensorimotor interaction with music, and specifies the underlying psychological and physiological principles. This 3Mo model is intended to provide a conceptual framework that guides future research on musical biofeedback systems in the domain of sports and motor rehabilitation.

Keywords: sonification, auditory biofeedback, music interaction, reinforcement learning, predictive processing

INTRODUCTION

In this article, we consider the potential of music as feedback system to support and influence the psychological and physical demands inherent to sports and motor rehabilitation tasks. Musical biofeedback may be considered a particular case of auditory biofeedback, or “sonification,” Sonification is commonly defined as the transfer of data, and data relationships, into non-speech audio for the purpose of communication and interpretation (Kramer et al., 1999; Hermann et al., 2011). Starting in the 1970s (Zaichkowsky, 1982), the use of sonification, or auditory feedback, in the domain of sports and motor rehabilitation has substantially progressed over the past 20 years (Huang et al., 2006; Dubus and Bresin, 2013; Giggins et al., 2013; Sigrist et al., 2013; Kos et al., 2015). Typically, sonification is used to enhance self-awareness of physiological processes and physical motion in order to regulate and control these in most optimal ways. Thereby, auditory biofeedback systems commonly map physiological and physical quantities to psychoacoustic (sound) parameters, such as loudness, pitch, timbre, and rhythm (Hermann and Hunt, 2005; Dubus and Bresin, 2013). Studies that explicitly use music as auditory biofeedback however are relatively scarce (Bergstrom et al., 2014; Moens et al., 2014; Van Dyck et al., 2015).

In the current article, we put forward the hypothesis that music is highly convenient as real-time feedback of physiological processes (cardiovascular, respiratory, electro-dermal, etc.), motor kinematic and kinetic processes, and performance parameter output (speed, force, height, etc.).

Used as biofeedback system, we assert that music can have a major added value, on top of mere sound, to the benefit of psychological and physical support and optimization of sports and motor rehabilitation tasks. The sports and motor rehabilitation tasks relate mainly to motor exercise, learning, relearning, and actual performance (including warming-up and cooling down). Argumentation in support of our hypothesis is structured according to three main functions of music and musical biofeedback: the power of music to motivate physical activity (i.e., motivation), the ability of musical biofeedback to monitor physiological and motor processes (i.e., monitoring), and the potential to use music to modify (i.e., optimize) these processes (i.e., modification). These three core functions—motivation, monitoring, and modification—outline the three pillars of our model, hence the name “3Mo model” of which a schematic overview is presented in **Figure 1**.

The first function concerns the “power” of music to motivate people; music stimulates people to get physically active, it induces emotions and moods, and it modulates attention and feelings of pain and exertion through “deep listening” (“trancing”). A second function is a more common one in the field of sonification research and relates to monitoring. The real-time monitoring of physiological processes, motor processes, and performance output parameters may sharpen self-awareness and drive self-regulation. In this section, we demonstrate that music is particularly relevant for providing

real-time biofeedback of multiple, concurrent (i.e., multilayered) physiological and physical processes. A third function pertains to the possibility of music to reliably modify physiological processes and motor behavior toward specific goals through reinforcement learning. This function relies on principles related to brainstem responses and sensorimotor predictive processing.

In support of each function, we collect concepts, theories, and empirical evidence. In that respect, the article introduces a novel model that brings together already existing theories. On top of that, some novel ideas are included that have received only limited evidence so far in the context of musical sonification for sports and motor rehabilitation purposes. This relates to ideas of multilayered musical biofeedback (see Monitoring by Musical Biofeedback: “Monitoring”) and the use of prediction and musical reward principles to endow motivational qualities (see Motivation by Musical Biofeedback: “Motivation”) and modify physiological and motor processes (see Modification by Musical Biofeedback: “Modification”).

Our approach is based on the idea that interaction with music is empowering (Leman, 2016). It gives music a central role in the development of expressive interactive machines that work with biofeedback. The concepts and theories used have strong links with musicological research, touching upon music performance, music emotions, music analysis, and

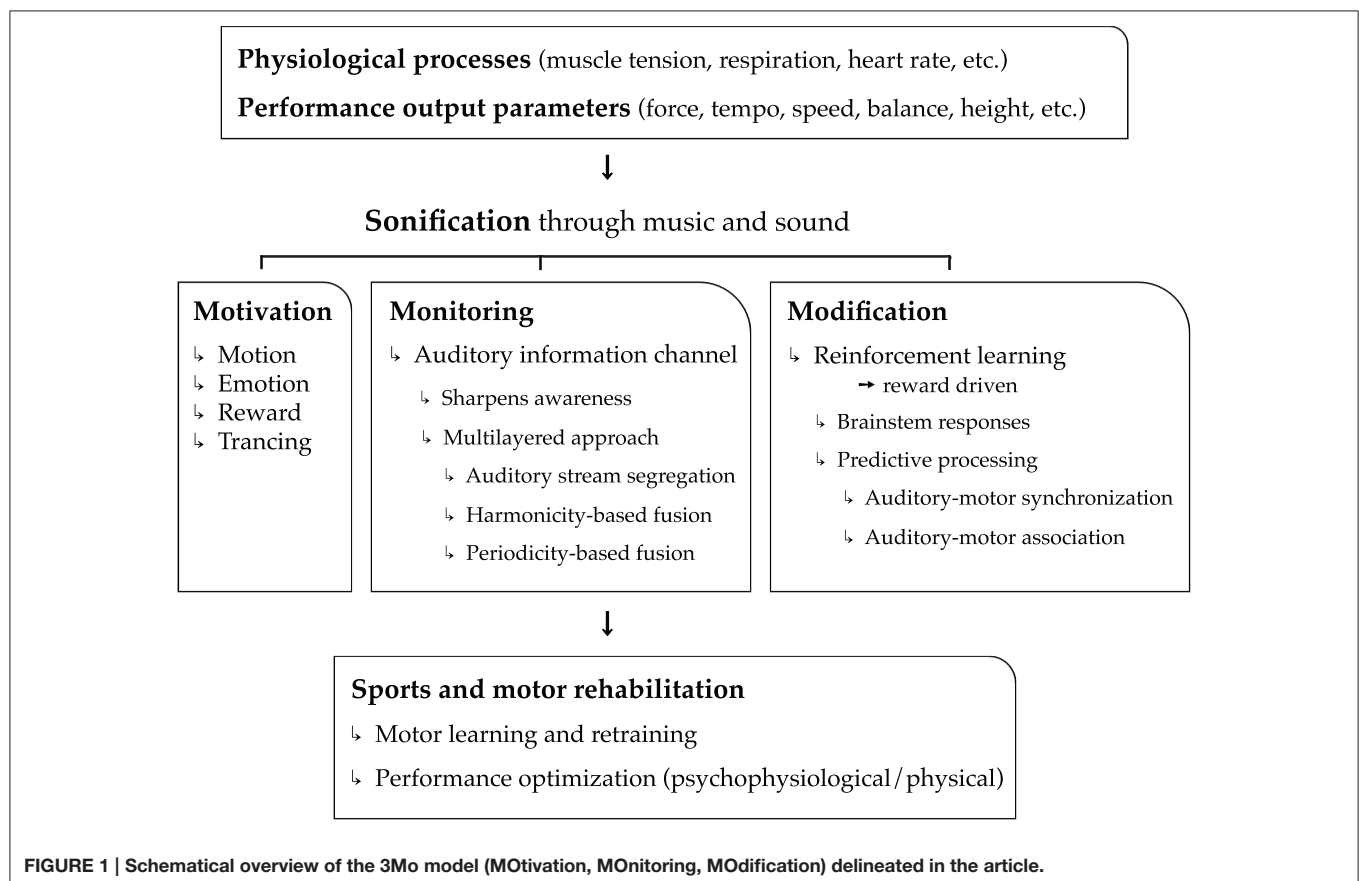


FIGURE 1 | Schematic overview of the 3Mo model (MOTivation, MOnitoring, MODification) delineated in the article.

even ethnomusicology. However, they are rooted in evidence coming from broader academic disciplines—including cognitive science, (neuro)physiology, and motor control—to provide an explanation of underlying psychological and physiological principles. These concepts and theories are brought together into a general model to make a strong case for the role of music in biofeedback systems. This model is intended to provide a conceptual framework that guides future research and practice. However, the model is still preliminary and needs further testing to proof its validity.

MOTIVATION BY MUSICAL BIOFEEDBACK

In the context of sports and motor rehabilitation, situations of high endurance, pain, fatigue, and “repetition-to-boredom” are ubiquitous. Motivation—or, the *will* to act—is therefore an indispensable aspect to persevere in these situations. We argue that the use of music and sounds coupled to motor behaviors—i.e., sonification—is particularly powerful as it may take advantage of the strong motivational qualities inherent to people’s interactions with music. In the following, we show how music may stimulate active behavior, elicit strong emotions, elicit feelings of reward, and induce altered states of consciousness (trance), contributing to motivation.

Music and Motion

A prominent characteristic of music is that it motivates people to get physically active. Almost everyone has experienced the compelling drive to tap the feet, nod the head, sway arms and hips, or dance along with music. Gesturing along with music is very common and bodily renderings of musical and sonic features may be quite elaborate, encompassing musical beat, melody, dynamics, phrasing, etc. In this paragraph, we outline two neurophysiological mechanisms that contribute to these phenomena. A first one is an *arousal* mechanism within the central nervous system, a second one a *motor resonance* mechanism.

Arousal has been attributed to the functioning of the brainstem, more in particular the reticular formation (Pfaff, 2006; Juslin and Västfjäll, 2008; Pfaff et al., 2012). It is demonstrated that increased arousal occurs in response to salient, unexpected sensory events and promote increased sensory alertness, emotional reactivity, and instinctive or learned motor activity.

On the other hand, bodily renderings of musical and sonic features may rely on a motor resonance mechanism. Apart from overt movement responses to music, there is ample neurophysiological evidence that passive listening to music automatically co-activates motor regions within the brain (for examples, see Maes et al., 2014). A particular interesting musical feature shown to elicit body movements and motor cortex excitability in listeners is musical groove (Janata et al., 2012). These phenomena are considered motor resonance or ideomotor effects (related to theories on mirror neurons), referring to the process whereby perceptual (here, auditory) events trigger automatic muscular reactions based on previously established

action-perception relationships (for more details, we refer to Maes et al., 2014). In addition to this effect of becoming physically active, motor resonance may create the illusion of having control over the actual skillful production of music and sounds (cf. sense of agency).

Although arousal and motor resonance mechanisms are presented here as different mechanisms, they link in that they rely both on prediction processes in the brain. We argue that it is exactly because of the dynamics of predictability and surprise (and, tension and release) inherent to most music that arousal and motor resonance mechanisms are “set into action,” become relevant in explaining music-induced body movement, and eventually can be deployed for sports and motor rehabilitation purposes.

Other than affecting the timing of people’s movements, music has also been shown to affect the amount of vigor in people’s movements. A walking experiment, where people were instructed to synchronize their steps to the beat of songs at 130 BPM, revealed a significant effect of the type of music on step size and thus on walking velocity (Leman et al., 2013). It seemed that even though all songs were at the same tempo, some music had a relaxing effect, decreasing the step size, and other music had an activating effect, actually increasing the step size compared to the average step size of walking to a metronome.

This effect of music was also found in a self-paced walking experiment where people were not instructed to synchronize with tempo-matched music (Buhmann et al., 2016). Analysis of the musical features attributing to the velocity effect of the music, showed that music with a recurring pattern every four beats had an activating effect on kinematic responses, resulting in bigger stride lengths compared to walking without music. On the other hand, music with recurring patterns every three or six beats, had a relaxing effect on the kinematic responses, resulting in smaller stride lengths compared to walking without music. Such emphasis on ternary aspects of the meter could either be due to a 3/4 meter in songs, or it could be the result of syncopating melodies. In both cases this ternary recurring pattern in the music seems to counteract the regular flow of a binary walking pattern. Results indicate, that although musical groove elicits body movements, these movements may not always contribute to the forward movement in walking or running. The expressiveness in high-groove music is often represented by non-binary recurring patterns in the rhythm or melody of the music. In order to boost performance in cyclic, binary movements, such as walking or running, caution is needed to select the most suitable kind of music.

Another outcome of the study by Buhmann et al. (2016) emphasizes the motivational aspects of music-movement interaction. Subjects were asked to rate different motivational aspects of all the songs they heard during their walk. This was done with a BMRI-2 test (Karageorghis et al., 2006). Songs that increased walking velocity were rated significantly higher with respect to motivation, thus revealing a close link between the effect of music on walking velocity and the power of music to motivate.

Music and Emotion

Next to motion stimulation, music is well-known for eliciting emotions. More even, emotional expression and regulation is often heard as prime motivation for people's engagement with music. There is ample research evidence showing that listening to music can have an effect on activity in the limbic system, considered to be the brain's emotional core (Blood and Zatorre, 2001; Koelsch, 2010; Peretz et al., 2013), and might lead to that "spreading gooseflesh, hair-on-end feeling" better known as "chills" (Panksepp, 1995). Used as a biofeedback system, it has been shown that music may modulate physiological arousal (Bergstrom et al., 2014). It is important to acknowledge that emotional responses to music and sound are tied to individuals' personal traits, preferences familiarity, and music-related autobiographic memories (Kreutz et al., 2007; Barrett et al., 2010). Therefore, it is recommended to take these aspects into account in sonification designs for sports and motor rehabilitation. However, research also delineated surface features in music and sound, such as consonance, tempo, mode, and texture, that affect musical responses more universally, or at least cultural-wide (Webster and Weir, 2005; Fritz et al., 2009; Weninger et al., 2013). An interesting phenomenon is the unpleasant sounding of dissonance, in contrast to consonance. Low-level, sensory dissonance is based on the sensation of "beating" and "roughness" of interacting partials of (musical) sounds. Although it is still an ongoing debate whether a dislike of dissonance is a truly innate or rather learned phenomenon, Juslin and Västfjäll (2008) attribute this phenomenon along with our responses to other basic acoustic qualities, such as fast, loud, noisy, and very low- or high-frequenced sound—to brainstem reflexes. These researchers consider these reflexes innate and automatic causing emotions, which may further act as positive or negative reinforcement of behavior. This aspect of automatic behavioral reinforcement mediated by emotional responses is highly relevant for sonification purposes in the context of sports and rehabilitation (see Modification by Musical Biofeedback). As motor behavior is attracted toward pleasant sounding qualities, one may link these qualities to desired motor behavior (positive reinforcement), while unwanted motor behavior is then matched to unpleasant sounding qualities (negative reinforcement).

Reward

In the context of sports performance and motor rehabilitation, people need to be motivated to learn in order to develop new behaviors and solve problems in order to reach specific goals. In this context, the just-mentioned concept of reinforcement in relation to music and musical emotion may be highly relevant. The aspect of reinforcement in relation to learning of new motor skills is closely tied to the concept of "reward." In reinforcement learning, people are not told exactly what to, but it is assumed that people will act and behave so as to maximize their received reward. In other words, reward is considered a prime motivator, "reinforce," or "attractor" of actions and motor behavior by inducing feelings of pleasure and happiness. Hence, learning environments are setup in a way that wanted behavior yields maximal reward and people are stimulated to discover this

by themselves. Reward and punishment are thus considered constraints guiding motor behavior toward specific goals. This principle of reinforcement learning will make out the core of the third component of our model, namely motor modification. In the corresponding section, we will go into more detail on the physiological principles of reward, and on the musical features and principles that can exert a strong rewarding force in people's engagement with music.

Trancing

Complementary to emotions and mood induced by music listening are aspects relating to absorption, dissociation, and trance (Becker, 2004; Herbert, 2011; Schäfer et al., 2013; Clarke, 2014). These aspects point to changes of awareness and consciousness that occur when people are "deeply" engaged with music, either by listening or performing. In a seminal work by Herbert (2011), the concepts of absorption and dissociation are thoroughly defined. They are determined subjective qualities subsumed within the experience of altered consciousness, or trance experience. Absorption refers to the experience of being completely occupied with the musical stimulus, without requiring any mental effort. Dissociation involves mentally cutting off from surroundings and extraneous thoughts. The benefits of using these subjective qualities in sonification designs in the context of sports and motor rehabilitation are ample. Being detached from internal and external "concerns" may lead to a calming, pleasant, and effortless experience, in which a focus could be maintained on the physical or mental task at hand. At the same time, as Herbert emphasizes, trance experience helps focusing on music that, via both emotional and formal qualities, may interact with the physical or mental task at hand. More specifically, Becker refers to DeNora's (2000) concept of music (or read, sonification) as a prosthetic technology of the body that provides organizing properties to which a range of bodily and mental processes can be entrained. Also, highly relevant in the context of sports and rehabilitation, is the link made between trance experience and the insensitivity to pain and fatigue, and high physical endurance. Becker (2007) provided an explanation of altered pain responses in trance experiences based on theories of the biology and neurophysiology of consciousness (e.g., Damasio, 1999). According to Damasio, the term "emotion" designates a specific autonomic physiological response—i.e., not under conscious control. Becker (2007) believes that persons engaged in a trance experience take voluntarily control over the physiology of emotional arousal, leading to a reduction of normal pain response and fatigue, and high physical endurance.

In the field of sports, a range of studies on music and locomotion has revealed positive effects of music in general, and synchronous music in particular, similar to the effects of a trance experience. With respect to psychophysical outcomes, significant increases in time-to-exhaustion (TTE) were uncovered when running to both motivational and neutral music, compared to running without music (Terry et al., 2012). In addition, significant effects of music on ratings of perceived exertion (RPE) were found when running at sub-maximal intensities (Bood et al., 2013). Typically, there are two common explanations

given to account for these findings. A first one is based on an informational processing model claiming that proprioception information (effort sense) and affective information (emotional quality of music) is preprocessed in parallel (Rejeski, 1985; Boucher and Trenske, 1990). As only a certain portion of information can be processed at once, music may distract from focusing attention on the internal sense of effort in vigorous activities. A second explanation is based on the concept of agency. As shown by Fritz et al. (2013), musical agency, defined as the performance of bodily movement with the intention to modulate expressive features of the musical feedback (timbre, loudness), may have effects on perceived exertion during a physically strenuous task.

Concerning physiological outcomes, lower oxygen consumption has been reported for athletes running with music compared to running without acoustic stimuli (Terry et al., 2012). The use of music is associated with better running economy although more evidence is needed to support this claim. Another effect on physiology has been revealed on heart rate: treadmill running with music at (near-) maximal RPE resulted in increased heart rates (Bood et al., 2013). This implies that music helps runners to perform at higher intensities. Furthermore, a study on the psychophysiological effects of synchronous vs. asynchronous music during cycling reported positive effects of synchronous music over asynchronous music (Lim et al., 2014). Although synchronizing movement to rhythmic stimuli did not reduce metabolic cost, it did lower limb discomfort and had a stronger effect on arousal than compared to asynchronous music.

MONITORING BY MUSICAL BIOFEEDBACK

According to its definition, sonification pertains to “the use of non-speech audio to convey information,” whereby data relations are transformed into “perceived relations in an acoustical signal for the purpose of facilitating communication or interpretation” (Kramer et al., 1999). In the present article, we approach sonification in relation to physical activities in the domain of sports and rehabilitation. In this context, one often speaks about auditory biofeedback instead of sonification. Thereby, sound and music is considered an auditory information channel that may—through real-time feedback—sharpen the awareness of physiological processes (cardiovascular, respiratory, electro-dermal, etc.), motor kinetic and kinematic processes, performance parameter output (speed, force, height, etc.), and their relationships. New technologies enable measuring these processes and parameters, and translating them into auditory information streams. Accordingly, attention might be directed to previously non-conscious operations of the body (“body schema”) and it becomes possible to affect these operations (Metzinger, 2003). This is particularly interesting, as physiological processes, motor processes, and performance output parameters tend to efface itself from conscious experience in most behavior. Augmenting the natural monitoring mechanism with technologies that sonify these

processes and parameters may assist self-regulation leading to optimal performance quality and efficiency in the domain of sports and rehabilitation.

An early experiment, illustrating the potential of sound to make people aware of muscle activation was conducted by Basmajian (1963). Slight contractions of skeletal muscles often recruit only a few motor units that may not be apparent through normal “proprioceptive” feedback. In Basmajian’s experiment, gentle contractions of a hand muscle (*abductor pollicis brevis*) were made apparent to the subjects through combined auditory-visual feedback. Basmajian found that, over time, subjects obtained fine voluntary control of individual motor units, so as to be able to perform various “tricks,” such as the production of rhythms, doublets, and roll effects. Interestingly, aural feedback proved more useful than visual feedback in learning and retaining these skills. In sports, biofeedback was introduced in the 1970s for the purpose of stress regulation, motor retraining, and motor performance optimization (Zaichkowsky, 1982). In these early experiments, measurement tools were used to monitor a wide range of physiological functions (e.g., heart rate, blood pressure, and muscle tension) in order for subjects to learn to voluntarily control these functions most optimally, leading to lower anxiety levels and/or improved motor performance. Most of the auditory biofeedback systems that have been used in previous research and practice are limited to providing auditory feedback of one specific process or parameter. However, in the following, we pinpoint the relevance in sports and motor rehabilitation to be able to coordinate multiple biological and/or motor processes simultaneously. We argue that the use of sound and music is particularly suited to assist in this task, through multilayered auditory feedback strategies.

Multilayered Musical and Auditory Biofeedback Strategies

Basically, music hits our eardrum as a highly complex mixture of air vibrations coming from different sound sources and sound layers, and affected by the acoustic environment (e.g., reflections). In the following, we describe in more detail how the human auditory processing system deals with this complex mixture of sounds and consequently, how sonification strategies may capitalize on these principles for the benefits of sports and rehabilitation research and practice.

Auditory Stream Segregation and Integration

Auditory stream segregation and integration concerns the fusion and fission of sound streams. The human auditory apparatus is particularly well accustomed to decompose a complex mixture of sounds into separate units (cf. auditory perceptual objects), through a psychoacoustic process called “auditory stream segregation,” “fission,” or “auditory scene analysis” (Bregman, 1990). Auditory streams thereby appear as emergent patterns: they result both from the acoustical configuration from sound and from the auditory-brain disposition to process those patterns and generate percepts. In speech perception, this process relates to the “cocktail party problem,” denoting human’s ability to focus on a single voice within an auditory noisy environment (McDermott, 2009). In the context of music, auditory streaming

segregation enables for instance to discern different instrumental sections in a symphonic work, a rock song, or electronic dance music. Earlier experimental research has pinpointed specific physical properties that support the separation of music into different distinct sources, such as pitch, timbre, loudness, tempo, and rhythm (Miller and Heise, 1950; van Noorden, 1975; McAdams and Bregman, 1979). A simple example is a sequence (ABABAB) of subsequent low (A) and high (B) pitches, which segregates as separate streams of A A A and B B B depending on duration and pitch interval. New insights into auditory anatomy and physiological processes have deepened our understanding of auditory stream segregation (Snyder and Alain, 2007; Bizley and Cohen, 2013), with an important role attributed to predictive information processing (Grossberg, 1999; Winkler and Schröger, 2015).

Pitch-Harmony-Tonality-Based Emergence

Pitch, chord harmony and tonality are based on the fusion of acoustical harmonic components. Consequently, they can be considered as outcomes of an emergent process based on acoustical configurations and auditory disposition. The main effect is that different auditory perceptual objects may blend together into a single auditory perceptual object; cf. harmonicity-based fusion principle (McAdams et al., 2004). Exemplary for this principle are the tones produced by musical instruments. Typical for vibrating strings (cf. string instruments) and air columns (cf. wind and brass instruments) is that they create multiple, harmonically related frequency components (partials) that combine together into complex tones. More in particular, such components contribute to the perception of *pitch* (based on the lowest harmonic fundamental if present, or based on the harmonic residue if absent)—and a series of overtone frequencies or harmonics whose amplitudes determine perceptual *timbre*. Another example is a major chord, which is composed of three instrumental tones (e.g., piano, guitar, etc.) being a root tone (e.g., C4), with on top a major third (E4), and perfect fifth (G4). These tones fuse because the harmonic patterns of these tones fuse into a harmonic residue that supports a bass note (C2) as root of the chord.

Rhythm-Periodicity-Based Emergence

Rhythm, or periodicity in general, has similar emergent characteristics as harmonicity, but they occur in the temporal domain (cf. Stockhausen, 1957/1959). Music contains overlapping, periodically repeated acoustic-perceptible patterns, such as beats, measures, and phrases that define overall hierarchical temporal structures. Typically, the periodicities of these patterns are related, in the sense that they are proportional to integer multiples (Butler, 2006). Pulses or beats, groups of two or three beats that define a measure, and even sequences of irregular meter (containing beats in successive patterns of 2, 3, and 3 beats) are quite common in music. What emerges at this temporal level is again based on the acoustical configuration and the auditory-brain disposition. Making an abstraction of the actual musical content, it is possible to consider the manifold of periodic acoustic-perceptible patterns as a series of phase-locked oscillators, with an integer relationship of frequency.

Although these different layers may be distinguished in terms of source and timbre, their temporal integration contributes to the perceptual grouping and fusion of the stream of auditory events.

Research demonstrated that musical patterns characterized by integer-related temporal periods (i.e., metrical layers) are manifested in spontaneous movement or dance to music (Naveda and Leman, 2009, 2010; Leman and Naveda, 2010; Toiviainen et al., 2010). In their studies, Naveda and Leman investigated the relationship between Samba music and dance. Using the method of Periodicity Transforms (PT), they decomposed Samba dance movement into its constituting periodicities, matching the metrical layers within the music. Samba movements were found to have their energy level mainly concentrated within 2-beat periods. Based on these “fundamental” periods, repetitive spatial movement trajectories (i.e., “basic gestures”) could be extracted for each body part. Based on this, the authors proposed a sonification of the resultant basic gestures, which led to an auditory reproduction of traditional rhythmic structures of Samba music (Naveda and Leman, 2008).

Relevance for Sonification Purposes

Based on the above-described emergent principles that are typical for music—respectively, auditory stream segregation, Pitch-Harmony-Tonality-based emergence, and Rhythm-Periodicity-based emergence—we outline three strategies for the purpose of multilayer sonification of physiological and/or motor processes.

A first strategy is based on the human ability to process different auditory layers on top of each other. By assigning different layers of auditory feedback simultaneously to different physiological processes, motor processes, and performance parameters, one may achieve to raise awareness of how these processes work together in relation to performed output and feeling. Coordinated behavior can thus be sonified by means of different interacting layers (Naveda and Leman, 2008).

Muscle synergy, or ensemble muscular activation, describes the complex integration of muscle groups and nerves that underlie the most common motor behaviors. Current research gained insights into how muscle synergies are encoded into the nervous system (Bizzi and Cheung, 2013). The topic of muscle synergy is of major importance for sports and motor rehabilitation (Barroso et al., 2015; Ting et al., 2015). Imbalances in multi-muscle coordination may lead to impaired performance and even injuries. We claim here that the “orchestration” of muscle synergies may be realized through a coupled, well-balanced auditory “symphony” (i.e., multilayer auditory biofeedback). Sonification may enhance awareness of the relationship between the different muscles, and may hint for changes in motor behavior to perform optimally and avoid muscle congestion and injury.

This first strategy focuses on making different physiological/motor processes and performance parameters explicit in corresponding auditory biofeedback streams, thereby relying on human auditory stream segregation skills.

In a second strategy, we rely on before-mentioned fusion effects, in which different auditory layers blend together into

a single auditory perceptual object. Here the focus is on the “end-product” (i.e., the outcome) of processes working together, instead of on the explicit contribution of each process.

An example of a pleasant sounding outcome is the combination of three pure tones that are harmonically related, e.g., 440, 880, and 1320 Hz. Another example is a musical major chord composed of a root note, major third, and perfect fifth. In the fourth section, we go into more details about musical and psychoacoustic principles that relate to reward. This approach is representative for the reinforcement learning method mentioned before. The idea is that optimal coordination of processes lead to a pleasing auditory outcome (cf. reward), while less optimal coordination of processes lead to gradual degrading of this pleasing auditory outcome (cf. “punishment”). This idea may also be applied to the sonification of the performance outcome of different persons, for example when they are engaged in a synchronized activity. The degree of synchronization may be rewarded by sonification (Varni et al., 2010; Demey et al., 2013).

A third strategy relates specifically to temporal periodicity-based fusion. We mentioned earlier that music often contains repeated patterns of which periodicities are proportional to integer multiples. Hence, we proposed to represent musical content as a series of phase-locked oscillators. This idea of phase-locked oscillators is interesting in light of the various physiological and motor processes engaged in sports practice. Many physiological processes (such as breathing and heart beat) and motor processes in sports are characterized by periodic patterns, and may well be considered as oscillatory systems and biological oscillators. Interestingly, research indicates that some of these biological oscillators tend to synchronize to each other at different period proportions; e.g., cardiovascular-respiratory (Schäfer et al., 1999), cardiovascular-motor (Niizeki and Saitoh, 2014), and respiratory-motor (McConnell, 2011). McConnell (2011) pinpointed the importance of coordinating breathing and motor rhythms to maximize comfort and (metabolism) efficiency in running, cycling, and rowing performances (Hoffmann et al., 2012). The fact that both music and biological processes within an individual may be conceived as a multilayered oscillatory system is interesting. It suggests that music may assist in simultaneously synchronizing multiple biological oscillators at different periods. Note however that the proportions of the periodicities of biological oscillators change depending on exercise intensities. For instance, The breathing-pedal cadence ratio of competitive cyclists shifts from 1:3 to 1:2 as exercise intensity or duration increases (McConnell, 2011). This entails that, for synchronization purposes, proportions of acoustic-perceptible periodicities within the music should change accordingly in order to match corresponding biological periodicities.

Although different periodicity layers in a song might be used to sonify different physiological and motor processes in movement, these periodicity layers might all somehow be related (integer multiples) and impact each other. Depending on the sort of relation, they either enhance one another or not. Evidence for this effect was given in the studies by Leman et al. (2013) and Buhmann et al. (2016), where musical layers with certain

multiples of the beat tempo had an increasing effect on velocity (e.g., a periodicity of once every four beats), whereas other multiples had a decreasing effect on velocity (e.g., periodicities of once every three or six beats). In regard of using music to manipulate specific physiological or motor processes, we therefore need to consider that different multi-layer tempi can enhance or diminish the amount of strength a subject applies in the movement or breathing that is being monitored by the music.

MODIFICATION BY MUSICAL BIOFEEDBACK

The monitoring of ongoing motor behavior by means of music and sound information refers merely to the direct transfer of behavioral data (basically physiological, movement, and performance output data) into auditory form for the purpose of increasing self-awareness of one's own behavior. However, monitoring may often become an assistive component of biofeedback systems that target the modification of motor behavior. Basically, two approaches can be distinguished here.

The first approach requires that the learner has an explicit representation of the target behavior (i.e., goal) to which ongoing behavior can be compared. Both the target behavior and the actual, ongoing behavior can be represented through music or sound patterns to allow such comparison. Learning and adaptation then consists in reducing the error between ongoing and target behavior; a form of goal-driven learning (Ram and Leake, 1995). As Ram and Leake (1995, p.1) point out, the process of goal-driven learning is guided by reasoning, in order to make good decisions about when and what to learn, to select appropriate strategies for achieving the desired learning, and to guide the application of the chosen strategies. In this regard, monitoring—i.e., sharpening self-awareness—becomes an essential precondition for people to modify their behavior. And therefore, the process of modification cannot be disconnected from the ability to monitor oneself. This is the typical approach in motor learning and rehabilitation research.

In the following, we propose and focus on an alternative approach to behavior modification that does not rely on self-monitoring; in other words it does not require that the learner has an explicit representation neither of the own behavior nor of the target behavior. In that regard, modification becomes disconnected from a direct transfer of behavior parameters (physiology, movement, and performance output) into audible form. Instead of relying on such explicit representations, learning and adaptation is reward-based, using reinforcement principles. In reinforcement learning, people are not told exactly what to do (i.e., the goal), but it is assumed that people will act and behave so as to maximize their received reward. The human reward system is understood as a collection of neural structures that contain dopamine-secreting neurons in the midbrain with pathways to other brain structures, such as the striatum, the hippocampus, and the prefrontal cortex (Schultz, 1999; Wise, 2004). Music is a particularly relevant phenomenon in this context as it is a potent source of pleasure and reward for most people (Dubé and Le Bel,

2003). Previous studies have demonstrated that music listening can activate the human reward system (Blood and Zatorre, 2001; Menon and Levitin, 2005; Alluri et al., 2015). In line with the core idea of reinforcement learning, we argue that pleasant and rewarding states promoted by music may function as an attractive force (“attractor”) of motor behavior. In the following, we delineate two principles regulating musical reward that can be exploited in strategies for modification of motor behavior. One is based on auditory brainstem responses (“brainstem-driven reward”), the other on predictive processing (“prediction-driven reward”).

Brainstem-Driven Reward

The human brainstem is an evolutionary old part of the nervous system. Automatic brainstem responses occur very early in the brain’s processing of auditory information. They are typically responses to auditory events that signal alert to the presence of a potential threat. These auditory events involve sounds that are sudden, loud, dissonant, noisy, very low- or high-frequenced, or feature fast temporal patterns (Juslin and Västfjäll, 2008). Mediated by brainstem activation, these sounds have effects on attention and physiological arousal and eventually, on human behavior (e.g., fight-or-flight response). According to Juslin and Västfjäll (2008), if arousal is too high, listeners will experience the sounds and music as unpleasant and reject them. Hence, listeners will be attracted to music that induces an “optimum” level of physiological arousal. Interestingly, this mechanism may be used as a sonification strategy to influence behavior toward specific goals. Central in this strategy is to associate wanted behavior to pleasant auditory states and processes (cf. reward), while undesired behavior would then prompt unpleasant auditory states and processes (cf. “punishment”). We assume then that motor behavior is spontaneously attracted toward these rewarding (pleasant) states and processes, without explicit knowledge of the target behavior being required (cf. reinforcement learning).

Predictive Processing-Driven Reward

Within cognitive science, it becomes common to think of the brain as a prediction machine that draws upon learned statistical regularities about the world (Friston and Kiebel, 2009; Clark, 2013). In perceiving, acting, and taking decisions in our daily environment, we are constantly in the process of making predictions of ensuing sensory events, of the probable causes of these sensory events, and of the consequences of actions taken. Theories of music cognition (Huron, 2006) and musical expression (Leman, 2016) draw heavily upon prediction. Importantly, in music research, the ability to anticipate and predict musical events has been acknowledged a potent source of pleasure and reward (Huron, 2006; Gebauer et al., 2012; Zatorre and Salimpoor, 2013). In the following, we consider prediction in relation to auditory-motor synchronization and association learning. We demonstrate, based on empirical evidence, how sonification strategies can be developed based on these principles for the purpose of motor modification.

Auditory-Motor Synchronization

We define synchronization or temporal rhythmic entrainment as the process of adjustment of a motor rhythm (e.g., moving, breathing, etc.) to an external periodic force typically the musical beat. Synchronization to a musical stimulus requires the ability to predict the time at which a musical beat is about to occur. Successful prediction may lead to strong feelings of pleasure and control, such as in dance. Presumably because of that, people often exhibit a spontaneous tendency to synchronize motor rhythms to the musical beat. In that sense, musical beat may function as an attractor of temporal coordinated behavior of humans, such as human locomotion. Sonification through manipulating the timing of the beats can therefore “attract” and influence movement behavior.

In the context of movement rehabilitation, rhythmic auditory-motor entrainment is often referred to as “auditory rhythmic cuing.” There is extensive evidence showing that rhythmic cuing may support timing of upper-limb movement control and gait in persons with motor disorders (Schaefer, 2014; Thaut et al., 2014; Yoo and Kim, 2016). More than temporal control, research points out that rhythmic cuing may lead to adaptation and optimization of spatio-dynamic parameters of motor control, such as smoothing of velocity and acceleration profiles of cyclical movement. Thaut et al. (1999) have provided an explanation of these effects rooted in period synchronization, rather than in phase synchronization. By aligning a repetitive movement pattern to a fixed (and thus, anticipated) rhythmic interval, one obtains accurate and consistent time information throughout the complete movement cycle. In other words, period synchronization allows time to be calibrated against a repeated motor pattern (cf. Maes et al., 2015 discussed further in this article). Correspondingly, the brain may use this time information to optimize spatio-dynamic motor control. Seen from this perspective, beneficial effects of rhythmic cuing becomes evident for the purpose of sports and rehabilitative training (Schaefer, 2014; Thaut et al., 2014).

The fact that musical beat exerts an attractor force on rhythmical motor behavior may also be exploited differently for the purpose of movement monitoring, and eventually movement adaptation and optimization. To exemplify this point, we refer to a music application, called the D-Jogger, that uses sonification of gait tempo for manipulating entrainment (Moens et al., 2014). The D-Jogger system enables a real-time transfer of walking or running cadence data into a corresponding musical tempo. More in particular, the D-Jogger extracts discrete cues in runners’ motor behavior—namely, footfalls—to which discrete cues in the music—namely, musical beats—are aligned, both in period and phase. Through continuously adapting the tempo and phase of the music to the footfalls of a runner, D-Jogger monitors the running behavior and creates an audio image of the movement. This monitoring can be done with different alignment strategies that can set a range of coupling strengths between music and movement. In all strategies the initial tempo of a song would be aligned such that it equals the walking or running cadence. In a first strategy, the music is not started in phase. Over the course of the song, music tempo is continuously adjusted to the movement period. In a second strategy, the music does not start in phase as

well, but during the song, the music tempo remains stable and the subject entrains to the phase of the song. In a third strategy, the music starts in phase with the movement of the exerciser. During the song, the phase remains fixed, and the subject entrains to the phase of the song, while the music tempo continuously adapts to the period of the movement. In a fourth and last strategy, the music starts in phase and over the course of the song, both music tempo and phase are adapted continuously according to the exerciser's cadence, as such ensuring perfect synchronization between these two rhythms.

Experiments with these different D-Jogger alignment strategies reveal evidence for clear distinctions between different stages of synchronization (Moens et al., 2014). The first stage consists of recognizing the beat and the beat tempo, which has been shown to be the most problematic part of entrainment. The second stage consists of imitating or synchronizing with the beat tempo, a more straightforward component, since a temporal scheme has been established. In the final stage of being in phase, entrainment is no longer needed to stay synchronized with the musical beat. That's why strategies that directly phase-lock the exerciser's movement with the music are preferred: they allow one to accurately predict the beat from the start. In contrast, strategies that require an exerciser to find the beat by him- or her-self are more demanding since they are more prone to erroneous beat prediction: phase-correction adjustments may require some effort and possibly take time to be accurate.

In addition to monitoring, the D-Jogger system can also be used to modify movement. Recently, an experiment was conducted to investigate whether runners could be sped up or slowed down by spontaneous auditory-motor synchronization (Van Dyck et al., 2015). Recreational runners were asked to run four laps of 200 m, a task that was repeated 11 times with a short break in between each running sequence. During each first lap, participants ran at their own preferred tempo without musical accompaniment. The average running tempo during the first lap was measured and served as a reference for the tempo-matched music—realized by the D-Jogger system—that was played during the second lap. In the last two laps, the music tempo was either increased or decreased by 3.00, 2.50, 2.00, 1.50, or 1.00 percent or was kept stable. In general, findings of this study showed that recreational runners are able to adapt their running cadence to tempo changes in music without being instructed to do so and even without being aware of this attunement. Evidence for an entrainment basin was discovered: the degree of entrainment with the tempo of the music dropped significantly as soon as tempo increases of 2.50 percent were introduced, and also when tempo decreases of 3.00 percent were introduced.

The study by Van Dyck et al. (2015) is based on the concept of deliberately introducing small errors in a runner's beat prediction, by slightly deviating the tempo of the music from the subject's running cadence. The human innate tendency to minimize beat prediction errors causes the subject to entrain the movement with the new musical tempo. Introducing beat-prediction errors was hereby managed through time stretching; compressing or stretching the time between the beats, resulting in faster or slower music, respectively. Another method for

introducing beat-prediction errors is shifting the audio signal in such a way that the upcoming beat no longer coincides with the next predicted footfall. This concept is based on manipulating the phase angle between the moment of the beat and the footfall. Currently, studies are being conducted to assess the effects of these manipulations on human locomotion, running, and other types of rhythmical physical activity.

A crucial assumption underlying the use of the principle of rhythmic entrainment as attractor of rhythmic motor behavior is that people exhibit a spontaneous tendency to synchronize rhythmic movement to external rhythmical sound patterns. However, studies show that people may have weak beat perception (Leow et al., 2014), and may not always exhibit the spontaneous tendency to phase-synchronize to a musical beat (Buhmann et al., 2016).

The D-Jogger method described above relies on a two-step process; a musical rhythm becomes first automatically aligned to an individual's motor rhythm (cf. sonification) while successively, period and/or phase manipulations are applied to the musical rhythm. In this second step, the music becomes an external stimulus, and one relies on the principle of spontaneous entrainment to guide people's motor rhythms. However, it would be of interest to further study how an additional sonification of motor rhythms during tempo/phase manipulations of external music may contribute to spontaneous entrainment to this music. This additional sonification, which would stay aligned to motor rhythms, makes the discrepancy between the external musical rhythm and internal motor rhythms more evident, which in turn may reinforce motor adaptation in order to resolve this discrepancy. On top of that, a well-considered sonification design could instigate the feeling of actually participating in the musical outcome, as in a music performance. This feeling of control relates to the concept of "agency," which refers to the subjective sense of having voluntarily control over actions and their outcome.

Auditory-Motor Association Learning

Interaction with a biofeedback system can exploit auditory-motor association schemes. These schemes, or "internal models," can be conceived as predictions of auditory consequences of actions (Maes et al., 2014). Based on a learning process, one can reliably predict the auditory outcome of performed actions, allowing the intentional production of certain sounds. This ability of prediction and control may lead to intense feelings of pleasure and reward, as in (social) music performance. Now, when a mismatch occurs between the expected and the actual auditory outcome of performed actions (i.e., a prediction error), one has the spontaneous tendency to adapt one's actions to minimize prediction errors and realize the intended outcome (cf. reward) (Lalazar and Vaadia, 2008; Krakauer and Mazzoni, 2011; Van Der Steen and Keller, 2013). In that sense, learned auditory-motor associations may function as "attractors" of coordinated behavior of humans. Based on this principle, we propose a sonification strategy that is implemented into two distinct steps. In a first step, actions become coupled to perceptual (i.e., auditory) outcomes through an associative learning process. In a second step, the auditory outcomes become altered, which

is assumed to lead to spontaneous motor adaptation in a well-specified manner.

To test this idea, we conducted an experiment using a finger-tapping paradigm (Maes et al., 2015). Participants were instructed to perform on a typical synchronization-continuation task. Each finger-tap triggered a piano tone of which the amplitude's decay curve exponentially decreased in a way the tone exactly fitted the target interval (i.e., 1100 ms). Thus, the duration of the target interval was represented/sonified by this piano tone. In the synchronization phase, participants learned to associate this relationship between the interval that needed to be tapped out (action), and the auditory tone (perception). In the continuation phase, we assumed that participants could rely on the tone's decay curve to time their actions; namely that they would tap at the moment that the previous tone ceased (cf. "sensorimotor timing strategy"). Now, if people would actually deploy this sensorimotor timing strategy, this entailed that a gradual alteration of the tones' duration (shorter/longer) throughout the continuation phase, would lead to corresponding changes in interval production rate. Namely, a gradual shortening of the tones' duration would entail a speeding up, while a lengthening would entail a slowing down. The results of the experiment provided evidence for the former, and were interpreted based on error-correction and corresponding motor adaptation mechanisms. This study outlines important mechanisms underlying temporal sensorimotor coordination, which are of interest for further implementation in sonification strategies in the domain of sports and rehabilitation.

The Role of Novelty and Surprise

Although the ability to anticipate and control auditory events is a potent source of reward, music that is too repetitive, simple and conventional will not sustain reward responses. Current learning theories suggest that learning processes are grounded in occurring discrepancies (errors) between what is expected/predicted and what actually happens (Schultz, 1998, 2007; Waelti et al., 2001; Hazy et al., 2010). In these studies, it is shown that dopamine neurons encode reward learning prediction errors; a positive response of dopamine neurons occurs when a reward is given unexpectedly, while this response gradually decreases as that reward becomes increasingly predictable. In line with this finding, Fiorillo et al. (2003) found that dopamine response is maximal when the uncertainty of a given reward outcome is highest, and decreases when reward outcome becomes more predictable. These findings advocate the importance of tension and uncertainty in musical compositions in order for them to be experienced interesting and pleasant (Gebauer et al., 2012). In the context of musical biofeedback systems for sports and motor rehabilitation, these findings stress the necessity to include aspects of surprise and novelty into sonification designs and strategies in order to support learning, self-regulation, and motivation processes.

The Role of Expression

Leman (2016) defines music as emergent patterns endowed with expression. Expressive cues in music have a relationship with sound-encoded human gestures and have a special appeal as biosocial signal that promotes the formation of an interaction

pattern between partners involved (often called sender and receiver). Expression in music works as an affordance, that is, an opportunity for expressive responding. The assumption is that musical patterns with expressive cues have a higher affordance for expressive responses than musical patterns without these expressive cues. A good example is the difference between deadpan piano music (e.g., a MIDI score played on a MIDI grand piano) and the same piano music onto which expressive cues are added (Flossmann et al., 2012). Listeners tend to respond more to the expressive music than to non-expressive music. Leman et al. (2013) and Buhmann et al. (2016) provide clear evidence that musical expression (activating vs. relaxing expression) can affect the velocity of walking, with a maximum effect size of about plus or minus 10% velocity increase or decrease in task-related settings and 3% velocity increase or decrease in spontaneous settings. Whether this effect is obtained by means of arousal-mediation, or by means of cross-model audio-motor overlapping brain regions is a topic of future research. The main observation is that expression seems to function as a facilitator for establishing a human-music interaction using the principles described above.

CONCLUSION

Auditory feedback of physiological processes, motor processes, and performance parameters has been proven particularly useful in the domain of sports and motor rehabilitation, for the purpose of stress regulation, motor retraining, and motor performance optimization. Also, research has demonstrated effects of (background) music on physiological, psychological, and motor aspects. However, the use of music in biofeedback systems has been much less explored. In the current article, we asserted that the use of music in biofeedback systems might have a major added value to the benefit of psychological and physical optimization of sports and motor rehabilitation tasks. Therefore, we presented the 3Mo model to describe three main functions of music that contribute to these benefits; namely Motivation, Monitoring, and Modification. The main idea was to delineate important components and principles to take into account and exploit in sonification designs. These components and principles are explicitly oriented toward music, and fundamental knowledge on human's sensorimotor interaction with music, and on the underlying physiological and psychological processes.

Although the model delineates important concepts and principles, many challenges lie ahead in order to fully realize the potential of musical biofeedback. For instance, further research is required to define the musical and acoustic parameters giving music motivational or relaxing qualities. Also, new methods need to be developed to (automatically) analyze multilayered periodicity patterns in music. In addition, there is a need to test and further develop strategies for modification of physical movement behavior. These strategies should incorporate temporal aspects, as well as spatial and spatiotemporal aspects. Another important challenge is the further integration of physiological and physical processes. Finally, progress needs to be made in the domain of hardware and software in order to provide reliable signals to be sonified. The 3Mo model indicates that the use of sonification in the field of sports and motor rehabilitation requires close collaboration

with other fields of research, including musicology, psychology, neuroscience, physiology, engineering, and art. It is only through actual collaborative research that the full potential of musical biofeedback will be discovered and put into practice.

AUTHOR CONTRIBUTIONS

PM: Outlined the general structure of the article, and wrote the main part of the article. JB: Contributed substantially to the

writing of the article with a focus on the description of empirical studies. ML: Contributed substantially with the development and shaping of the conceptual framework outlined in the article.

FUNDING

This research was conducted in the framework of the EmcoMetecca II project, granted by Ghent University (Methusalem-BOF council) to ML.

REFERENCES

- Alluri, V., Brattico, E., Toiviainen, P., Burunat, I., Bogert, B., Numminen, J., et al. (2015). Musical expertise modulates functional connectivity of limbic regions during continuous music listening. *Psychomusicology* 25, 443–454. doi: 10.1037/pmu0000124
- Barrett, F. S., Grimm, K. J., Robins, R. W., Wildschut, T., Sedikides, C., and Janata, P. (2010). Music-evoked nostalgia: affect, memory, and personality. *Emotion* 10, 390–403. doi: 10.1037/a0019006
- Barroso, F. O., Torricelli, D., Bravo-Esteban, E., Taylor, J., Gómez-Soriano, J., Santos, C., et al. (2015). Muscle synergies in cycling after incomplete spinal cord injury: correlation with clinical measures of motor function and spasticity. *Front. Hum. Neurosci.* 9:706. doi: 10.3389/fnhum.2015.00706
- Basmajian, J. V. (1963). Control and training of individual motor units. *Science* 141, 440–441. doi: 10.1126/science.141.3579.440
- Becker, J. O. (2004). *Deep Listeners: Music, Emotion, and Trancing*. Bloomington, IN: Indiana University Press.
- Becker, J. O. (2007). “Music, trancing, and the absence of pain,” in *Pain and Its Transformations: The Interface of Biology and Culture*, eds S. Coakley and K. Fauiman Shlemay (Cambridge, MA: Harvard University Press), 166–194.
- Bergstrom, I., Seinfeld, S., Arroyo-Palacios, J., Slater, M., and Sanchez-Vives, M. V. (2014). Using music as a signal for biofeedback. *Int. J. Psychophysiol.* 93, 140–149. doi: 10.1016/j.ijpsycho.2013.04.013
- Bizley, J. K., and Cohen, Y. E. (2013). The what, where and how of auditory-object perception. *Nat. Rev. Neurosci.* 14, 693–707. doi: 10.1038/nrn3565
- Bizzi, E., and Cheung, V. C. K. (2013). The neural origin of muscle synergies. *Front. Comput. Neurosci.* 7:51. doi: 10.3389/fncom.2013.00051
- Blood, A. J., and Zatorre, R. J. (2001). Intensely pleasurable responses to music correlate with activity in brain regions implicated in reward and emotion. *Proc. Natl. Acad. Sci. U.S.A.* 98, 11818–11823. doi: 10.1073/pnas.191355898
- Bood, R. J., Nijssen, M., van der Kamp, J., and Roerdink, M. (2013). The power of auditory-motor synchronization in sports: enhancing running performance by coupling cadence with the right beats. *PLoS ONE* 8:e70758. doi: 10.1371/journal.pone.0070758
- Boutcher, S. H., and Tremske, M. (1990). The effects of sensory deprivation and music on perceived exertion and affect during exercise. *J. Sport Exerc. Psychol.* 12, 167–176. doi: 10.1123/jsep.12.2.167
- Bregman, A. S. (1990). *Auditory Scene Analysis: The Perceptual Organization of Sound*. Cambridge, MA: MIT press.
- Buhmann, J., Desmet, F., Moens, B., Van Dyck, E., and Leman, M. (2016). Spontaneous velocity effects of musical expression on self-paced walking. *PLoS ONE* 11:e0154414. doi: 10.1371/journal.pone.0154414
- Butler, M. J. (2006). *Unlocking the Groove*. Bloomington, IN: Indiana University Press
- Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behav. Brain Sci.* 36, 181–204. doi: 10.1017/S0140525X12000477
- Clarke, E. F. (2014). Lost and found in music: music, consciousness and subjectivity. *Musicae Sci.* 18, 354–368. doi: 10.1177/1029864914533812
- Damasio, A. R. (1999). *The Feeling of what Happens: Body and Emotion in the Making of Consciousness*. New York, NY: Harcourt Brace & Company.
- Demey, M., Muller, C., and Leman, M. (2013). “DanSync: a platform to study entrainment and joint-action during spontaneous dance in the context of a social music game,” in *Intelligent Technologies for Interactive Entertainment*, eds M. Mancas, N. d’Allesandro, X. Siebert, B. Gosselin, C. Valderama, and T. Dutoit (Berlin: Springer), 124–135.
- DeNora, T. (2000). *Music In Everyday Life*. Cambridge: Cambridge University Press.
- Dubé, L., and Le Bel, J. (2003). The content and structure of laypeople’s concept of pleasure. *Cogn. Emot.* 17, 263–295. doi: 10.1080/02699930302295
- Dubus, G., and Bresin, R. (2013). A systematic review of mapping strategies for the sonification of physical quantities. *PLoS ONE* 8:e82491. doi: 10.1371/journal.pone.0082491
- Fiorillo, C. D., Tobler, P. N., and Schultz, W. (2003). Discrete coding of reward probability and uncertainty by dopamine neurons. *Science* 299, 1898–1902. doi: 10.1126/science.1077349
- Flossmann, S., Grachten, M., and Widmer, G. (2012). “Expressive performance rendering with probabilistic models,” in *Guide to Computing for Expressive Music Performance*, eds A. Kirke and E. R. Miranda (Berlin: Springer), 75–98.
- Friston, K., and Kiebel, S. (2009). Predictive coding under the free-energy principle. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 364, 1211–1221. doi: 10.1098/rstb.2008.0300
- 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
- Fritz, T. H., 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
- Gebauer, L., Kringelbach, M. L., and Vuust, P. (2012). Ever-changing cycles of musical pleasure: the role of dopamine and anticipation. *Psychomusicology* 22, 152–167. doi: 10.1037/a0031126
- Giggins, O. M., Persson, U. M., and Caulfield, B. (2013). Biofeedback in rehabilitation. *J. Neuroeng. Rehabil.* 10:60. doi: 10.1186/1743-0003-10-60
- Grossberg, S. (1999). “Pitch-based streaming in auditory perception,” in *Musical Networks: Parallel Distributed Perception and Performance*, eds N. Griffith and P. Todd (Cambridge, MA: MIT Press), 117–140.
- Hazy, T. E., Frank, M. J., and O’Reilly, R. C. (2010). Neural mechanisms of acquired phasic dopamine responses in learning. *Neurosci. Biobehav. Rev.* 34, 701–720. doi: 10.1016/j.neubiorev.2009.11.019
- Herbert, R. (2011). *Everyday Music Listening: Absorption, Dissociation and Trancing*. Oxford: Ashgate Publishing, Ltd.
- Hermann, T., and Hunt, A. (2005). An introduction to interactive sonification. *IEEE Multimedia* 12, 20–24. doi: 10.1109/MMUL.2005.26
- Hermann, T., Hunt, A., and Neuhoff, J. G. (eds.). (2011). *The Sonification Handbook*. Berlin: Logos Publishing House.
- Hoffmann, C. P., Torregrosa, G., and Bardy, B. G. (2012). Sound stabilizes locomotor-respiratory coupling and reduces energy cost. *PLoS ONE* 7:e45206. doi: 10.1371/journal.pone.0045206
- Huang, H., Wolf, S. L., and He, J. (2006). Recent developments in biofeedback for neuromotor rehabilitation. *J. Neuroeng. Rehabil.* 3:11. doi: 10.1186/1743-0003-3-11
- Huron, D. B. (2006). *Sweet Anticipation: Music and the Psychology of Expectation*. Cambridge, MA: MIT Press.
- Janata, P., Tomic, S. T., and Haberman, J. M. (2012). Sensorimotor coupling in music and the psychology of the groove. *J. Exp. Psychol. Gen.* 141, 54–75. doi: 10.1037/a0024208

- Juslin, P. N., and Västfjäll, D. (2008). Emotional responses to music: the need to consider underlying mechanisms. *Behav. Brain Sci.* 31, 559–575. doi: 10.1017/S0140525X08005293
- Karageorghis, C. I., Priest, D. L., Terry, P. C., Chatzisarantis, N. L. D., and Lane, A. M. (2006). Redesign and initial validation of an instrument to assess the motivational qualities of music in exercise: The Brunel Music Rating Inventory-2. *J. Sports Sci.* 24, 899–909. doi: 10.1080/02640410500298107
- Koelsch, S. (2010). Towards a neural basis of music-evoked emotions. *Trends Cogn. Sci.* 14, 131–137. doi: 10.1016/j.tics.2010.01.002
- Kos, A., Umek, A., and Tomazic, S. (2015). “Biofeedback in sport: challenges in real-time motion tracking and processing,” in *Paper presented at the IEEE 15th International Conference on Bioinformatics and Bioengineering (BIBE)* (Belgrade).
- Krakauer, J. W., and Mazzoni, P. (2011). Human sensorimotor learning: adaptation, skill, and beyond. *Curr. Opin. Neurobiol.* 21, 636–644. doi: 10.1016/j.conb.2011.06.012
- Kramer, G., Walker, B. N., Bonebright, T. L., Cook, P., and Flowers, J. H. (1999). *Sonification Report: Status of the Field and Research Agenda: International Community for Auditory Display (ICAD)*. Santa Fe, NM: International Community for Auditory Display (ICAD).
- Kreutz, G., Ott, U., Teichmann, D., Osawa, P., and Vaitl, D. (2007). Using music to induce emotions: Influences of musical preference and absorption. *Psychol. Music* 36, 101–126. doi: 10.1177/0305735607082623
- Lalazar, H., and Vaadia, E. (2008). Neural basis of sensorimotor learning: modifying internal models. *Curr. Opin. Neurobiol.* 18, 573–581. doi: 10.1016/j.conb.2008.11.003
- Leman, M. (2016). *The Expressive Moment: How Interaction (with Music) Shapes Human Empowerment*. Cambridge, MA: MIT press.
- Leman, M., Moelants, D., Warewyck, M., Styns, F., van Noorden, L., and Martens, J.-P. (2013). Activating and relaxing music entrains the speed of beat synchronized walking. *PLoS ONE* 8:e67932. doi: 10.1371/journal.pone.0067932
- Leman, M., and Naveda, L. (2010). Basic gestures as spatiotemporal reference frames for repetitive dance/music patterns in Samba and Charleston. *Music Percept.* 28, 71–91. doi: 10.1525/mp.2010.28.1.71
- Leow, L.-A., Parrott, T., and Grahn, J. A. (2014). Individual differences in beat perception affect gait responses to low-and high-groove music. *Front. Hum. Neurosci.* 8:811. doi: 10.3389/fnhum.2014.00811
- Lim, H. B. T., Karageorghis, C. I., Romer, L. M., and Bishop, D. T. (2014). Psychophysiological effects of synchronous versus asynchronous music during cycling. *Med. Sci. Sports Exerc.* 46, 407–413. doi: 10.1249/MSS.0b013e3182a6378c
- Maes, P.-J., Giacofci, M., and Leman, M. (2015). Auditory and motor contributions to the timing of melodies under cognitive load. *J. Exp. Psychol. Hum. Percept. Perform.* 41, 1336–1352. doi: 10.1037/xhp0000085
- Maes, P.-J., Leman, M., Palmer, C., and Wanderley, M. M. (2014). Action-based effects on music perception. *Front. Psychol.* 4:1008. doi: 10.3389/fpsyg.2013.01008
- McAdams, S., and Bregman, A. (1979). Hearing musical streams. *Comput. Music J.* 3, 26–60.
- McAdams, S., Depalle, P., and Clarke, E. (2004). “Analyzing musical sound,” in *Empirical Musicology: Aims, Methods, Prospects*, eds E. Clarke and N. Cook (Oxford: Oxford University Press), 157–196.
- McConnell, A. (2011). *Breathe Strong, Perform Better*. Champaign, IL: Human Kinetics.
- McDermott, J. H. (2009). The cocktail party problem. *Curr. Biol.* 19, R1024–R1027. doi: 10.1016/j.cub.2009.09.005
- Menon, V., and Levitin, D. J. (2005). The rewards of music listening: response and physiological connectivity of the mesolimbic system. *Neuroimage* 28, 175–184. doi: 10.1016/j.neuroimage.2005.05.053
- Metzinger, T. (2003). Phenomenal transparency and cognitive self-reference. *Phenomenol. Cogn. Sci.* 2, 353–393. doi: 10.1023/B:PHEN.0000007366.42918.eb
- Miller, G. A., and Heise, G. A. (1950). The trill threshold. *J. Acoust. Soc. Am.* 22, 637–638. doi: 10.1121/1.1906663
- Moens, B., Muller, C., van Noorden, L., Franěk, M., Celie, B., Boone, J., et al. (2014). Encouraging spontaneous synchronisation with D-Jogger, an adaptive music player that aligns movement and music. *PLoS ONE* 9:e114234. doi: 10.1371/journal.pone.0114234
- Naveda, L., and Leman, M. (2008). “Sonification of samba dance using periodic pattern analysis,” in *Paper presented at the 4th International Conference on Digital Arts (ARTECH 2008)* (Porto).
- Naveda, L., and Leman, M. (2009). A cross-modal heuristic for periodic pattern analysis of samba music and dance. *J. New Music Res.* 38, 255–283. doi: 10.1080/09298210903105432
- Naveda, L., and Leman, M. (2010). The spatiotemporal representation of dance and music gestures using topological gesture analysis (TGA). *Music Percept.* 28, 93–111. doi: 10.1525/mp.2010.28.1.93
- Niizeki, K., and Saitoh, T. (2014). Cardiolocomotor phase synchronization during rhythmic exercise. *J. Phys. Fitness Sports Med.* 3, 11–20. doi: 10.7600/jpfsm.3.11
- Panksepp, J. (1995). The emotional sources of “chills” induced by music. *Music Percept.* 13, 171–207. doi: 10.2307/40285693
- Peretz, I., Aubé, W., and Armony, J. L. (2013). “Towards a neurobiology of musical emotions,” in *The Evolution of Emotional Communication: From Sounds in Nonhuman Mammals to Speech and Music in Man*, eds E. Attenmüller, S. Schmidt and E. Zimmerman (Oxford: Oxford University Press), 277–299.
- Pfaff, D. W. (2006). *Brain Arousal and Information Theory*. Cambridge, MA: Harvard University Press.
- Pfaff, D. W., Martin, E. M., and Faber, D. (2012). Origins of arousal: roles for medullary reticular neurons. *Trends Neurosci.* 35, 468–476. doi: 10.1016/j.tins.2012.04.008
- Ram, A., and Leake, D. B. (1995). “Learning, goals, and learning goals,” in *Goal-Driven Learning*, eds A. Ram and D. B. Leake (Cambridge, MA: MIT Press), 1–38.
- Rejeski, W. J. (1985). Perceived exertion: an active or passive process. *J. Sport Psychol.* 7, 371–378. doi: 10.1123/jsp.7.4.371
- Schaefer, R. S. (2014). Auditory rhythmic cueing in movement rehabilitation: findings and possible mechanisms. *Philos. Trans. R. Soc. B Biol. Sci.* 369:20130402. doi: 10.1098/rstb.2013.0402
- Schäfer, C., Rosenblum, M. G., Abel, H.-H., and Kurths, J. (1999). Synchronization in the human cardiorespiratory system. *Phys. Rev. E Stat. Phys. Plasmas Fluids Relat. Interdiscip. Topics* 60, 857–870. doi: 10.1103/PhysRevE.60.857
- Schäfer, T., Fachner, J., and Smukalla, M. (2013). Changes in the representation of space and time while listening to music. *Front. Psychol.* 4:508. doi: 10.3389/fpsyg.2013.00508
- Schultz, W. (1998). Predictive reward signal of dopamine neurons. *J. Neurophysiol.* 80, 1–27.
- Schultz, W. (1999). The reward signal of midbrain dopamine neurons. *News Physiol. Sci.* 14, 249–255.
- Schultz, W. (2007). Behavioral dopamine signals. *Trends Neurosci.* 30, 203–210. doi: 10.1016/j.tins.2007.03.007
- Sigrist, R., Rauter, G., Riener, R., and Wolf, P. (2013). Augmented visual, auditory, haptic, and multimodal feedback in motor learning: a review. *Psychon. Bull. Rev.* 20, 21–53. doi: 10.3758/s13423-012-0333-8
- Snyder, J. S., and Alain, C. (2007). Toward a neurophysiological theory of auditory stream segregation. *Psychol. Bull.* 133, 780–799. doi: 10.1037/0033-2909.133.5.780
- Stockhausen, K. (1957/1959). How time passes by. *Die Reihe*, 3, 10–41.
- Terry, P. C., Karageorghis, C. I., 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. H., Kenyon, G. P., Schauer, M. L., and McIntosh, G. C. (1999). The connection between rhythmicity and brain function. *IEEE Eng. Med. Biol. Mag.* 18, 101–108. doi: 10.1109/51.752991
- Thaut, M. H., McIntosh, G. C., and Hoemberg, V. (2014). Neurobiological foundations of neurologic music therapy: rhythmic entrainment and the motor system. *Front. Psychol.* 5:1185. doi: 10.3389/fpsyg.2014.01185
- Ting, L. H., Chiel, H. J., Trumbower, R. D., Allen, J. L., McKay, J. L., Hackney, M. E., et al. (2015). Neuromechanical principles underlying movement modularity and their implications for rehabilitation. *Neuron* 86, 38–54. doi: 10.1016/j.neuron.2015.02.042
- Toivainen, P., Luck, G., and Thompson, M. R. (2010). Embodied meter: hierarchical eigenmodes in music-induced movement. *Music Percept. Interdiscip. J.* 28, 59–70. doi: 10.1525/mp.2010.28.1.59

- Van Der Steen, M. C., and Keller, P. E. (2013). The ADaptation and Anticipation Model (ADAM) of sensorimotor synchronization. *Front. Hum. Neurosci.* 7:253. doi: 10.3389/fnhum.2013.00253
- Van Dyck, E., Moens, B., Buhmann, J., Demey, M., Coorevits, E., Dalla Bella, S., et al. (2015). Spontaneous entrainment of running cadence to music tempo. *Sports Med. Open* 1:15. doi: 10.1186/s40798-015-0025-9
- van Noorden, L. (1975). *Temporal Coherence in the Perception of Tone Sequences*. Ph.D. dissertation, Institute for Perception Research, Eindhoven.
- Varni, G., Volpe, G., and Camurri, A. (2010). A system for real-time multimodal analysis of nonverbal affective social interaction in user-centric media. *IEEE Trans. Multimedia* 12, 576–590. doi: 10.1109/TMM.2010.2052592
- Waelti, P., Dickinson, A., and Schultz, W. (2001). Dopamine responses comply with basic assumptions of formal learning theory. *Nature* 412, 43–48. doi: 10.1038/35083500
- Webster, G. D., and Weir, C. G. (2005). Emotional responses to music: interactive effects of mode, texture, and tempo. *Motiv. Emot.* 29, 19–39. doi: 10.1007/s11031-005-4414-0
- Weninger, F., Eyben, F., Schuller, B. W., Mortillaro, M., and Scherer, K. R. (2013). On the acoustics of emotion in audio: what speech, music, and sound have in common. *Front. Psychol.* 4:292. doi: 10.3389/fpsyg.2013.00292
- Winkler, I., and Schröger, E. (2015). Auditory perceptual objects as generative models: setting the stage for communication by sound. *Brain Lang.* 148, 1–22. doi: 10.1016/j.bandl.2015.05.003
- Wise, R. A. (2004). Dopamine, learning and motivation. *Nat. Rev. Neurosci.* 5, 483–494. doi: 10.1038/nrn1406
- Yoo, G. E., and Kim, S. J. (2016). Rhythmic auditory cueing in motor rehabilitation for stroke patients: systematic review and meta-analysis. *J. Music Ther.* 53, 149–177. doi: 10.1093/jmt/thw003
- Zaichkowsky, L. D. (1982). “Biofeedback for self-regulation of competitive stress,” in *Stress Management for Sport*, eds L. D. Zaichkowsky and W. E. Sime (Reston, VA: American Alliance for Health, Physical Education, Recreation and Dance), 55–64.
- Zatorre, R. J., and Salimpoor, V. N. (2013). From perception to pleasure: music and its neural substrates. *Proc. Natl. Acad. Sci. U.S.A.* 110(Suppl. 2), 10430–10437. doi: 10.1073/pnas.1301228110

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.

Copyright © 2016 Maes, Buhmann and Leman. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Commentary: Environmental Sound Training in Cochlear Implant Users

Nicholas Altieri*

Communication Sciences and Disorders, ISU Multimodal Language Processing Lab, Idaho State University, Pocatello, ID, USA

Keywords: cochlear implants, multimodal perception, environmental sound training, individual differences, generalization

A commentary on

Environmental Sound Training in Cochlear Implant Users

by Shafiro, V., Sheft, S., Kuvadia, S., and Gygi, B. (2015). *J. Speech Hear. Lang. Res.* 58, 509–519. doi: 10.1044/2015_JSLHR-H-14-0312

INTRODUCTION

Cochlear implants (CIs) are prosthetic devices developed for listeners with profound bilateral hearing loss. Despite considerable advances in CI hearing technology allowing for improved speech and language recognition, several studies have reported that the identification of common environmental sounds—even years after implantation plus high speech perception scores—prove difficult for most listeners (e.g., Loebach and Pisoni, 2008; Shafiro et al., 2011, 2015). The literature suggests explanations for environmental sound identification difficulty in CI users: The chief difficulty is that CI signals are highly degraded compared to the frequency-rich neural signal in normal-hearing (NH) listeners. CIs typically include 4 to 22 electrodes; this electrode array, while constituting a drastic improvement from early CIs containing 1 to 4 electrodes, still represents less than 1% of hair cells in a healthy cochlea contributing to sound-frequency information (Wilson, 2004). Besides the degraded signal provided by even state-of-the-art CIs, Shafiro et al. (2015) described other factors complicating environmental sound identification, namely, the likelihood of degraded representations of memory for environmental sounds caused by years of hearing loss.

To help address these concerns, I propose a modification of the environmental sound training procedure initially developed by Shafiro et al. (2015). The aim is to utilize multisensory cues, sounds presented in noise to enhance ecological validity, and a same-different discrimination phase prior to closed-set identification. This modified procedure should enhance neural plasticity, and consequently reconstruct auditory representations that have become degraded after years of CI use.

TRAINING PROGRAM OVERVIEW

Shafiro et al. (2015) reviewed studies utilizing training programs involving presenting post-lingually deafened CI users with environmental sounds (e.g., Inverso and Limb, 2010; Looi and Arnephy, 2010), or alternatively, presenting NH listeners with either 4 or 8-channel simulated CI signals (Loebach and Pisoni, 2008; Shafiro et al., 2012). Results consistently showed evidence for significant improvement in listeners' ability to identify environmental sounds subsequent to closed-set training. Interestingly, evidence for generalization to other categories was reported, including improved scores in speech recognition by Loebach and Pisoni (2008) and Shafiro et al. (2012) (who examined simulated sounds in NH listeners).

OPEN ACCESS

Edited by:

Diego Minciocchi,
University of Florence, Italy

Reviewed by:

Angelica Perez Fornos,
University of Geneva, Switzerland

*Correspondence:

Nicholas Altieri
altinich@isu.edu

Specialty section:

This article was submitted to
Neuroprosthetics,
a section of the journal
Frontiers in Neuroscience

Received: 01 November 2016

Accepted: 18 January 2017

Published: 01 February 2017

Citation:

Altieri N (2017) Commentary:
Environmental Sound Training in
Cochlear Implant Users.
Front. Neurosci. 11:36.
doi: 10.3389/fnins.2017.00036

TABLE 1 | Comparison of proposed training procedures.

Shafiro's training program	Proposed modification
Include Multiple Tokens of Sounds	Include Multiple Tokens of Sounds
Closed-Set Identification	Phase 1: Same-Different Discrimination
Stimuli presented in quiet	Stimuli presented in noise to improve ecological validity
Feedback for incorrect responses: Repeat stimulus three times	Feedback for incorrect responses: Step 1—present sound with video Step 2—present sound in quiet Step 3—present sound in noise
Only one phase	Phase 2: Closed-Set Identification
	Stimuli presented in noise to improve ecological validity
	Feedback for incorrect responses: Step 1—present sound with video Step 2—present sound in quiet Step 3—present sound in noise

In light of this research showing evidence for improved sound identification, Shafiro et al. (2015) developed a program to train post-lingually deafened CI users on a large closed-set of common sounds, and provide of a short 1-week computerized training program. The procedure consisted of two Pre-Test sessions separated by a week, another week of Training, and two Post-Test sessions each separated by 1 week. Each of these four sessions included two speech recognition tests (the CNC word recognition test; Peterson and Lehiste, 1962, and speech-in-noise SPIN-R; Elliott, 1995). Additionally, the Familiar Environmental Sound Test (FEST) was administered (Shafiro, 2008); FEST includes closed-set identification of 60 familiar sounds (160 words total; four tokens each) across five categories.

Sound-training involved training listeners on a subset of sounds obtained from FEST. On each training trial, a sound was presented and the listener was required to make a closed-set identification response. Feedback was critical to training: When a listener responded incorrectly, the program repeated the correct response three times before advancing to the next trial.

Shafiro et al.'s (2015) results indicated improved performance. Trained items showed the largest degree of improvement. Generalization was reported for untrained items, although performance on untrained items was substantially lower. Generalization, however, failed to occur for word or sentence recognition. Significant individual variability in environmental sound recognition skills was reported subsequent to training.

REFERENCES

Abel, S., Weiller, C., Huber, W., Willmes, K., and Specht, K. (2015). Therapy-induced brain reorganization patterns in aphasia. *Brain* 138, 1097–1112. doi: 10.1093/brain/awv022

Unfortunately, the authors observed that neither CI brand, length of implantation, nor age accounted for the variability. Variability was also observed across stimuli, with five items receiving particularly low identification scores even after training (e.g., “brushing teeth,” “blowing nose,” “zipper,” and “airplane flying”). Such sounds are “inharmonic,” possessing unique envelope cues that prove difficult for CI users to access.

OPTIMIZING ENVIRONMENTAL SOUND-TRAINING

To remedy these concerns, I propose a modified multimodal training procedure designed to improve sound-cue acquisition in CI users. Importantly, Shafiro et al. (2015) training utilized feedback. Incorrect responses were repeated three times before continuing. The first proposed modification will involve hierarchically structuring feedback: Each time a listener responds incorrectly, the first cue reinforcement will be to present the (without noise) with a video clip of the sound source. Next, the video will be removed and the same sound (or another token of the same sound) will be presented (again, without noise). The third cue will simply be a presentation of the sound at the same level of background noise used in testing. Studies on a wide variety of topics, from stroke patients with aphasia to traumatic brain injury patient with cognitive deficits support the efficacy of hierarchical cueing (Constantinidou et al., 2008; Abel et al., 2015). In an fMRI study examining the influence of hierarchical cueing therapy on brain reorganization in aphasia patients, Abel et al. (2015) reported that therapy gains appeared were associated with a decrease in brain activation. The observed activation decrease in the experimental group suggests that therapy gains facilitated *efficient* brain reorganization; efficient in the sense that less brain activation was required to perform the task.

Next, I suggest modifying the procedure by including a same-different detection phase (Phase 1) to reinforce and help encode representations (using two tokens on *same* trials)—this is especially important for difficult sounds such as “zippers.” Distinguishing “same” vs. “different” requires a lower-level cognitive decision; the ability to distinguish “same” vs. “different” is necessary although not sufficient for identification. Phase 2 will include the identification phase used by Shafiro albeit with the modified cueing procedure (Table 1). In controlled studies, these modifications will hypothetically reinforce auditory representations, improve generalization scores, and reduce variability among listeners and stimulus items.

AUTHOR CONTRIBUTIONS

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

Constantinidou, F., Thomas, R. D., and Robinson, L. (2008). Benefits of categorization training in patients with traumatic brain injury during post-acute rehabilitation: additional evidence from a randomized controlled trial. *J. Head Trauma Rehabil.* 23, 312–328. doi: 10.1097/01.HTR.0000336844.99079.2c

- Elliott, L. L. (1995). Verbal auditory closure and the Speech Perception in Noise (SPIN) test. *J. Speech Hear. Res.* 38, 1363–1376. doi: 10.1044/jshr.3806.1363
- Inverso, Y., and Limb, C. J. (2010). Cochlear implant-mediated perception of nonlinguistic sounds. *Ear Hear.* 31, 505–514. doi: 10.1097/AUD.0b013e3181d99a52
- Loebach, J. L., and Pisoni, D. B. (2008). Perceptual learning of spectrally degraded speech and environmental sounds. *J. Acoust. Soc. Am.* 123, 1126–1139. doi: 10.1121/1.2823453
- Looi, V., and Arnehy, J. (2010). Environmental sound perception of cochlear implant users. *Cochlear Implants Int.* 11, 203–227. doi: 10.1002/cii.428
- Peterson, G., and Lehiste, I. (1962). Revised CNC list for auditory tests. *J. Speech Hear. Disord.* 27, 62–70. doi: 10.1044/jshd.2701.62
- Shafiro V. (2008). Development of a large-item environmental sound test and the effects of short-term training with spectrally-degraded stimuli. *Ear Hear.* 29, 775–790. doi: 10.1097/AUD.0b013e31817e08ea
- Shafiro, V., Gygi, B., Cheng, M. Y., Vachhani, J., and Mulvey, M. (2011). Perception of environmental sounds by experienced cochlear implant patients. *Ear Hear.* 32, 511–523. doi: 10.1097/AUD.0b013e3182064a87
- Shafiro, V., Sheft, S., Gygi, B., and Ho, K. (2012). The influence of environmental sound training on the perception of spectrally degraded speech and environmental sounds. *Trends Amplif.* 16, 83–101. doi: 10.1177/1084713812454225
- Shafiro, V., Sheft, S., Kuvadia, S., and Gygi, B. (2015). Environmental sound training in cochlear implant users. *J. Speech Hear. Lang. Res.* 58, 509–519. doi: 10.1044/2015_JSLHR-H-14-0312
- Wilson, B. S. (2004). “Engineering design of cochlear implant systems,” in *Auditory Prostheses: Cochlear Implants and Beyond*, eds F. G. Zeng, A. N. Popper, and R. R. Fay (New York, NY: Springer-Verlag), 14–52.

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.

Copyright © 2017 Altieri. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



On the Auditory-Proprioception Substitution Hypothesis: Movement Sonification in Two Deafferented Subjects Learning to Write New Characters

Jérémy Danna* and Jean-Luc Velay

Aix-Marseille Université, CNRS, Laboratoire de Neurosciences Cognitives (LNC), Marseille, France

OPEN ACCESS

Edited by:

Diego Minciocchi,
University of Florence, Italy

Reviewed by:

Kazutaka Takahashi,
University of Chicago, USA
Amy L. Orsborn,
New York University, USA

*Correspondence:

Jérémy Danna
jeremy.danna@univ-amu.fr

Specialty section:

This article was submitted to
Neuroprosthetics,
a section of the journal
Frontiers in Neuroscience

Received: 02 December 2016

Accepted: 06 March 2017

Published: 23 March 2017

Citation:

Danna J and Velay J-L (2017) On the Auditory-Proprioception Substitution Hypothesis: Movement Sonification in Two Deafferented Subjects Learning to Write New Characters. *Front. Neurosci.* 11:137. doi: 10.3389/fnins.2017.00137

The aim of this study was to evaluate the compensatory effects of real-time auditory feedback on two proprioceptively deafferented subjects. The real-time auditory feedback was based on a movement sonification approach, consisting of translating some movement variables into synthetic sounds to make them audible. The two deafferented subjects and 16 age-matched control participants were asked to learn four new characters. The characters were learned under two different conditions, one without sonification and one with sonification, respecting a within-subject protocol. The results revealed that characters learned with sonification were reproduced more quickly and more fluently than characters learned without and that the effects of sonification were larger in deafferented than in control subjects. Secondly, whereas control subjects were able to learn the characters without sounds the deafferented subjects were able to learn them only when they were trained with sonification. Thirdly, although the improvement was still present in controls, the performance of deafferented subjects came back to the pre-test level 2 h after the training with sounds. Finally, the two deafferented subjects performed differently from each other, highlighting the importance of studying at least two subjects to better understand the loss of proprioception and its impact on motor control and learning. To conclude, movement sonification may compensate for a lack of proprioception, supporting the auditory-proprioception substitution hypothesis. However, sonification would act as a “sensory prosthesis” helping deafferented subjects to better feel their movements, without permanently modifying their motor performance once the prosthesis is removed. Potential clinical applications for motor rehabilitation are numerous: people with a limb prosthesis, with a stroke, or with some peripheral nerve injury may potentially be interested.

Keywords: sonification, real-time auditory feedback, proprioception, compensation, motor control, handwriting

INTRODUCTION

When someone is suffering from a loss of a given sensory modality, another preserved modality is generally used to supply equivalent sensory signals (for a review, see Bach-y-Rita and Kercel, 2003). The main trans-sensory systems were developed for blind persons, via tactile-vision substitution (e.g., Bach-y-Rita et al., 1998) or auditory-vision substitution

(e.g., Renier et al., 2005). However, using the auditory modality to compensate for proprioception loss, i.e., the auditory-proprioception substitution, remains an unexplored question. To address this issue, this study proposed to assess the effects of supplementary auditory feedback in subjects having a loss of proprioception.

The proprioceptive system includes sensory signals arising from several different receptors located in different body tissues (i.e., skin, joint capsule, tendon, muscle, ligamentous, and connective tissue). The proprioceptive function is to inform about static-position and the movement of body parts. The consequences of proprioceptive loss and the question of how far vision may supplement it, has been extensively studied in sensorimotor control or adaptation in deafferented subjects (Lajoie et al., 1992; Ghez et al., 1995; Sainburg et al., 1995; Nougier et al., 1996; Krakauer et al., 1999; Scheidt et al., 2005; Pipereit et al., 2006; Sarlegna et al., 2006, 2010). Without proprioception, subjects are very reliant on vision, and can show near normal performance in visual tasks such as visual reaching. In a mirror-drawing task, whereas healthy participants pay attention to controlling the incongruent information provided by visual and proprioceptive feedback, deafferented subjects have less difficulty adapting their movement because the sensory conflict does not exist for them (Lajoie et al., 1992). However, their performance is very much affected in tasks performed without continuous visual information (Fournier et al., 2002) or in tasks requiring adaptation in musculoskeletal dynamics for which proprioceptive feedback is critical (Sainburg et al., 1995; Krakauer et al., 1999).

Interestingly, the impact of proprioceptive loss on motor learning has mainly been studied in adaptation tasks but, to the best of our knowledge, never during the motor learning of a new pattern under normal, non-biased, perceptual conditions. When a new pattern is learned, the task-relevant sensory information, provided from both the environment and the body, are integrated to allow the fluent execution of the pattern and its memorization as an internal model (for a review, Wolpert et al., 2011). At the beginning, the learners do not have a kinesthetic reference for the movement and hence they should control it more visually. In arm control for example, hand trajectories would initially be planned in spatial coordinates without taking account of the joint motions (Morasso, 1981).

Until now, theories on motor control and learning have mainly focused on the role of vision and proprioception, considering the movements as silent. Yet, human actions often generate sounds whose variations directly inform about the movements made. When these sounds are systematically present during motor learning, strong audio-motor associations are created in such a way that, after learning, the sounds alone will evoke the movement, and reciprocally the silent movement will recall its associated sound (Kohler et al., 2002; Zatorre et al., 2007). Consequently, when movements are naturally silent, adding auditory information during motor execution may improve their control and thus facilitate memorization (e.g., Effenberg et al., 2016). The method, so-called movement sonification, consists of translating, in real-time, some movement parameters into synthetic sounds. The multimodal (visual,

auditory, and proprioceptive) integration of sonified movements has been shown to be effective in motor control and learning (see Sigrist et al., 2013 for a review). Applications are various, from sports practice (e.g., Effenberg, 2005), clinical rehabilitation (e.g., Scholz et al., 2014), to school education (e.g., Danna et al., 2014).

Handwriting is particularly relevant for evaluating auditory-proprioception substitution. Despite the faint scratching of the pen, handwriting is considered as a silent activity, mainly controlled by vision and proprioception (for a review, see Danna and Velay, 2015). Handwriting is possible without proprioceptive feedback. Provided vision was available, the quality of the written trace of a deafferented subject was comparable to that of control subjects (Teasdale et al., 1993). Nevertheless, when comparing a deafferented subject (one of the two subjects) and control subjects, though the words written by the former remained legible, the kinematics of her handwriting movement were deeply affected (Hepp-Reymond et al., 2009). To conclude, vision and proprioception are complementary in handwriting control: spatial information about the static trace is mainly provided by vision and movement information is mainly provided by proprioception (Danna and Velay, 2015).

For these reasons, we decided to sonify handwriting for auditory-proprioception substitution. Providing spatial information by the means of sounds is not a relevant strategy, because spatial information is already well supplied by vision. The purpose was rather to translate information about the movement, usually provided by proprioception, into auditory information. In particular, velocity signals, mainly provided by muscle spindles (Cordo et al., 2011), play a crucial role in movement perception and control when precise and fluid movement is required. So the question was, how to sonify handwriting velocity? Actually, when listening carefully to the sound produced during handwriting, a friction sound generated by the pen-paper interaction can be heard (Thoret et al., 2014) compared this real friction sound to a synthetic friction sound whose timbre was related to the pen velocity and they observed that this velocity sonification adequately informed about the pen displacement. They concluded that velocity sonification enters into a natural mapping between the sound and the action in order to contribute to the building of a multimodal sensorimotor representation of handwriting (Thoret et al., 2014). Based on this assumption, a similar sonification strategy was tested and validated for handwriting assessment (Danna et al., 2015a), handwriting learning (Danna et al., 2015b), and the rehabilitation of dysgraphia (Danna et al., 2014).

The auditory-proprioception substitution hypothesis was proposed as a strategy for stroke rehabilitation (Scholz et al., 2014) but we have yet to see the experimental validation of this hypothesis (Ghez et al., 2000) made an encouraging pilot study but without actual control experiments. The purpose of the present experiment was thus to assess the auditory-proprioception substitution in both two deafferented subjects and 16 control participants who all had to learn new characters with and without associated sonification.

Four predictions could be made before the experiment:

- (A) Because the present strategy consists in translating velocity information into sounds, we predict that training with sonification will improve handwriting kinematics and not spatial accuracy of the written trace (as already observed by Danna et al., 2015b).
- (B) The loss of proprioception is known to affect handwriting kinematics (Hepp-Reymond et al., 2009), therefore training with sonification will have a larger effect in deafferented than in control participants.
- (C) If the supplementary auditory signals only help to control the movement during training, but are not integrated within an internal multisensory model of the character, then applying sonification during training will only have a short-term effect.
- (D) Conversely, if the auditory signals are actually incorporated into the internal model of the character, then they will improve the kinematics of character production at longer term.

METHODS

Participants

Two deafferented subjects, GL (right-handed female, 65 years) and IW (left-handed male, 61 years), participated in the experiment. The Edinburgh Inventory (10-item version, Oldfield, 1971) conducted by Lefumat et al. (2016) revealed a Laterality Quotient of +77% for GL and −100% for IW. Both suffer from a complete loss of touch, vibration, pressure and kinesthetic senses below the neck in IW and below the nose in GL (Cooke et al., 1985; Cole and Sedgwick, 1992). The sural nerve biopsy conducted by Cooke et al. (1985) on GL revealed that fibers larger than 6.5 μm represented only 1.6% of the total number of fibers. However, both subjects have perceptions of pain and temperature, indicating a selective impairment of the large diameter peripheral sensory myelinated fibers. Motor fibers are not affected as shown by motor nerve conduction velocities and needle electromyography investigation of the arm muscles. H-reflexes are absent, no sensory nerve action potentials can be registered in the arms, and no cortical response can be evoked by electrical stimulation of the peripheral nerves of either arm. GL has been suffering from a permanent and specific loss of the large peripheral myelinated sensory fibers since she was 31 (for more details about her history and disease characteristics, see Cole and Paillard, 1995). IW experienced a permanent and specific loss of the large peripheral myelinated sensory fibers when he was 19 (for more details about his history and disease characteristics, Cole and Sedgwick, 1992; Cole and Paillard, 1995).

Sixteen healthy, age-matched control subjects (8 right-handed women and 8 men, between 58 and 68 years) volunteered for the experiment. Two of the controls were left-handed. None of the controls reported any relevant medical history. This study received a prior approval from the Ethics Committee of Aix-Marseille University and the CNRS (N° RCB 2010-A00155-34). All participants signed a written informed consent before starting the experiment, in accordance with the ethical standards set out in the Declaration of Helsinki.

Task

The task consisted of learning to write four new characters (**Figure 1A**) on a sheet of paper (A4 format: 21.0 \times 29.7 cm) affixed to a graphic tablet (Wacom, Intuos3 A4, sampling frequency 200 Hz) using an ink pen. The characters were extracted from the Tamil script. Character 4 was slightly modified in order to be drawn without lifting the pen. Each character was presented at the top of the sheet. A gray point has been added to indicate the starting point on the characters. A square (4.0 \times 4.0 cm) was drawn for each repetition in order to produce a character of comparable size.

Procedure

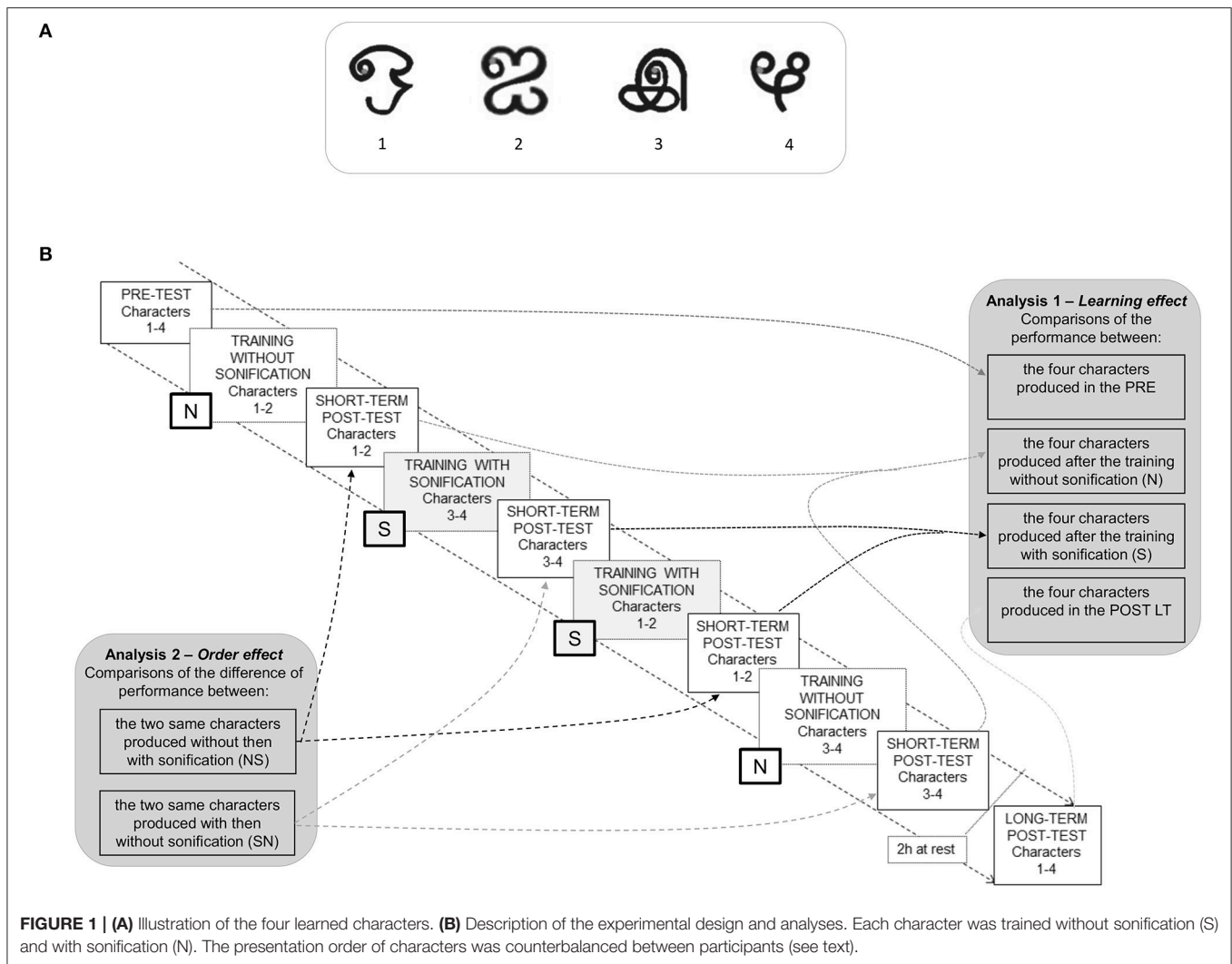
The experiment began with a short familiarization during which participants were asked to draw some simple geometric shapes with the auditory feedback in order to become informed about the meanings of the sonification. For the sake of clarity, the term “sonification” will be used rather than auditory feedback. The actual experimental design included a pre-test, a training session, and two post-tests. The first post-test (POST ST) was performed just after the training session of each character, and the second about 2 h later (POST LT). The pre-test and the two post-tests were exactly the same: The participants wrote each of the four characters once without sonification.

We used a classical within-subject ABBA protocol consisting of two different sessions (A and B) repeated in a different order. More precisely, the four characters were learned by pairs in two modes of training, one without sonification (session N) and one with sonification (session S), respecting the NSSN protocol. During the training sessions, the participants wrote each of the two characters 16 times. Two characters (characters 1 & 2) were learned first without, then with, sonification (order NS) and the other two (characters 3 & 4) were learned first with, then without, sonification (order SN; see **Figure 1B**). The order of characters written with sonification was counterbalanced between the two deafferented subjects and between controls in such a way that half of participants began the training sessions with characters 1–2 and the other half began with characters 3–4. Participants were asked to draw the characters in a single movement, without lifting the pen from the gray starting point to the end of the character.

Sonification Strategy

We applied the same sonification strategy already used in a previous study (Danna et al., 2015b), with the exception of impact sounds which were not present here. Sonification was generated in real time with Max software (<http://cycling74.com>). An example of sonified handwriting is available online in the Supplementary Material (**Supplementary Video File 1**).

A rubbing sound was associated to a correct handwriting velocity. This synthetic sound was close to the sound generated by writing with chalk on a blackboard. Technically, the synthesis was based on a source-resonator model which simulates the physical sound source as the result of successive impacts of a pencil on the asperities of a given surface. The surface roughness is modeled by a noise reflecting the height of the surface asperities while the



velocity profile of the pencil is modeled by low-pass filtering the noise with a time varying cutoff frequency that creates timbre variations according to the velocity profile (for more details, Conan et al., 2014).

When handwriting was too slow, the *rubbing* sound changed into *squeaking* sound. These squeaking sounds were based on non-linear (stick-slip) friction behavior (for more details, Thoret et al., 2013). This strategy is drawn from the metaphor of the squeaking of a door which naturally leads the writers to increase their movement speed in order to avoid this unpleasant noise. The synthesis model enabled sudden transitions between squeaking sounds and rubbing sounds. Transitions from the friction sound to the squeaking sound occurred when the instantaneous tangential velocity dropped below 1.5 cm s^{-1} (Danna et al., 2015b).

Finally, the pen pressure on the paper sheet, a measure directly provided by the tablet, was linearly associated to the sound volume, so that the greater the pen pressure the higher the volume.

Data Analysis

Four variables, three kinematic and one spatial, were computed from the (x,y) position of the pen on the tablet. The kinematic analyses were illustrated in Figure 2.

The movement velocity was the mean translational velocity from the starting point until the final lift, when the character was completed.

The number of abnormal velocity peaks was determined by the Signal-to-Noise velocity peaks difference (SNvpd). SNvpd is the difference between the number of velocity peaks after filtering the tangential velocity with a frequency cutoff (fc) of 10 Hz and the number of velocity peaks after filtering the tangential velocity with an fc of 5 Hz (Danna et al., 2013). Accordingly, the number of abnormal velocity peaks is an index of movement fluency: the less fluid the movement, the greater the number of abnormal velocity peaks and *vice versa*.

The number of stops was determined by counting the moments when the pen stopped during the drawing of the character. Note that stops are distinct from lifts of the pen: the

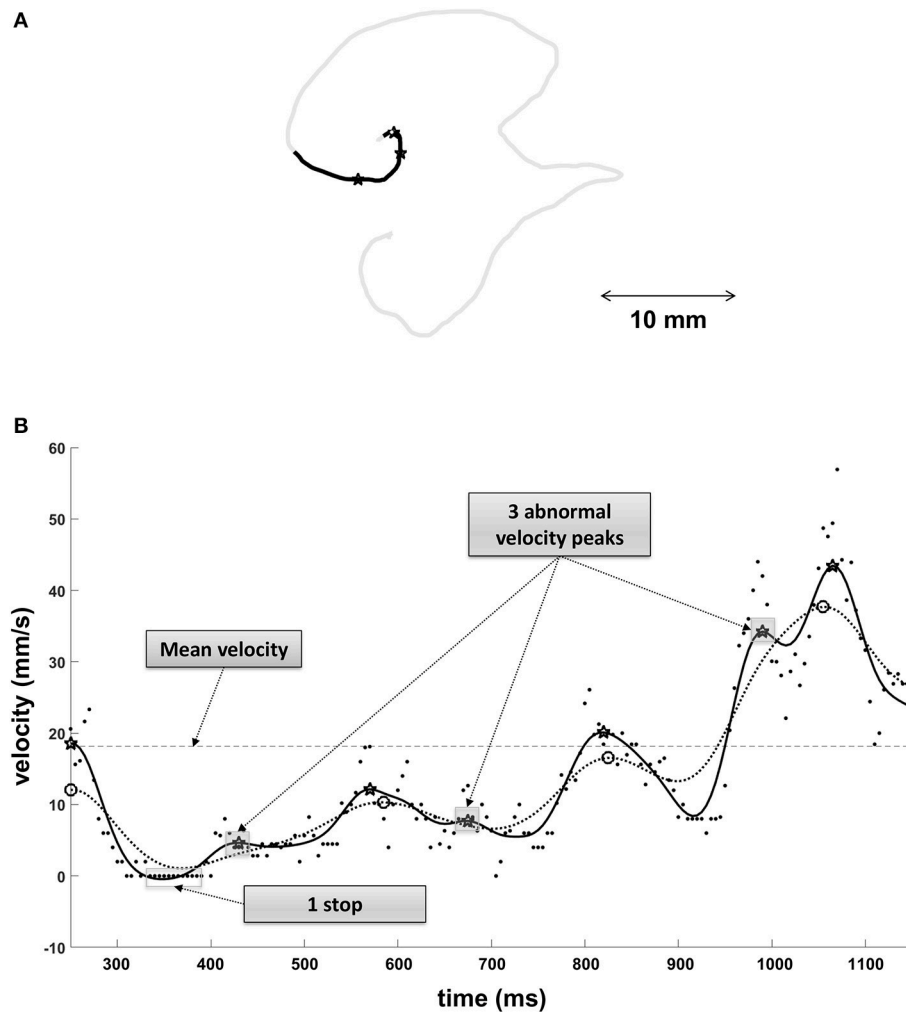


FIGURE 2 | (A) Illustration of a character written by GL. **(B)** Velocity profile from a portion of the written character (in black). The three kinematic variables are illustrated: The dotted line corresponds to the mean velocity. The three black stars correspond to the abnormal velocity peaks which are present only when filtering the velocity with a cutoff frequency at 10 Hz (black line) but not when filtering the velocity with a cutoff frequency at 5 Hz (dotted curve). Finally, the white rectangle corresponds to a stop (white dot in **A**) determined by a null (raw) velocity longer than 35 ms.

former occurred even though the pen was still in contact with the paper. Stops shorter than 35 ms were considered as normal stops (Paz-Villagrán et al., 2014). Therefore, here we only took into account those longer than 35 ms. Because the task consisted of writing the characters without lifting the pen, we assume that more stops were produced at the beginning, when participants had not yet memorized the characters and had to look at the model.

The spatial accuracy was determined by the Dynamic Time Warping (DTW) distance. DTW distance is the measurement of the spatial error between the character written by participants and a character prototype considered as a reference. More precisely, it corresponds to a point-to-point comparison between the two characters for which both spatial and temporal information is available. The DTW distance is computed as the average Euclidean distance between all pairs of matching points (for more details about criteria used for matching, see Niels et al.,

2007). The character prototypes were realized by a proficient adult who practiced writing each character with the aid of a model until the perfect shape was achieved. The series of (x,y) coordinates corresponding to the shape of each character were then filtered with a 4th order low-pass Butterworth filter with a fc of 5 Hz. These four characters were considered as “ideal” characters and the greater the disparity between them and the character drawn by a subject, the greater the DTW distance. For the sake of clarity, we took the inverse of DTW distance as an index of spatial accuracy: The better the character matched with the reference, the higher the score.

Statistics

Statistical analyses were conducted in two steps.

(1) *Learning effect*. As can be seen in **Figure 1B** (right), the effect of practice was assessed for the control group by computing the mean performance for the four characters written in the

pre-test (*PRE*), those written just after the training session with sonification (*POST ST—after S*), those written just after the training without sonification (*POST ST—after N*), and those written about 2 h afterwards (*POST LT*). These data were submitted to an analysis of variance (ANOVA) with the four Learning conditions as repeated measures and Bonferroni's post hoc tests when necessary. To compare GLs & IW's data to those of controls, we used *t*-test comparisons of a single value to a population sample (Nougier et al., 1996; Sarlegna et al., 2010) for the four learning conditions. The significance threshold was corrected to 0.0125 for the four *t*-test comparisons (Bonferroni's correction).

(2) *Order effect*. Within-subject ABBA protocol induces an order effect because some characters learned without sonification were learned after some characters learned with sonification. To evaluate the order effect, we averaged the performance in the same pairs of characters under the four conditions N, S, S, and N in the short term post-tests (**Figure 1B—left**). Then, we computed the difference of performance between the post-test of characters learned first with then without sonification, taking into account the presentation order, namely SN (without then with sonification) versus NS (with then without sonification). A difference significantly above or below zero revealed an effect of sonification and the order effect appeared if the difference in performance was observed in the NS order only. For that, we used *t*-test comparisons of a single value (0) to the controls' performance with Bonferroni's correction for the two presentation orders (significant threshold at 0.025). In order to assess whether sonification had a greater effect in deafferented subjects than in the controls, we also used *t*-tests comparisons of a single value to a population sample (with Bonferroni's correction) to compare the differences in performance of the controls to those of the deafferented subjects.

RESULTS

The effects of learning and sonification are presented in turn on each of the four variables analyzed.

Learning Effect

The performance of control and deafferented participants in the four learning tests without sonification are presented in **Figure 3**. Illustrations of the characters produced by GL and IW are supplied in the Supplementary Material (**Supplementary Figure 1**). Finally, the performance of control and deafferented subjects during the training phases with and without sonification are presented in **Figure 4**.

Movement velocity

The control group exhibited a main effect of learning, $F_{(3, 45)} = 15.24$, $p < 0.001$, $\eta_p^2 = 0.50$ (see **Figure 3A**). Bonferroni's *post-hoc* tests confirmed that the mean velocity in the three post-tests was higher than in the pre-test ($p < 0.001$). The comparison between the three post-tests was not significant.

The comparison between GL, IW, and controls revealed that GL was always slower than the controls ($p < 0.001$ for the four comparisons) whereas IW's velocity was comparable to that of

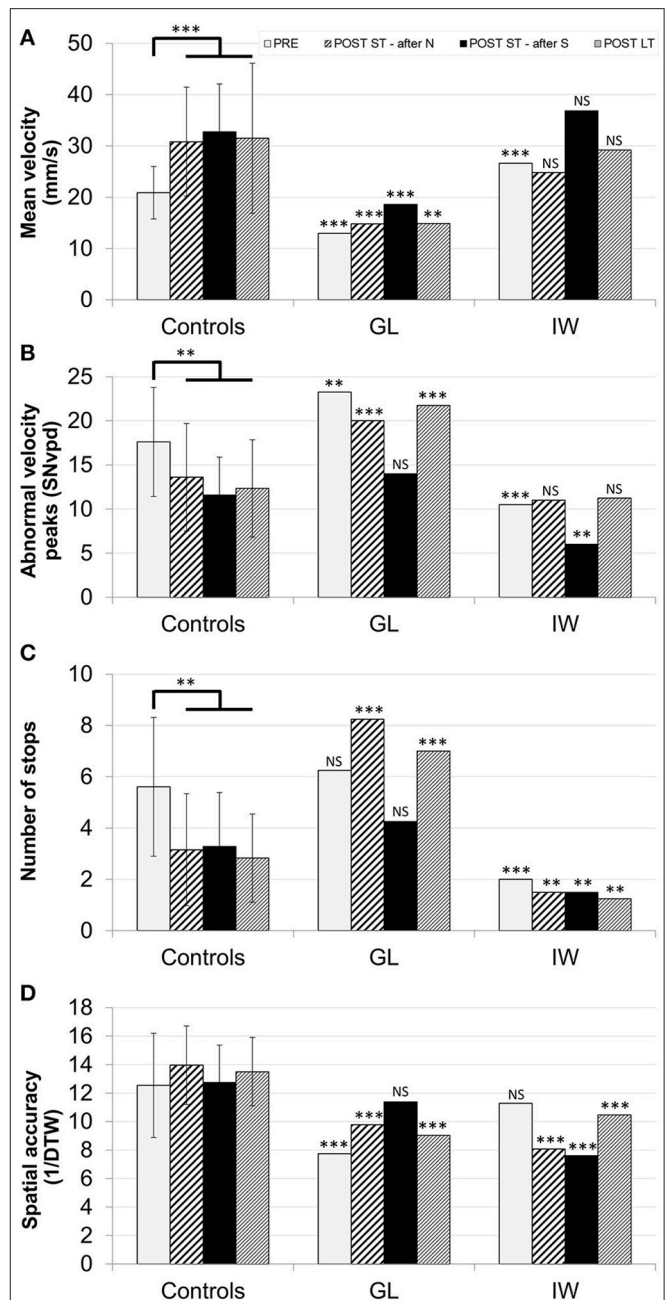


FIGURE 3 | Mean performance of the control group, GL and IW in the pre-test (*PRE*), the post-test following the learning sessions without sonification (*POST ST—after N*), the post-test following the learning sessions with sonification (*POST ST—after S*), and the post-test at T0 + 2 h (*POST LT*). Error bars correspond to between-participants SD of the control group. ** $p < 0.01$; *** $p < 0.001$. The performance was assessed with the mean velocity (**A**), the abnormal velocity peaks (**B**), the number of stops (**C**), and the spatial accuracy (**D**).

the controls except in the pre-test where it was even higher ($p < 0.001$). Moreover, contrary to the controls, both GLs and IW's velocities in the *POST ST—after N* and in the *POST LT* were not different from their initial velocity in the *PRE* (see **Figure 3A**).

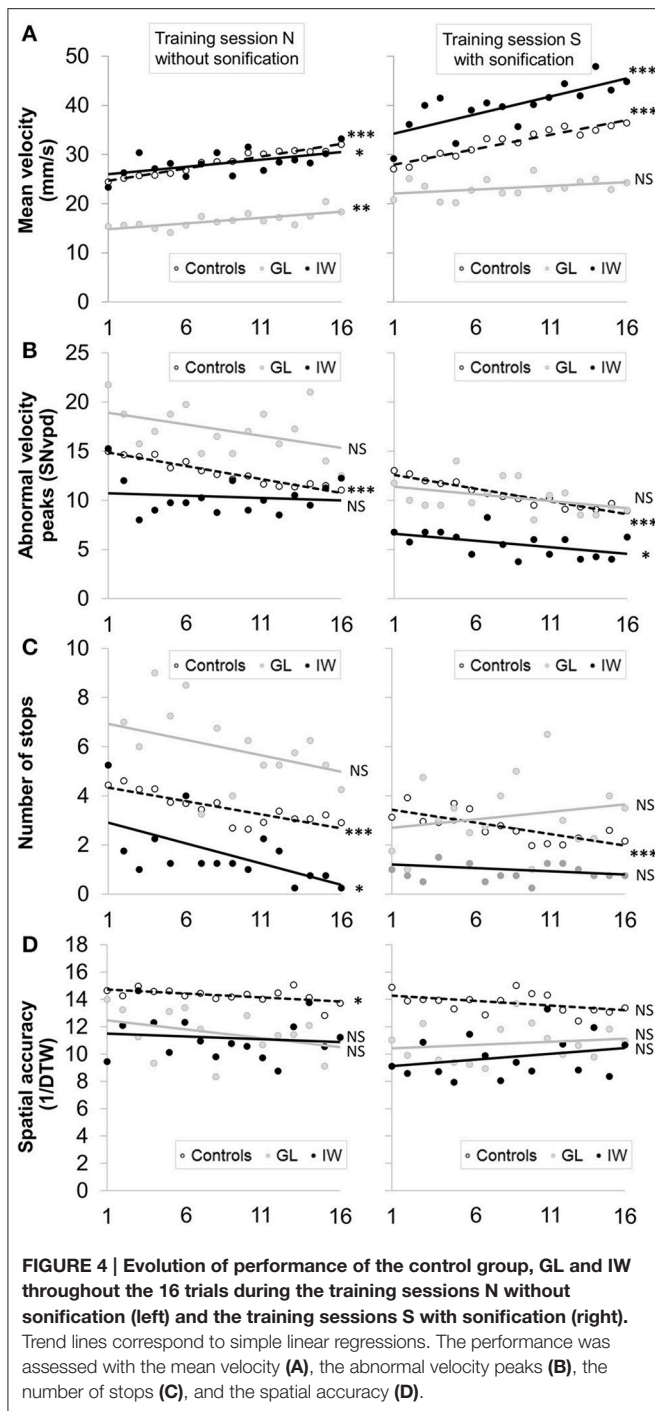


Figure 4A shows the evolution of velocity across the 16 repetitions within the training sessions (without and with sonification). When comparing the evolution of the velocity (trend lines), two observations can be made: (1) comparing the Y-intercept between sessions N and S gives an idea of the initial effect of sonification at the first trial, before learning. (2) comparing the slopes between sessions N and S informs about the sonification effect on learning progress over the 16 repetitions. In control participants, adding sounds during training (session S)

gave rise to a slight increase in writing speed (Y-intercept) but did not change the learning progression (identical slopes). GL was globally slower than the controls, but she benefited more than them from the presence of sonification at the first trial (Y-intercept). However, her learning slope was not modified by the sonification (null slope/non-significant regression in the session S). IW was quite similar to the controls without sonification (session N), however with the sonification (session S) both his initial speed and learning progression were greater than in control condition.

Number of Abnormal Velocity Peaks

The control group exhibited a main effect of learning, $F_{(3, 45)} = 11.35$, $p < 0.001$, $\eta_p^2 = 0.43$. Bonferroni's *post-hoc* tests confirmed that the number of abnormal velocity peaks in the three post-tests was lower than in the pre-test ($p < 0.01$, see **Figure 3B**). The two short-term post-tests were not different in the control group whatever the order of the training sessions they had followed.

As can be seen in **Figure 3B**, GL produced more abnormal velocity peaks than the control group in the *PRE*, the *POST ST—after N*, and the *POST LT* ($p < 0.01$ for the three comparisons), but not for the *POST ST—after S*. In other words, GL generally wrote the characters less fluently than the control participants, except when she had just learned the characters with sonification. The sonification effect on GL's movement fluency was larger than that on control participants.

IW wrote the characters with less abnormal velocity peaks than the control participants in the *PRE* and when he learned to write the characters with sonification ($p < 0.01$ for the two comparisons) but neither in the *POST ST—after N* nor in the *POST LT*.

Contrary to the control group, the movement fluency of GL and IW in the post-test of characters learned without sonification and in the post-test at $T0 + 2$ h were almost identical to their initial performance in the pre-test.

When comparing the evolution of the abnormal velocity peaks during the two modes of training in the control group (**Figure 4B**), a slight initial effect of sonification at the first trial (Y-intercept) was noted but no impact on learning progression (identical slopes) can be observed. Regarding GL, whereas she wrote the characters less fluently than the controls without sonification (with a great variability across repetitions), she performed close to the controls with sonification. However, due to her variable performance, the regression analysis did not reveal a significant evolution. Finally, IW was more fluent from the very beginning than the controls (lower Y-intercept) but he experienced no improvement across the 16 trials. With sonification, a positive effect at the first trial (Y-intercept) was noted, as well on his learning progression.

Number of Stops

The control group exhibited a main effect of learning, $F_{(3, 45)} = 8.51$, $p < 0.001$, $\eta_p^2 = 0.36$. The Bonferroni's *post-hoc* tests confirmed that the number of stops in the three post-tests was lower than in the pre-test ($p < 0.01$, see **Figure 3C**). The results of the comparison between the two short-term post-tests were not significantly different.

The deafferented subjects and the control groups all produced different results. GL had a significantly greater number of stops than the controls for the *POST ST—after N* and for the *POST LT* only ($p < 0.001$ for both comparisons, **Figure 3C**). During the *PRE* and the *POST ST—after S*, her stops number was comparable to controls. In other words, as for movement velocity and fluency, her stops numbers in the post-test of characters learned without sonification and in the *POST LT* were almost identical to her initial performance. This was not the case for the control group. Concerning the comparison of stops number between IW and the control group, the difference was significant for all tests. Even before learning, IW had seldom stopped during his movements, likely because he was using a different control strategy.

Regarding the evolution of the number of stops across the repetitions within the two modes of training (**Figure 4C**), the same observations as for abnormal velocity peaks can be made in the controls. Concerning GL, sonification allowed her to perform the task with a mean number of stops comparable to that of the controls but no learning progression was observed, whatever the training mode (N or S). IW produced very few stops (between 0 and 2, except at the first trial in the training session N), suggesting a feedforward control strategy.

Spatial Accuracy

In the control group, the spatial accuracy did not evolve from the pre-test to the long-term post-test, $F_{(3, 45)} = 1.07$, NS (see **Figure 3D**). GL drew the characters with a lower accuracy, except the characters which had been learned with sonification. Contrary to GL, IW displayed a spatial accuracy close to that of the control participants in the pre-test only. In all the post-tests, whatever the training mode, he showed a lower spatial accuracy than the control group, irrespective of the presence of sonification (see **Figure 3D**).

Comparing the evolution of spatial accuracy between the two training sessions indicates a slight effect of sonification on the performance variability in the control group (**Figure 4D**). In both GL and IW, 1- their spatial accuracy was lower than that of the controls, and 2- sonification had a slight negative effect on spatial accuracy (observed by a low Y-intercept in sessions S) but no effect on learning (null slope/non-significant regression in the sessions N and S).

Order Effect

As explained, our within-subject NSSN protocol induces an order effect. Differences of performance between the two *POST ST* (after S and after N) were thus computed in each order (NS vs. SN) and reported in **Table 1**.

Movement Velocity

In the control group, comparing the characters learned with versus without sonification (S-N) revealed that the velocity difference was significant in NS order ($p < 0.001$) but not in SN order ($p = 0.13$, see **Table 1**). As expected, this marked difference between the training orders suggests that when two characters were first learned with sonification, the gain in velocity was maintained afterwards when two new characters were trained without sonification.

TABLE 1 | Mean difference of performance (between-participants SD) between the post-test when characters were learned with sonification and the post-test when characters were learned without sonification (S-N) according to the order of presentation (NS: first without then with sonification vs. SN: first with then without sonification) for the control group, GL and IW.

		NS order: first without then with sonification	SN order: first with then without sonification
Difference in velocity	GL	8	−0.4
	IW	19.3	4.7
	controls	6.7 (±6.2)	−2.9 (±7.2)
Difference in velocity peaks	GL	−8.5	−3.5
	IW	−11	1
	controls	−3.2 (±3.5)	−0.8 (±6.6)
Difference in stops number	GL	−2.5	−5.5
	IW	−1.5	1.5
	controls	−0.4 (±2.5)	0.7 (±3.75)
Difference in spatial accuracy	GL	6.9	−3.7
	IW	−2.5	1.6
	controls	−1.3 (±3.3)	−1.1 (±3.7)

In bold: significant difference revealed by comparing the control group to zero (analysis 3). In bold and italic: significant difference revealed by comparing the control group to GL and to IW (analysis 4).

Does sonification have a greater effect on deafferented subjects than on the control group? Results revealed that, irrespective of the order (NS vs. SN), the difference of velocity between the post-test of characters learned with versus without sonification (S-N) was greater for IW than for the controls (see **Table 1**). Note that this was not the case for GL whose spontaneous velocity was much lower than that of IW and the control group.

Number of Abnormal Velocity Peaks

In the control group, comparing the fluency difference when characters were learned with vs. without sonification (S-N) revealed that the difference was significant in NS order ($p < 0.01$) but not in SN order ($p = 0.63$, see **Table 1**).

Does sonification have a greater effect on deafferented subjects than on the control group? Results showed that in NS order, i.e., when the characters were learned first without and then with sonification, the difference in abnormal velocity peaks between the two training sessions was larger in the deafferented subjects than in the control participants (see **Table 1**). In the reverse SN order, the difference of fluency was not significantly greater than in the control participants. This marked difference between the two orders of training suggests that, both in control and deafferented participants, the fluency increased following the training with sonification and stayed high, even though the subsequent characters were trained without sonification.

Number of Stops

In the control group, comparing the fluency of the characters learned with versus without sonification (S-N) revealed that,

irrespective of the order, the difference was not significant ($p = 0.49$ for NS order and $p = 0.47$ for SN order, see **Table 1**). Therefore, the number of stops was not influenced by sonification in the control group.

Whatever the order (NS vs. SN), the difference in the number of stops was greater for GL than for the control group. This was not true for IW who made few stops, whatever the learning task or the sonification condition.

Spatial Accuracy

In the control group, comparing the characters learned with versus without sonification (S-N) revealed that, irrespective of the order, the difference in spatial accuracy was not significant ($p = 0.13$ for NS order and $p = 0.24$ for SN order, see **Table 1**). These results confirmed that spatial accuracy was not influenced by sonification in the control group.

The increase in spatial accuracy was significantly greater in GL than in the control participants in the NS order only. IW's spatial accuracy was slightly greater than in the control participants in the reverse SN order (see **Table 1**).

DISCUSSION

The goal of this study was to evaluate auditory-proprioception substitution in two persons lacking proprioception. The effects of real-time auditory feedback were assessed during the motor learning of new graphic patterns. The results of this experiment can be summarized as follows:

In Control Participants

Overall, control participants were able to learn the new characters without sounds, but the sonification improved their learning: characters learned with sonification were reproduced more quickly and more fluently than those learned without. In other words, adding auditory kinematic signals during training lead to an improvement of kinematic variables when the characters were subsequently drawn without the sounds. These results are in agreement with those of a previous study where participants had to learn new characters with their non-dominant hand (Danna et al., 2015b). The improvement was present in the short term, but it was also observed in the longer term, 2 h after the end of the training sessions. However, this motor improvement was not accompanied by better spatial accuracy in the characters (prediction A). Note that the task consisted of reproducing graphic patterns with the dominant hand and in the presence of the model. We suppose that displaying the model allowed the participants to reproduce it accurately from the very first trial. Consequently, the learning consisted more of improving the kinematics than of improving the spatial accuracy, as children do when they learn how to write and free themselves from the models of the characters (Chartrel and Vinter, 2008).

The positive effects of sonification were present when the characters were first learned without then with sonification but not in the reverse order. This order effect, previously observed (Danna et al., 2015b), can be interpreted in the light of the theory of Event Coding (Prinz, 1997; Hommel et al., 2001): When characters have first been learned with sounds, a

multimodal (visual, proprioceptive and auditory) representation of the graphic pattern, including the internalized sounds, would have been created. Then, this multimodal representation would be reactivated even if the sounds associated with the movement are no longer supplied.

In Deafferented Subjects

Contrary to the control participants who performed better in all post-tests, whatever the sonification condition, the deafferented subjects were unable to learn the characters when training was attempted without sonification. In other words, they were unable to learn new kinematic properties leading to producing fluent graphic patterns whereas the controls were able to do so. This finding strongly suggests that without proprioceptive feedback, motor learning would be either longer or even impossible. This is consistent with the observation that handwriting automaticity in a deafferented patient (GL) was impaired and that proprioception would be a prerequisite to maintain a learned and automated complex motor behavior such as handwriting (Hepp-Reymond et al., 2009). More generally, it has been shown that proprioception plays an important role in the updating of an internal model of limb dynamics used to program motor commands (Sainburg et al., 1995; Krakauer et al., 1999; Pipereit et al., 2006), even if dynamic information may be inferred solely on the basis of vision (Fleury et al., 1995; Sarlegna et al., 2010).

Interestingly, movement sonification seems to be more efficient in deafferented persons than in control participants. In the short-term, the effects of sonification were larger in deafferented, than in control subjects for all kinematical variables (prediction B). More precisely, sonification gave rise to a larger improvement in movement fluency in both deafferented subjects, a larger improvement in velocity for IW than for the controls, and a larger decrease in stops for GL than for the controls. These findings support the hypothesis that translating kinematic information into auditory information substitutes for proprioceptive input. Hearing their sonified movement allowed the deafferented subjects to become informed about the kinematics of their movements that they can no longer feel through proprioception. As GL expressed after the experiment, they "feel their movement by hearing it." Another, more speculative, hypothesis to explain why deafferented subjects benefited more from the sonification could be that they process auditory information better than controls. It is known that sensory deprivation leads to significant cross-modal brain reorganization which is paralleled by enhanced perceptual abilities. For example, (Bavelier et al., 2006) showed enhancements in visual cognition in deaf subjects due to a reorganization of multisensory areas, highlighting cross-modal interactions as a fundamental feature of brain organization and cognitive processing. The symmetrical effect was observed by (Lessard et al., 1998) who showed that early-blind subjects were able to localize sound sources better than sighted subjects. However, sight and hearing both capture environmental information. Cross-modal enhancements between these two exteroceptive senses when one of them is missing is more likely than enhancements of auditory sensitivity in deafferented subjects although the reverse, enhancement of kinesthetic

sensitivity in deaf subjects, has been observed (Levänen and Hamdorf, 2001).

Although the sonification helped the deafferented subjects to learn the new characters in the short term, about 2 h after the training sessions, their performances were similar to those in the pre-test, contrary to the controls who maintained a higher performance. A first hypothesis is that applying auditory information only facilitates the control of ongoing movement in deafferented subjects but does not permit to learn a new motor pattern (prediction C). In other words, in the post-test following the learning sessions with sonification, they wrote better because they kept in short-term memory the movement they had performed just before, but not necessarily because they learned the motor pattern. This hypothesis is supported by their performance during the training sessions: Both deafferented subjects exhibited a fast effect of sonification, from the very first trials, but did not improve over the following repetitions. If this explanation holds, sonification would serve as “sensory prosthesis” helping the deafferented subjects to “feel” (by ear) their movement and to better produce it when the sounds are present, but would not able them to permanently change their motor performance without the prosthesis. Another hypothesis is that producing sonified movement during the training did lead the deafferented subjects to create a multimodal representation which was not maintained over the time in the present experiment because the training period was too short.

Between Deafferented Subjects

The initial performance differed between the two deafferented subjects who used opposite strategies. In the pre-test, GL was slower and less fluent than IW, confirming previous observations according to which GL would generally tend to use on-line visual feedback to guide her movement whereas IW would rely on forward motor planning (Cole and Paillard, 1995). These authors reported that both deafferented subjects can write, but their techniques for maintaining accuracy with their eyes shut differed: On the one hand, GL was very slow and, when drawing the letters, she tended to place them in the wrong area of the paper. On the other hand, IW moved fast across the page in an attempt to preserve both shape and correct framing of his writing space, at the cost of accuracy in the shape of the letters. If GL was slower than IW because of a greater reliance on visual control, why was she finally less accurate than him? It is likely that she tended to discretize her movements into many sub-movements (strokes) separated by stops (Ghez et al., 1990) have shown that the spatial accuracy of deafferented subjects was particularly affected at the endpoint of the movement, even under close visual control. We thus suppose that the stops made by GL in order to visually control her movement led her to be *in fine* less accurate.

Consequently, training and sonification had different effects in GL and IW. In GL, the learning curve is comparable to that of the controls (with more variability in her performance), with a greater effect on her. Usually, the poor performance of people beginning to learn to write is the consequence of a close visual control. This visual control gradually decreases with training, paving the way for a more automatic control (Danna and Velay, 2015). It is worth noting that audition is

available for the provision of supplementary information during the execution of silent movements, especially in deafferented subjects that use their vision for controlling and adapting their movements. Furthermore, according to the modality appropriateness hypothesis (Welch and Warren, 1980), audition would be more accurate than vision for the treatment of spatiotemporal information about the ongoing movements. Consequently, we hypothesize that training with sonification helped GL to decrease her visual control, leading her to write more fluently thanks to a shift from a *product-oriented* (the written trace) to a *process-oriented* (the movement that generates the trace) control. The initial performance of IW suggest a process-oriented control from the beginning of the learning task. Consequently, sonification during training would not change his initial feedforward strategy but led him to program faster movements to the detriment of spatial accuracy, suggesting a change in speed-accuracy tradeoff. In any case, the opposite results in GL and IW highlight the importance of studying two deafferented subjects to understand the impact of proprioception deprivation on motor control and learning.

Conclusion and Perspectives

This study confirms the potential of movement sonification for motor control and learning. Of course, sonifying the handwriting of people with total proprioceptive loss might appear anecdotal, but it demonstrates that auditory signals may act in substitution of proprioceptive deficit. Clinical applications may be numerous: people with a limb prosthesis, with a stroke, with some peripheral nerve injury, or parkinsonian patients with proprioceptive integration deficits (Schneider et al., 1987; e.g., Klockgether et al., 1995) may potentially be interested. Applied to other human movements, such as walking for example, sonification could be a new “prosthetic” device accessible at a much lower cost to millions of people. At a more fundamental level, neuroimaging and EEG studies must be conducted in order to determine the neural basis of auditory-proprioception substitution.

ETHICS STATEMENT

This study was carried out in accordance with the recommendations of the Aix-Marseille University and the CNRS (N°: RCB 2010-A00155-34) with written informed consent from all subjects. All subjects gave written informed consent in accordance with the Declaration of Helsinki. The protocol was approved by the Aix-Marseille University and the CNRS#39.

AUTHOR CONTRIBUTIONS

Conceived and designed the experiments: JD, JV. Performed the experiments: JD, JV. Analyzed the data: JD. Wrote the paper: JD, JV.

FUNDING

This work, carried out within the Labex BLRI (ANR-11-LABX-0036), has benefited from support from the French

Government, managed by the French National Agency for Research (ANR), under the project title Investments of the Future A/MIDEX (ANR-11-IDEX-0001-02) and under the CNRS project DEFISENS.

ACKNOWLEDGMENTS

We thank GL and IW for their participation, and Fabrice Sarlegna who organized their venue. We are grateful to Vietnam Paz-Villagrán and Olivia Vérove for helping in data recording, as well as Richard-Kronland Martinet, Sølvi Ystad, Mitsuko Aramaki and Charles Gondre of the Laboratoire de Mécanique et d'Acoustique (LMA, UPR 7051) for their collaboration in the project of handwriting sonification. Finally, we thank David Wood (English at

your Service, www.eays.eu) for revising the English of the paper.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fnins.2017.00137/full#supplementary-material>

Supplementary Video File 1 | Example of sonified handwriting of a deafferented subject (IW) in eight reproductions of a character during training session with sonification.

Supplementary Figure 1 | Presentation of the four characters written by the two deafferented subjects (GL and IW) in the pre-test (PRE), in the short-term post-tests following the training phases (POST ST) and in the long-term post-test (POST LT). The characters learned with sonification just before the POST ST are surrounded.

REFERENCES

- Bach-y-Rita, P., Kaczmarek, K. A., Tyler, M. E., and Garcia-Lara, J. (1998). Form perception with a 49-point electrotactile stimulus array on the tongue: a technical note. *J. Rehabil. Res. Dev.* 35, 427–430.
- Bach-y-Rita, P., and Kercel, S. W. (2003). Sensory substitution and the human-machine interface. *Trends Cogn. Sci.* 7, 541–546. doi: 10.1016/j.tics.2003.10.013
- Bavelier, D., Dye, M. W. G., and Hauser, P. C. (2006). Do deaf individuals see better? *Trends Cogn. Sci.* 10, 512–518. doi: 10.1016/j.tics.2006.09.006
- Chartrel, E., and Vinter, A. (2008). The impact of spatio-temporal constraints on cursive letter handwriting in children. *Learn. Instr.* 18, 537–547. doi: 10.1016/j.learninstruc.2007.11.003
- Cole, J. D., and Sedgwick, E. M. (1992). The perception of force and of movement in a man without large myelinated sensory afferents below the neck. *J. Physiol. Lond.* 449, 503–515. doi: 10.1113/jphysiol.1992.sp019099
- Cole, J., and Paillard, J. (1995). "Living without touch and peripheral information about body position and movement: studies with deafferented subjects," in *The Body and Self*, eds J. L. Bermudez, N. Eilan, and A. J. Marcel (Cambridge, MA: MIT Press), 245–266.
- Conan, S., Thoret, E., Aramaki, M., Derrien, O., Gondre, C., Kronland-Martin, R., et al. (2014). An intuitive synthesizer of continuous interaction sounds: rubbing, scratching and rolling. *Comput. Music J.* 38, 24–37. doi: 10.1162/COMJ_a_00266
- Cooke, J. D., Brown, S., Forget, R., and Lamarre, Y. (1985). Initial agonist burst duration changes with movement amplitude in a deafferented patient. *Exp. Brain Res.* 60, 184–187. doi: 10.1007/BF00237030
- Cordo, P. J., Horn, J. L., Künster, D., Cherry, A., Bratt, A., and Gurfinkel, V. (2011). Contributions of skin and muscle afferent input to movement sense in the human hand. *J. Neurophysiol.* 105, 1879–1888. doi: 10.1152/jn.00201.2010
- Danna, J., Fontaine, M., Paz-Villagrán, V., Gondre, C., Thoret, E., Aramaki, M., et al. (2015b). The effect of real-time auditory feedback on learning new characters. *Hum. Mov. Sci.* 43, 216–228. doi: 10.1016/j.humov.2014.12.002
- Danna, J., Paz-Villagrán, V., Capel, A., Pétoz, C., Gondre, C., Pinto, S., et al. (2014). "Handwriting movement sonification for the diagnosis and the rehabilitation of graphomotor disorders," in *Sound, Music and Motion*, eds M. Aramaki, O. Derrien, R. Kronland-Martin, and S. Ystad (Berlin; Heidelberg: Springer), 246–255.
- Danna, J., Paz-Villagrán, V., Gondre, C., Aramaki, M., Kronland-Martin, R., Ystad, S., et al. (2015a). Let me hear your handwriting!. Evaluating the movement quality from its sonification. *PLoS ONE* 10:e0128388. doi: 10.1016/j.humov.2014.12.002
- Danna, J., Paz-Villagrán, V., and Velay, J. L. (2013). Signal-to-Noise velocity peak difference: a new method for evaluating the handwriting movement fluency in children with dysgraphia. *Res. Dev. Disabil.* 34, 4375–4384. doi: 10.1016/j.ridd.2013.09.012
- Danna, J., and Velay, J. L. (2015). Basic and supplementary sensory feedback in handwriting. *Front. Psychol.* 6:169. doi: 10.3389/fpsyg.2015.00169
- Effenberg, A. O. (2005). Movement sonification: effects on perception and action. *IEEE Multimedia* 12, 53–59. doi: 10.1109/MMUL.2005.31
- Effenberg, A. O., Fehse, U., Schmitz, G., Krueger, B., and Mechling, H. (2016). Movement sonification: effects on motor learning beyond rhythmic adjustments. *Front. Neurosci.* 10:219. doi: 10.3389/fnins.2016.00219
- Fléury, M., Bard, C., Teasdale, N., Paillard, J., Cole, J., Lajoie, Y., et al. (1995). Weight judgment. The discrimination capacity of a deafferented subject. *Brain* 118, 1149–1156. doi: 10.1093/brain/118.5.1149
- Fournier, P., Paillard, J., Lamarre, Y., Cole, J., and Jeannerod, M. (2002). Lack of conscious recognition of one's own actions in a haptically deafferented patient. *Neuroreport* 13, 541–547. doi: 10.1097/00001756-200203250-00036
- Ghez, C., Gordon, J., and Ghilardi, M. F. (1995). Impairments of reaching movements in patients without proprioception. II. Effects of visual information on accuracy. *J. Neurophysiol.* 73, 361–372.
- Ghez, C., Gordon, J., Ghilardi, M. F., Christakos, C. N., and Cooper, S. E. (1990). Roles of proprioceptive input in the programming of arm trajectories. *Cold Spring Harb. Symp.* 55, 837–847. doi: 10.1101/SQB.1990.055.01.079
- Ghez, C., Rikakis, T., DuBois, R. L., and Cook, P. R. (2000). "An auditory display system for aiding interjoint coordination," in *Proceedings of the International Conference on Auditory Displays (ICAD)*, (Nashua, NH: ICAD).
- Hepp-Reymond, M. C., Charakov, V., Schulte-Mönting, J., Hueth, F., and Kristeva, R. (2009). Role of proprioception and vision in handwriting. *Brain Res. Bull.* 79, 365–370. doi: 10.1016/j.brainresbull.2009.05.013
- Hommel, B., Müsseler, J., Aschersleben, G., and Prinz, W. (2001). The theory of event coding (TEC): a framework for perception and action planning. *Behav. Brain Sci.* 24, 849–878. doi: 10.1017/S0140525X01000103
- Klockgether, T., Borutta, M., Rapp, H., Spieker, S., and Dichgans, J. (1995). A defect of kinesthesia in Parkinson's disease. *Mov. Disord.* 10, 460–465. doi: 10.1002/mds.870100410
- 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
- Krakauer, J. W., Ghilardi, M. F., and Ghez, C. (1999). Independent learning of internal models for kinematic and dynamic control of reaching. *Nat. Neurosci.* 2, 1026–1031. doi: 10.1038/14826
- Lajoie, Y., Paillard, J., Teasdale, N., Bard, C., Fléury, M., Forget, R., et al. (1992). Mirror drawing in a deafferented patient and normal subjects: visuoproprioceptive conflict. *Neurology* 42, 1104–1106. doi: 10.1212/WNL.42.5.1104
- Lefumat, H. Z., Miall, R. C., Cole, J. D., Bringoux, L., Bourdin, C., Vercher, J. L., et al. (2016). Generalization of force-field adaptation in proprioceptively-deafferented subjects. *Neurosci. Lett.* 616, 160–165. doi: 10.1016/j.neulet.2016.01.040

- Lessard, N., Paré, M., Lepore, F., and Lassonde, M. (1998). Early-blind human subjects localize sound sources better than sighted subjects. *Nature* 395, 278–280. doi: 10.1038/26228
- Levänen, S., and Hamdorf, D. (2001). Feeling vibrations: enhanced tactile sensitivity in congenitally deaf humans. *Neurosci. Lett.* 301, 75–77. doi: 10.1016/S0304-3940(01)01597-X
- Morasso, P. (1981). Spatial control of arm movements. *Exp. Brain Res.* 42, 223–227. doi: 10.1007/BF00236911
- Niels, R., Vuurpijl, L., and Schomaker, L. (2007). Automatic allograph matching in forensic writer identification. *Int. J. Pattern Recogn.* 21, 61–81. doi: 10.1142/S0218001407005302
- Nougier, V., Bard, C., Fleury, M., Teasdale, N., Cole, J., Forget, R., et al. (1996). Control of single-joint movements in deafferented patients: evidence for amplitude coding rather than position control. *Exp. Brain Res.* 109, 473–482. doi: 10.1007/BF00229632
- 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
- Paz-Villagrán, V., Danna, J., and Velay, J. L. (2014). Lifts and stops in proficient and dysgraphic handwriting. *Hum. Mov. Sci.* 33, 381–394. doi: 10.1016/j.humov.2013.11.005
- Pipereit, K., Bock, O., and Vercher, J. L. (2006). The contribution of proprioceptive feedback to sensorimotor adaptation. *Exp. Brain Res.* 174, 45–52. doi: 10.1007/s00221-006-0417-7
- Prinz, W. (1997). Perception and action planning. *Eur. J. Cogn. Psychol.* 9, 129–154. doi: 10.1080/713752551
- Renier, L., Collignon, O., Poirier, C., Tranduy, D., Vanlierde, A., Bol, A., et al. (2005). Cross-modal activation of visual cortex during depth perception using auditory substitution of vision. *Neuroimage* 26, 573–580. doi: 10.1016/j.neuroimage.2005.01.047
- Sainburg, R. L., Ghilardi, M. F., Poizner, H., and Ghez, C. (1995). Control of limb dynamics in normal subjects and patients without proprioception. *J. Neurophysiol.* 73, 820–835.
- Sarlegna, F. R., Gauthier, G. M., Bourdin, C., Vercher, J. L., and Blouin, J. (2006). Internally-driven control of reaching movements: a study on a proprioceptively deafferented subject. *Brain Res. Bull.* 69, 404–415. doi: 10.1016/j.brainresbull.2006.02.005
- Sarlegna, F. R., Malfait, N., Bringoux, L., Bourdin, C., and Vercher, J. L. (2010). Force-field adaptation without proprioception: can vision be used to model limb dynamics? *Neuropsychologia* 48, 60–67. doi: 10.1016/j.neuropsychologia.2009.08.011
- Scheidt, R. A., Conditt, M. A., Secco, E. L., and Mussa-Ivaldi, F. A. (2005). Interaction of visual and proprioceptive feedback during adaptation of human reaching movements. *J. Neurophysiol.* 93, 3200–3213. doi: 10.1152/jn.00947.2004
- Schneider, J. S., Diamond, S. G., and Markham, C. H. (1987). Parkinson's disease: sensory and motor problems in arms and hands. *Neurology* 37, 951–956. doi: 10.1212/WNL.37.6.951
- Scholz, D. S., Wu, L., Pirzer, J., Schneider, J., Rollnik, J. D., Großbach, M., et al. (2014). Sonification as a possible stroke rehabilitation strategy. *Front. Neurosci.* 8:332. doi: 10.3389/fnins.2014.00332
- Sigrist, R., Rauter, G., Riener, R., and Wolf, P. (2013). Augmented visual, auditory, haptic, and multimodal feedback in motor learning: a review. *Psychon. Bull. Rev.* 20, 21–53. doi: 10.3758/s13423-012-0333-8
- Teasdale, N., Forget, R., Bard, C., Paillard, J., Fleury, M., and Lamarre, Y. (1993). The role of proprioceptive information for the production of isometric forces and for handwriting tasks. *Acta Psychol.* 82, 179–191. doi: 10.1016/0001-6918(93)90011-F
- Thoret, E., Aramaki, M., Gondre, C., Kronland-Martinet, R., and Ystad, S. (2013). “Controlling a non-linear friction model for evocative sound synthesis applications,” in *Proceedings of the 16th International Conference on Digital Audio Effects*, (Maynooth).
- Thoret, E., Aramaki, M., Kronland-Martinet, R., Velay, J. L., and Ystad, S. (2014). From sound to shape: auditory perception of drawing movements. *J. Exp. Psychol. Hum.* 40, 983–994. doi: 10.1037/a0035441
- Welch, R. B., and Warren, D. H. (1980). Immediate perceptual response to inter-sensory discrepancy. *Psychol. Bull.* 88, 638–667. doi: 10.1037/0033-2909.88.3.638
- Wolpert, D. M., Diedrichsen, J., and Flanagan, J. R. (2011). Principles of sensorimotor learning. *Nat. Rev. Neurosci.* 12, 739–751. doi: 10.1038/nrn3112
- Zatorre, R. J., Chen, J. L., and Penhune, V. B. (2007). When the brain plays music: auditory-motor interactions in music perception and production. *Nat. Rev. Neurosci.* 8, 547–558. doi: 10.1038/nrn2152

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.

Copyright © 2017 Danna and Velay. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Mapping Sonification for Perception and Action in Motor Skill Learning

John F. Dyer^{1*}, Paul Stapleton² and Matthew Rodger¹

¹ School of Psychology, Queen's University Belfast, Antrim, United Kingdom, ² Sonic Arts Research Centre, School of Arts, English and Languages, Queen's University Belfast, Antrim, United Kingdom

Keywords: sonification, motor skill learning, augmented feedback, concurrent feedback, guidance effect, perception and action

INTRODUCTION, GOAL, AND SCOPE

Real-time sonification of human movement (conversion of motion signals into sound) can be used as augmented feedback for motor skill learning. With sonification, motor skills can, in some instances, be learned more quickly and successfully (Sigrist et al., 2013a). The goal of such sonification systems is a permanent (or at least, lasting) improvement in performance at a physical task or skill, which persists in the absence of augmented feedback. Many experimental investigations of feedback, however, show that when performance is tested without feedback, a decline occurs (Park et al., 2000; Maslovat et al., 2009). This finding has become known as the “guidance effect” (Buchanan and Wang, 2012). It has been suggested that this effect is a consequence of learner overreliance on the “guidance” provided by augmented feedback, at the expense of task-intrinsic sensory feedback. For effective learning, this is clearly not desirable.

In this paper, we advocate a perception-action approach to sonification when used as feedback for skill learning, which may lead researchers and trainers to design more effective prototypes. We highlight three main issues: 1. The learner's task should be conceived as perception-action based and sonification designed accordingly, 2. Sonification should provide Ecological information for perception rather than propositional knowledge-of-performance, and 3. Ecologically meaningful sound morphologies should be harnessed effectively.

PERFORMANCE

Successful coordination requires the pickup and use of event-structured information through perception and action (Gibson, 1972). Perceptual information available to a moving agent can be said to specify the state of the agent-environment system (i.e., the task); this enables the skilled agent to control movement and perceive its outcome “directly” (Warren, 2006; for a formal explanation of the relation between Ecological information and task dynamics, see Turvey et al., 1981). Through repeated interactions with a task, novices can become selectively sensitive to informational variables which best serve task goals, and become more adept at bringing these variables into use for coordination—a process described by Eleanor Gibson as “education of attention” (Gibson, 1969; Jacobs and Michaels, 2007). Motor skill learning is therefore characterized by the “tuning in” of perception and action—a progression toward the active pickup of better-specifying and more useful informational variables (Huys et al., 2009; Gray, 2010; Wilson et al., 2010a). Stoffregen and Bardy (2001) propose that action is controlled via the pickup of multisensory informational variables, which better specify the state of the perception-action system than can unimodal variables. Sonification can therefore provide higher-order information—available via interaction and specific to that interaction—which enables better perception and control of movement. With sonification, novices can practice with an enhanced, more responsive perceptual-motor workspace (defined as the emergent resources and constraints of organism and environment in the context of a task, which

OPEN ACCESS

Edited by:

Diego Minciaccchi,
University of Florence, Italy

Reviewed by:

Andrew D. Wilson,
Leeds Beckett University,
United Kingdom
Alfred Oliver Effenberg,
Leibniz University of Hanover,
Germany

*Correspondence:

John F. Dyer
jdye01@qub.ac.uk

Specialty section:

This article was submitted to
Neuroprosthetics,
a section of the journal
Frontiers in Neuroscience

Received: 20 February 2017

Accepted: 07 August 2017

Published: 21 August 2017

Citation:

Dyer JF, Stapleton P and Rodger M
(2017) Mapping Sonification for
Perception and Action in Motor Skill
Learning. *Front. Neurosci.* 11:463.
doi: 10.3389/fnins.2017.00463

are perceptually available through dynamic interaction: see Newell et al., 1991), which employs sound as a helpful constraint on action.

A model of the informational variables available and useful for the learner in a task can be guided by existing literature, or refined by pilot testing. However, the most useful informational variable(s) for the perceiving learner may not necessarily correspond to the motor variable being tracked by the researcher as a measure of performance. In other words, measurement and experience are not isomorphic. As an example, consider research on sonified reaching and target tracking (Oscari et al., 2012; Schmitz and Bock, 2014; Boyer et al., 2016). The task as instantiated here is to track or reach for a target, while using whatever information is provided by the system to guide one's effector/pointer. The variable of interest for measurement in this task (and others) is the absolute positional difference between hand/pointer position and target position (error). This variable is frequently sonified, by mapping error to a sonic variable such as pitch, amplitude or inter-aural panning, with mixed results (Konttinen et al., 2004; Rosati et al., 2012; Sigrist et al., 2013b). It is not certain that instantaneous positional error is a relevant variable for a moving individual in an everyday context. Everyday pointing for example, is primarily a visuomotor task with a criterion for success often defined in social terms (Kennedy, 1985). It makes sense from the detached perspective of an experimenter to measure positional error as an objective performance index, but perhaps another, possibly higher-order variable might be more useful for the learner as a perceiver (Runeson, 1977). The choice of what to sonify may have additional implications for the guidance effect, as the next section will detail.

LEARNING AND THE GUIDANCE EFFECT

An analysis of the task can enable identification of the perception-action resources used by a skilled performer in an everyday context, including important informational variables and control parameters (Wilson and Golonka, 2013; Bruineberg and Rietveld, 2014). Sonification systems can then be designed to *highlight* these same useful variables/parameters, rather than to create new parameters—control of which might be independent of how the task would be performed without feedback. The value of highlighting task-intrinsic information lies in the possibility to avoid the guidance effect of augmented feedback. As an example, Ronsse et al. (2011) provided direct sonification of changes in hand-movement direction with a set of two tones. Participants were required to learn a 90° out-of-phase bimanual wrist coordination task in which a two-tone isochronous galloping rhythm was produced by perfect performance. When sonification was withdrawn, participant performance remained stable. In contrast, a second group of participants who had practiced with movement-coupled graphical feedback showed a decline in performance following withdrawal. In this task, sonification preserved the spatio-temporal structure of relevant task-intrinsic events in the perceptual-motor workspace, therefore the information required to control movement was perceivable with

or without feedback. Sonification had acted as a guide for its pickup. Conversely, graphical feedback provided information for the direct control of bimanual phase-relationship; its removal meant that coordination was no longer possible as the required information was absent (see Wilson et al., 2010b). If learning is seen as education of attention (Jacobs and Michaels, 2007), the need to sonify task-intrinsic events to avoid the guidance effect is clear (for similar findings, see Dyer et al., 2017).

KNOWLEDGE AND INFORMATION

Current understanding of augmented feedback and its role in performance enhancement has its foundations in classic studies on knowledge-of-results and knowledge-of-performance (KR/KP) feedback (Adams, 1971). Today, sonification in Psychology is still widely discussed using these terms (Konttinen et al., 2004; Dyer et al., 2015; Sors et al., 2015; Fujii et al., 2016). However, this continuity belies a subtle shift in what these terms have come to mean over time as technology has improved to the point where real-time sonification as KP is possible. In the late twentieth Century, both KP and KR meant explicit, propositional knowledge—typically verbal (or verbalisable) knowledge about movement outcome (KR) or performance (KP). Older reviews of KP/KR research (Adams, 1971; Salmoni et al., 1984) show that motor skill learning was explicitly conceptualized as a knowledge and memory-based, problem-solving task, soluble by the application of explicit knowledge and rules (typically, coach-provided guidance, or scores/graphs of performance and error). The goal was to improve performance by delivering instructions, which could be applied to programming of motor output “intellectually,” i.e., independently of perception and action. Thomas and Thomas (1994) have argued that the traditional knowledge-based approach to motor skill learning underplays the role of selective sensitivity to perceptual information in skilled performance, catering mostly to the earliest “cognitive stage” of learning (Fitts and Posner, 1967).

Today, augmented feedback (including sonification) is often delivered concurrently with movement (Sigrist et al., 2013a), and could be considered something more like augmented information for online, perceptual control of action. However, the older style of thinking about feedback as explicit knowledge is evident in many modern implementations. This thinking manifests in the design of mappings intended to transmit a signal to the learner, which can be said to *contain* knowledge—in a description of current performance relative to an ideal (often a sonified error score). This signal must be parsed by the learner with the application of a remembered mapping rule, and the decoded knowledge applied to update ongoing movement. This bears directly on sonified feedback given the requirement for learner interpretation associated with such rule-based mappings. Time required to interpret the knowledge contained in an auditory error signal puts an intellectual barrier between perception and action, which may not be conducive to fluid, skilful performance. The perception-action approach contends that “knowledge” related to skill is primarily enacted via perception-action engagement with the dynamics of a

task—rather than through the rote application of schemas and rules (for this argument, see Ingold, 2000, 2001; van Dijk et al., 2015). Effenberg and colleagues (Vinken et al., 2013; Effenberg et al., 2016) argue similarly that a “direct” approach to mapping in which sound quality perceptually correlates with the dynamics of ongoing movement is most appropriate for motor skill learning. Learners can learn to use sonic information to perceive movement directly, with no need for cognitive elaboration, as the control of movement is directly related to the sensory consequences of movement (for a related example, see Stienstra et al., 2011). This approach preserves the immediacy of Ecological perception, and “knowledge” emerges from two-way interaction rather than being translated from an incoming coded signal.

ECOLOGICALLY-MEANINGFUL SOUND

A theme in this article has been that sonification researchers interested in motor skill learning should understand that learners are primarily perceivers, with pre-existing skills. Perception of something meaningful, or action-relevant in a sonic experience can be conceptualized as an active listening skill which is related to the sociocultural context of its development (Steenson and Rodger, 2015). It follows then that listeners might already know how to listen and interact with certain sound morphologies, in certain contexts/tasks. The use of sounds which cater to existing bodily skills, such as physical modeling of metallic scraping in a writing task (Danna et al., 2015) constrain the learner’s relation to the task and guide perception and action more effectively than might be possible with more basic pitch mapping (see also Roddy and Furlong, 2014). However, Dubus and Bresin (2013) show that pitch mapping (of a pure tone or the center frequency of filtered noise) remains a common strategy in sonification generally, but also in sonification of motor tasks. Most individuals have little experience using a pure tone for movement coordination, therefore such a mapping may be challenging and require extensive training before it can be used. The use of already-familiar sound morphologies (e.g., melodies, rhythms, sounds of real-life noisy interactions) may produce more “intuitive” feedback systems. What we advocate here is

not a distinction which is sometimes made, between “ecological” sounds of the natural world on the one hand and “artificial,” synthetic sounds on the other. “Meaningfulness,” in an Ecological sense, is defined relative to a perceiver’s experience using the information which a sound source provides.

CONCLUSION

In this paper, we have argued for a perception-action approach to motor skill learning as the basis for understanding the utility of sonification as augmented feedback. If information supports performance, then sonification should highlight task-intrinsic information to counter the guidance effect. A clearer definition of what “information” is can help guide the design of sonified feedback whereby knowledge is a product of interaction rather than transmission (e.g., see Wilson and Golonka, 2013). Lastly, learners have abundant socioculturally-situated listening experience already; it is therefore undershooting the potential of sonification as feedback to rely only on unfamiliar or esoteric sound morphologies like pure tones. To speculate, the common root of these three issues for design may be related to the existence of different frames of reference for the experimenter/designer and the learner. It is more helpful to see the learner as a situated agent with a repertoire of existing perception-action skills than an engine that must apply propositional knowledge to enact a desired change in state.

AUTHOR CONTRIBUTIONS

JD conceived the topic in discussion with PS and MR and also drafted the manuscript. All three authors were involved in redrafting and editing of the manuscript.

FUNDING

This work was funded in part by a grant from the Northern Ireland Department of Employment and learning awarded to the lead author to undertake this research as part of his Ph.D.

REFERENCES

- Adams, J. A. (1971). A closed-loop theory of motor learning. *J. Motor Behav.* 3, 111–150. doi: 10.1080/00222895.1971.10734898
- Boyer, É. O., Bevilacqua, F., Susini, P., and Hanneton, S. (2016). Investigating three types of continuous auditory feedback in visuo-manual tracking. *Exp. Brain Res.* 235, 691–701. doi: 10.1007/s00221-016-4827-x
- Bruineberg, J., and Rietveld, E. (2014). Self-organization, free energy minimization, and optimal grip on a field of affordances. *Front. Hum. Neurosci.* 8:599. doi: 10.3389/fnhum.2014.00599
- Buchanan, J. J., and Wang, C. (2012). Overcoming the guidance effect in motor skill learning: feedback all the time can be beneficial. *Exp. Brain Res.* 219, 305–320. doi: 10.1007/s00221-012-3092-x
- Danna, J., Fontaine, M., Paz-Villagrán, V., Gondre, C., Thoret, E., Aramaki, M., et al. (2015). The effect of real-time auditory feedback on learning new characters. *Hum. Mov. Sci.* 43, 216–228. doi: 10.1016/j.humov.2014.12.002
- Dubus, G., and Bresin, R. (2013). A systematic review of mapping strategies for the sonification of physical quantities. *PLoS ONE* 8:e82491. doi: 10.1371/journal.pone.0082491
- Dyer, J., Stapleton, P., and Rodger, M. W. M. (2015). Sonification as concurrent augmented feedback for motor skill learning and the importance of mapping design. *Open Psychol. J.* 8, 1–11. doi: 10.2174/1874350101508010192
- Dyer, J., Stapleton, P., and Rodger, M. W. M. (2017). Transposing musical skill: sonification of movement as concurrent augmented feedback enhances learning in a bimanual task. *Psychol. Res.* 81, 850–862. doi: 10.1007/s00426-016-0775-0
- Effenberg, A. O., Fehse, U., Schmitz, G., Krueger, B., and Mechling, H. (2016). Movement sonification: effects on motor learning beyond rhythmic adjustments. *Front. Neurosci.* 10:219. doi: 10.3389/fnins.2016.00219
- Fitts, P. M., and Posner, M. I. (1967). *Human Performance*. Belmont, CA: Brooks/Cole Publishing Company.
- Fujii, S., Lulic, T., and Chen, J. L. (2016). More feedback is better than less: learning a novel upper limb joint coordination pattern with augmented auditory feedback. *Front. Neurosci.* 10:251. doi: 10.3389/fnins.2016.00251
- Gibson, E. J. (1969). *Principles of Perceptual Learning and Development*. New York, NY: Appleton-Century-Crofts.

- Gibson, J. J. (1972). "A theory of direct visual perception," in *The Psychology of Knowing*, eds J. Royce and W. Rozenboom (New York, NY: Gordon & Breach), 76–89.
- Gray, R. (2010). Expert baseball batters have greater sensitivity in making swing decisions. *Res. Q. Exerc. Sport* 81, 373–378. doi: 10.5641/027013610X13088600028897
- Huys, R., Cañal-Bruland, R., Hagemann, N., Beek, P. J., Smeeton, N. J., and Williams, A. M. (2009). Global information pickup underpins anticipation of tennis shot direction. *J. Mot. Behav.* 41, 158–171. doi: 10.3200/JMBR.41.2.158-171
- Ingold, T. (2000). *The Perception of the Environment*. Abingdon: Taylor & Francis.
- Ingold, T. (2001). "Beyond art and technology: the Anthropology of Skill," in *Anthropological Perspectives on Technology*, ed H. Schiffer (Albuquerque, NM: University of New Mexico Press), 17–33.
- Jacobs, D. M., and Michaels, C. F. (2007). Direct learning. *Ecol. Psychol.* 19, 321–349. doi: 10.1080/10407410701432337
- Kennedy, J. M. (1985). Convergence principle in blind people's pointing. *Int. J. Rehabil. Res.* 8, 207–210.
- Kontinen, N., Mononen, K., Viitasalo, J. T., and Mets, T. (2004). The effects of augmented auditory feedback on psychomotor skill learning in precision shooting. *J. Sport Exerc. Psychol.* 26, 306–316. doi: 10.1123/jsep.26.2.306
- Maslovat, D., Brunke, K. M., Chua, R., and Franks, I. M. (2009). Feedback effects on learning a novel bimanual coordination pattern: support for the guidance hypothesis. *J. Mot. Behav.* 41, 45–54. doi: 10.1080/00222895.2009.10125923
- Newell, K. M., McDonald, P. V., and Kugler, P. N. (1991). "The perceptual-motor workspace and the acquisition of skill," in *Tutorials in Motor Neuroscience*, eds J. Requin and G. E. Stelmach (Dordrecht: Springer), 95–108.
- Oscari, F., Secoli, R., Avanzini, F., Rosati, G., and Reinkensmeyer, D. J. (2012). Substituting auditory for visual feedback to adapt to altered dynamic and kinematic environments during reaching. *Exp. Brain Res.* 221, 33–41. doi: 10.1007/s00221-012-3144-2
- Park, J. H., Shea, C. H., and Wright, D. L. (2000). Reduced-frequency concurrent and terminal feedback: a test of the guidance hypothesis. *J. Mot. Behav.* 32, 287–296. doi: 10.1080/00222890009601379
- Roddy, S., and Furlong, D. (2014). Embodied aesthetics in auditory display. *Organ. Sound* 19, 70–77. doi: 10.1017/S1355771813000423
- Ronsse, R., Puttemans, V., Coxon, J. P., Goble, D. J., Wagemans, J., Wenderoth, N., et al. (2011). Motor learning with augmented feedback: modality-dependent behavioral and neural consequences. *Cereb. Cortex* 21, 1283–1294. doi: 10.1093/cercor/bhq209
- Rosati, G., Oscari, F., Spagnol, S., Avanzini, F., and Masiero, S. (2012). Effect of task-related continuous auditory feedback during learning of tracking motion exercises. *J. Neuroeng. Rehabil.* 9, 1–13. doi: 10.1186/1743-0003-9-79
- Runeson, S. (1977). On the possibility of "smart" perceptual mechanisms. *Scand. J. Psychol.* 18, 172–179. doi: 10.1111/j.1467-9450.1977.tb00274.x
- Salmoni, A. W., Schmidt, R. A., and Walter, C. B. (1984). Knowledge of results and motor learning: a review and critical reappraisal. *Psychol. Bull.* 95, 355–386.
- Schmitz, G., and Bock, O. (2014). A comparison of sensorimotor adaptation in the visual and in the auditory modality. *PLoS ONE* 9:e107834. doi: 10.1371/journal.pone.0107834
- Sigrist, R., Rauter, G., Riener, R., and Wolf, P. (2013a). Augmented visual, auditory, haptic, and multimodal feedback in motor learning: a review. *Psychon. Bull. Rev.* 20, 21–53. doi: 10.3758/s13423-012-0333-8
- Sigrist, R., Rauter, G., Riener, R., and Wolf, P. (2013b). Terminal feedback outperforms concurrent visual, auditory, and haptic feedback in learning a complex rowing-type task. *J. Mot. Behav.* 45, 455–472. doi: 10.1080/00222895.2013.826169
- Sors, F., Murgia, M., Santoro, I., and Agostini, T. (2015). Audio-Based Interventions in Sport. *Open Psychol. J.* 8, 212–219. doi: 10.2174/1874350101508010212
- Stenson, C., and Rodger, M. W. M. (2015). Bringing sounds into use: thinking of sounds as materials and a sketch of auditory affordances. *Open Psychol. J.* 8(Suppl 3), 174–182. doi: 10.2174/1874350101508010174
- Stienstra, J., Overbeeke, K., and Wensveen, S. (2011). "Embodying complexity through movement sonification," in *Proceedings of the 9th ACM SIGCHI Italian Chapter International Conference on Computer-Human Interaction: Facing Complexity* (New York, NY: ACM Press), 39–44.
- Stoffregen, T. A., and Bardy, B. G. (2001). On specification and the senses. *Behav. Brain Sci.* 24, 195–213. doi: 10.1017/S0140525X01003946
- Thomas, T., and Thomas, J. R. (1994). Developing expertise in sport: the relation of knowledge and performance. *Int. J. Sport Psychol.* 25, 295–312.
- Turvey, M. T., Shaw, R. E., Reed, E. S., and Mace, W. M. (1981). Ecological laws of perceiving and acting: in reply to Fodor and Pylyshyn. *Cognition* 9, 237–304. doi: 10.1016/0010-0277(81)90002-0
- van Dijk, L., Withagen, R., and Bongers, R. M. (2015). Information without content: a Gibsonian reply to enactivists' worries. *Cognition* 134, 210–214. doi: 10.1016/j.cognition.2014.10.012
- Vinken, P. M., Kröger, D., Fehse, U., Schmitz, G., Brock, H., and Effenberg, A. O. (2013). Auditory coding of human movement kinematics. *Multisens. Res.* 26, 533–552. doi: 10.1163/22134808-00002435
- Warren, W. H. (2006). The dynamics of perception and action. *Psychol. Rev.* 113, 358–389. doi: 10.1037/0033-295X.113.2.358
- Wilson, A. D., and Golonka, S. (2013). Embodied cognition is not what you think it is. *Front. Psychol.* 4:58. doi: 10.3389/fpsyg.2013.00058
- Wilson, A. D., Snapp-Childs, W., and Bingham, G. P. (2010a). Perceptual learning immediately yields new stable motor coordination. *J. Exp. Psychol. Hum. Percept. Perform.* 36, 1508–1514. doi: 10.1037/a0020412
- Wilson, A. D., Snapp-Childs, W., Coats, R., and Bingham, G. P. (2010b). Learning a coordinated rhythmic movement with task-appropriate coordination feedback. *Exp. Brain Res.* 205, 513–520. doi: 10.1007/s00221-010-2388-y

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.

Copyright © 2017 Dyer, Stapleton and Rodger. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



A Novel Sonification Approach to Support the Diagnosis of Alzheimer's Dementia

Letizia Gionfrida^{1,2*} and Agnieszka Roginska^{3,4}

¹Centre of Performance Science, Royal College of Music, London, United Kingdom, ²BICV Group, Bioengineering Department, Faculty of Engineering, Imperial College London, London, United Kingdom, ³Music and Audio Research Laboratory, Music Technology, New York University, New York, NY, United States, ⁴Division of Nuclear Medicine, Department of Radiology, NYU School of Medicine, New York, NY, United States

OPEN ACCESS

Edited by:

Diego Minicacchi,
University of Florence, Italy

Reviewed by:

Karl Egger,
Universitätsklinikum Freiburg,
Germany

Giuseppe Curcio,
University of L'Aquila, Italy

*Correspondence:

Letizia Gionfrida
letizia.gionfrida@rcm.ac.uk,
l.gionfrida17@imperial.ac.uk

Specialty section:

This article was submitted
to Neuroprosthetics,
a section of the journal
Frontiers in Neurology

Received: 04 September 2017

Accepted: 16 November 2017

Published: 07 December 2017

Citation:

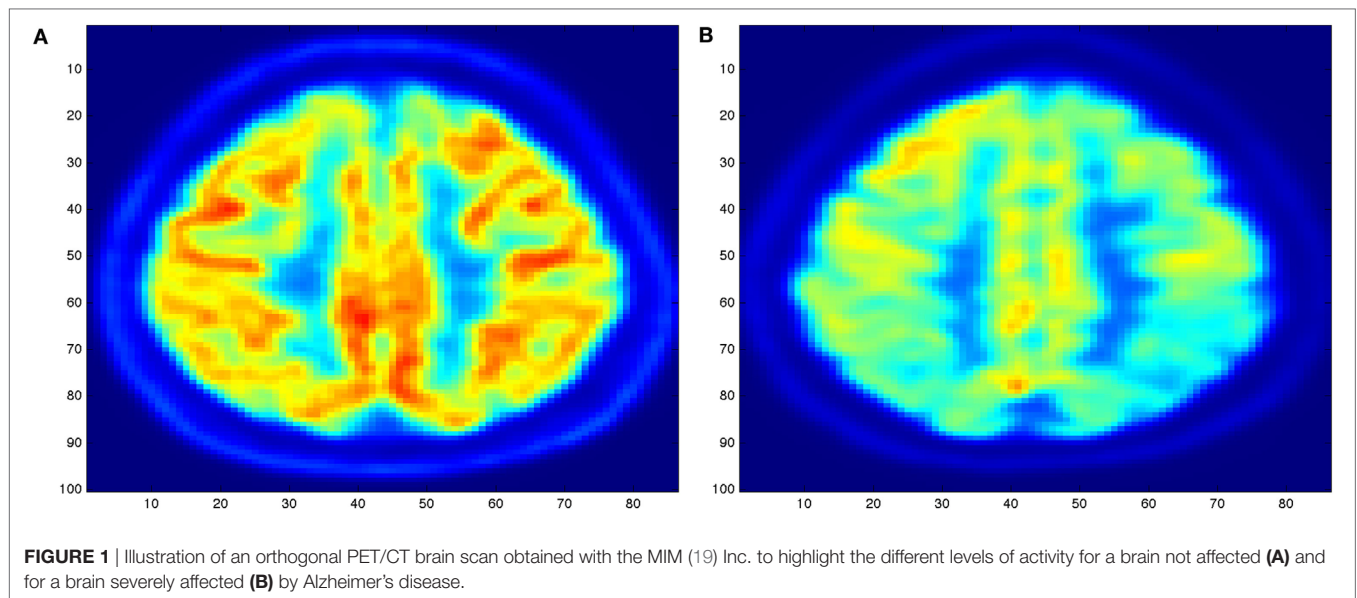
Gionfrida L and Roginska A (2017)
A Novel Sonification Approach
to Support the Diagnosis of
Alzheimer's Dementia.
Front. Neurol. 8:647.
doi: 10.3389/fneur.2017.00647

Alzheimer's disease is the most common neurodegenerative form of dementia that steadily worsens and eventually leads to death. Its set of symptoms include loss of cognitive function and memory decline. Structural and functional imaging methods such as CT, MRI, and PET scans play an essential role in the diagnosis process, being able to identify specific areas of cerebral damages. While the accuracy of these imaging techniques increases over time, the severity assessment of dementia remains challenging and susceptible to cognitive and perceptual errors due to intra-reader variability among physicians. Doctors have not agreed upon standardized measurement of cell loss used to specifically diagnose dementia among individuals. These limitations have led researchers to look for supportive diagnosis tools to enhance the spectrum of diseases characteristics and peculiarities. Here is presented a supportive auditory tool to aid in diagnosing patients with different levels of Alzheimer's. This tool introduces an audible parameter mapped upon three different brain's lobes. The motivating force behind this supportive auditory technique arise from the fact that AD is distinguished by a decrease of the metabolic activity (hypometabolism) in the parietal and temporal lobes of the brain. The diagnosis is then performed by comparing metabolic activity of the affected lobes to the metabolic activity of other lobes that are not generally affected by AD (i.e., sensorimotor cortex). Results from the diagnosis process compared with the ground truth show that physicians were able to categorize different levels of AD using the sonification generated in this study with higher accuracy than using a standard diagnosis procedure, based on the visualization alone.

Keywords: sonification, medical imaging, PET scan, Alzheimer's dementia, computer-aided diagnosis

1. INTRODUCTION

The word dementia indicates a collective group of neurodegenerative disorders characterized by impairment of cognitive and functional abilities together with behavioral symptoms (1). It is also characterized by loss of neurons and synapses, atrophy of brain regions (particularly in the temporal and parietal lobes), and metabolic dysfunction of brain activity. Among different forms of dementia, Alzheimer's disease is a progressive neurodegenerative disease that accounts for around 80% of dementia cases (2), and it is characterized by difficulties with thinking and, in more severe cases, loss of self-awareness. This is reflected in a decreased metabolic activity (hypometabolism) in the parietal and temporal lobes of the brain (3), while more severe cases also present hypometabolism in the frontal lobe of the brain (4).



The diagnosis of AD (2) can definitely be diagnosed after death, with a histopathologic confirmation, but it can be assessed using neuroimaging techniques. Alzheimer's is also diagnosed based on the person's medical history, laboratory assessments, and behavioral observations. In particular, medical neuroimaging refers to several different techniques that are used to produce different visualizations of human bodies, tissues, and organs. A standard AD procedure to exclude other cerebral pathology or subtypes of dementia often includes functional and structural imaging techniques as magnetic resonance imaging (MRI) and computed tomography (CT), but such techniques are not able to detect the stages of the disease, especially due to volume loss is not apparent early in the course of the disease (5). In addition to structural imaging, molecular imaging techniques as positron emission tomography (PET) helps in differentiating dementia syndromes for their peculiarity of reflecting cerebral metabolic rates of the brain. **Figure 1** shows an orthogonal PET/CT brain scan with a brain not affected and a brain affected by AD. In addition hybrid nuclear/structural imaging analysis, as PET-MRI and PET-CT, have shown the potential to further improve image registration and reduced radiation exposure (6). The advantages of these hybrid techniques are the high resolution and sensitivity, and the simultaneous acquisition of different brain areas (7).

These methods are reliable and advanced during the diagnosis process, but there is still large inconsistency in the diagnosis process among physicians upon different visual inspections provided by these medical imaging (8). The main problem is that the diagnosis is highly dependent on subjective judgment (9) and clinicians continue to have difficulties when the visual analysis leads to imperceptible differences between health and disease (8).

These problems have led physicians and scientists to investigate for assistive diagnosis procedures that together with standard visualizations from imaging techniques can convey information in a supportive way that can aid the diagnosis process (10).

Several studies have explored the possibility of aided diagnosis using auditory feedback in medical domains.

The technique that involves non-verbal information that can be transformed into audio feedback and used to facilitate data communication and/or interpretation is called sonification (11). A broad variety of sonification designs have been explored in the past two decades; work such as Walker and Kramer (12) and Walker and Mauney (13), have been successfully employed for esthetic representations, while for Polli (14, 15) have used sonification to describe atmospherically/geographical data. Particularly, expression of medical data into audio domains as supplementary tool has been largely explored, i.e., to inspect multivariate medical data as EEG data (16), or as a supporting method for medical imaging in the diagnosis process. Starting from the assumption that radiologists have to cost their efforts to investigate these medical images (17) used a large amount of MR images providing pieces of evidence that sonification helps in attention rousing and fatigue reduction during medical imaging diagnosis.

In spite of the success in applying auditory information to the previous mentioned datasets, no work has been conducted on the use of sonification to facilitate AD diagnosis process using PET/CT scan. In light of the challenges in diagnosis at assessing Alzheimer's diagnosis in living patients, we explore a novel sonification tool that enhances the diagnosis of dementia using a PET/CT scan. In this study, we present a sonification tool that improves diagnosis accuracy. We investigate the diagnosis of different stages of AD using a PET/CT scan and how the use of a sonification method added upon the visual inspection can increase accuracy in distinguish between brains of different levels of AD.

2. MATERIALS AND METHODS

The sonification technique to enhance the diagnosis of dementia used a hybrid PET/CT scan for the acquisition process of the medical dataset. Once acquired, the medical dataset was pre-processed

by medical physicians to standardize the output to be analyzed. After the pre-processing session, the dataset was analyzed using a tool that allowed the sonification of the medical dataset.

The sonification technique implemented in this tool is called Triple-Tone Sonification (TTS). It aims to enhance explicitness in the diagnosis process taking advantages from the fact that the human ear, as investigated in Ref. (18) article, has a distinct advantage in hearing two close frequencies beating against each other in a phenomenon known as frequency beating. Following the TTS technique, the metabolic activity of the segmented lobes was mapped to the frequency that creates the tone. Particularly, brain lobes detected with a PET/CT scan was segmented, using the MIM (19), into three key areas: frontal lobe, parietal lobes, which are highly affected by AD (4), and sensorimotor cortex (SMC), that usually remains unaffected (2). Each region was then mapped to a different audible frequency and the interaction of these tones' frequencies resulted in beating patterns. From the acquisition to the sonification, the pre-processing to the analysis, the entire methodology of the PET/CT scan for the 32 de-identified human brains is explained in the following subsections.

2.1. Dataset and PET/CT Pre-processing

The dataset used to evaluate the performances of the TTS was obtained from the radiology department of the New York University Langone Medical Center. It consisted of 32 de-identified PET/CT scans of human brains that were diagnosed with different stages of Alzheimer's dementia. The entire dataset was acquired using the vendor-neural (19) medical software that allows the visualizations of general nuclear medicine imaging. The collection of 32 brain scans was distributed as of eight brain scans in each of four different categories of AD. These categories included eight subjects' brain not influenced by dementia, eight subjects mildly affected by Alzheimer's, eight brains with moderate level of AD, and eight subjects' brain severely affected by Alzheimer's dementia.

Given the nature of the image-based diagnosis process, as common particle in biomedical investigation, the gold standard (such as the presence of AD and the eventual level of severity) was provided by the nuclear medicine section chief of the NYU Langone Medical Center Dr. Kent P. Friedman, based on his medical and professional experience. The assessment of presence of AD and eventual stage of severity for each subject was made after at minimum of 6 months investigation for each patient in the

dataset. Dr. Friedman, utilizing the visualization obtained with the MIM (19) software from the PET/CT scan, together with the medical history, such as age, sex, previous analysis, and blood test, for the panel of the 32 patients, was capable of providing the ground truth for each subject utilized in the entire study.

As standard medical procedure, using the MIM (19) software, all the datasets were spatially warped to a standard brain model so that all the 32 brain scans used in the testing were spatially consistent (20). The size of each dataset was $86 \times 100 \times 86$ voxels, with each voxel corresponding to a physical size $2 \text{ mm} \times 2 \text{ mm} \times 2 \text{ mm}$. These adaptations made also possible to sonify the same two-dimensional subset of data points within each three-dimensional dataset by simply choosing the same lateral slice each time for sonification. Hence, for this study the entire dataset was not utilized, but one slice per brain with its 3D projection was pulled out to be sonified. The preferable sliced chosen for the sonification was the 30th slice (from the top) that passes through representative regions of the frontal lobe, parietal lobe, and sensorimotor cortex. Consequently, for each patient, the 30th slice was selected to be the lateral slice utilized to be sonified for the purpose for this study. This corresponds to a lateral slice through the brain between 58 and 60 mm from the top edge of the dataset with size of 86×100 voxels. The second subset chosen for the study was a 3D projection view of the brain generated using the MIM (19) software, as illustrated in **Figure 2A** and in **Figure 2B**.

In a standard medical diagnosis physicians inspect the view of the brain generated with the MIM (19) to assess the overall picture regarding the diagnosis of a brain. Since does not allow to export the 3D projection view in the DICOM (Digital Imaging and Communication in Medicine) file format, screen captures of a gray scale representation of the 3D projection view, as shown in **Figure 2B**, were brought into play as base datasets for the sonification. As the 3D projection datasets are screen captures, they contain extraneous elements, such as axes, labels, and unnecessary white space. All extraneous elements were removed before being saved as a dataset, this made possible to use our sonification approach on this dataset.

Boundary coordinates selected using the MIM (19) are shown **Figure 6** and are used for lobe segmentation performed using the TTS. The visualization medical software performs its own automatic lobe segmentation of the spatially normalized brain datasets and provides segmentation information to the user in

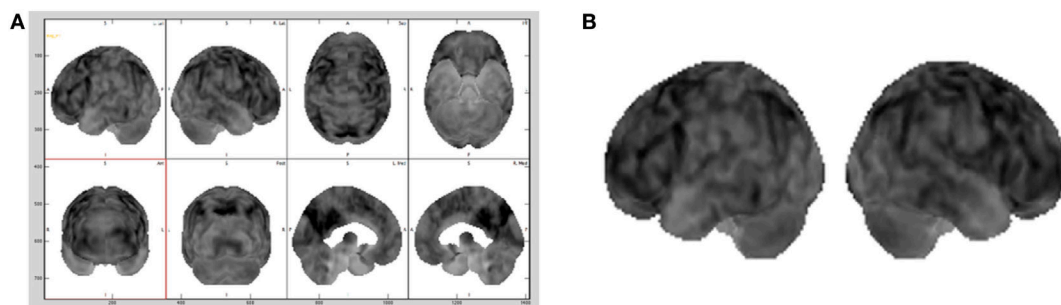


FIGURE 2 | Different brains visualization using the MIM (19) software with a typical Advanced visualization (A) and its 3D projection (B) extracted.

the form of the DICOM standard RTSTRUCT files. Once these coordinates are known, the datasets in output from the MIM (19), such as DICOM standard RTSTRUCT file format, goes in input into the data analysis TTS tool that performed the frontal, parietal, and SMC lobes segmentation. In the TTS tool, at this stage, the contours were approximated to straight lines and the results for lateral slices and 3D projections are shown in **Figure 3A,B**.

In both the lateral slice and 3D projection view remained some irrelevant data points. In PET/CT scans, data points can be irrelevant for multiple reasons, as outlined by Piper (20) and Marcus et al. (21). Some data points in the lateral slice can lie outside the actual brain area, while some other irrelevant data points consist of those that lie within the brain, but are medically irrelevant (20). The brain can for the most part be divided into white matter and gray matter (1, 22). The relevant metabolic activity for the diagnosis of AD is that of the gray matter of the brain (8). Hence, the white matter content is medically irrelevant and should not be included as part of the sonification. Therefore, all voxels whose intensity fell below a certain threshold were masked from being sonified. This threshold was set to 45% of the maximum allowable intensity, considering the bit-depth of the dataset. Hence, for datasets with a bit-depth of 15 bits, the masking threshold is set to 14,745.15 out of a maximum allowable value of 32,767.

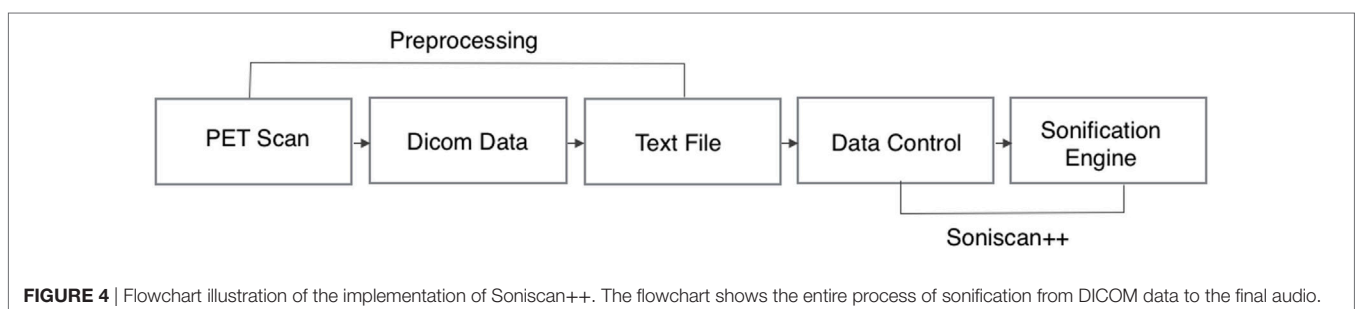
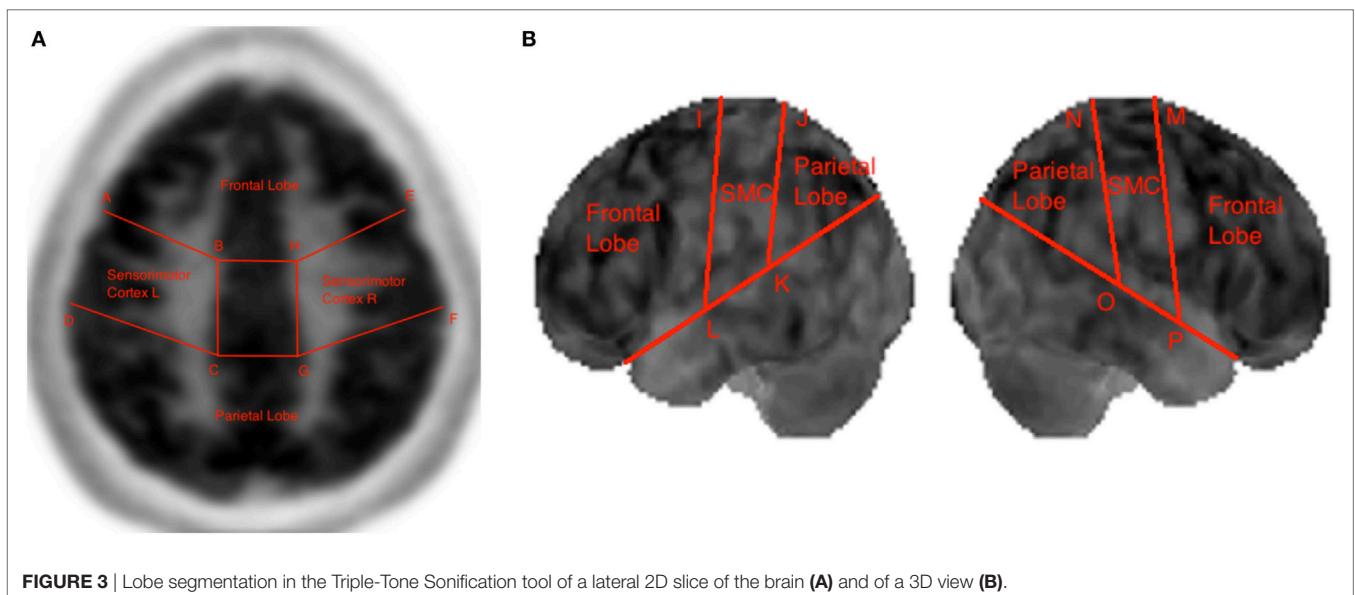
In this data preparation, as usual procedure in diagnosis, physicians using the MIM (19), selected the two-dimensional subset

of data to be sonified and the boundary coordinates for the lobes segmentation. Also as standard research medical operating routine, during this first phase of preprocessing, the multi-paradigm numerical computing environment (23) was also utilized to read the DICOM datasets in output for the MIM software through its Image Processing Toolbox and to translate and store them into arrays. All of brain scans were stored in the standard .mat format together with a text file format. The boundary coordinates for the lobes segmentation were stored in a separate text file. After the pre-processing, the following section will discuss the data analysis tool developed in this study.

2.2. Soniscan++

Once the medical dataset was pre-processed, it was then analyzed using a tool developed at the NYU Music Technology for the purpose of this research, named Soniscan++. This tool was introduced to translate raw medical images, stored as text files, into sound in a standardized manner that allowed reproducibility and rigor into the analysis process. The data analysis tool was developed using the object oriented programming language C++, given real-time sound synthesis speed and versatility of this programming language. The main purpose of this tool was to sonify the spatially normalized medical dataset utilizing the TTS.

The complete workflow of the process that translates brain scans into sound is summarized in **Figure 4**. As shown,



SoniScan++ consists of two major blocks: the Data Control block and Sonification Engine block. Soniscan++ also output files in a.scd format so they could be easily played by a graphical user interface developed using Supercollider (24) shown in **Figure 5**.

The Data Control block is responsible for reading the dataset in text file format as three-dimensional array, performing masking of the data, and extracting the 30th slice to be sonified. As output from the medical device the text file consists of several lines of text. As illustrated in **Figure 6**, each line of text refers to a single voxel, corresponding to the voxel's x-location, y-location, z-location, and value.

Once the slice and its coordinates are extracted, the Sonification Engine block is responsible of segmenting the lobes into the three regions to be sonified, based on the given boundary coordinate points. It then removes of all irrelevant data points from the sub-datasets, using the TTS to map the difference in metabolic activity between the lobes of interest (frontal and parietal) and the reference lobe (SMC) to an easily perceivable auditory parameter. Finally, it outputs the frequencies' differences for each slice and

writes all the output into an audible format. During this stage, the Sonification Engine is also responsible for extracting randomly from the dataset the 30th slice for one participant in each category for the training session of the evaluation and avoiding that the randomly selected slices is then used for the testing phase in the evaluation process, that will be following mentioned.

The TTS technique, implemented in Sonification Engine block, assigns an oscillator to each of the three brain lobes of the lateral slices selected. As illustrated in **Figure 7**, the TTS assigns, for each slice selected, a triangular wave oscillator to the frontal lobe, parietal lobe, and sensorimotor cortex. The frequencies of these oscillators are mapped to the average metabolic activity of these regions. Specifically, to the sensorimotor cortex is given a base frequency of 440 Hz (A above middle C), and the other two lobes are deviations from that frequency. These deviation values are determined by how far the metabolic regional activity deviates from the sensorimotor cortex. Hence, differences in metabolic activity, reflecting the progress of the AD in between lobes, result in different frequencies for the oscillators. The more pronounced

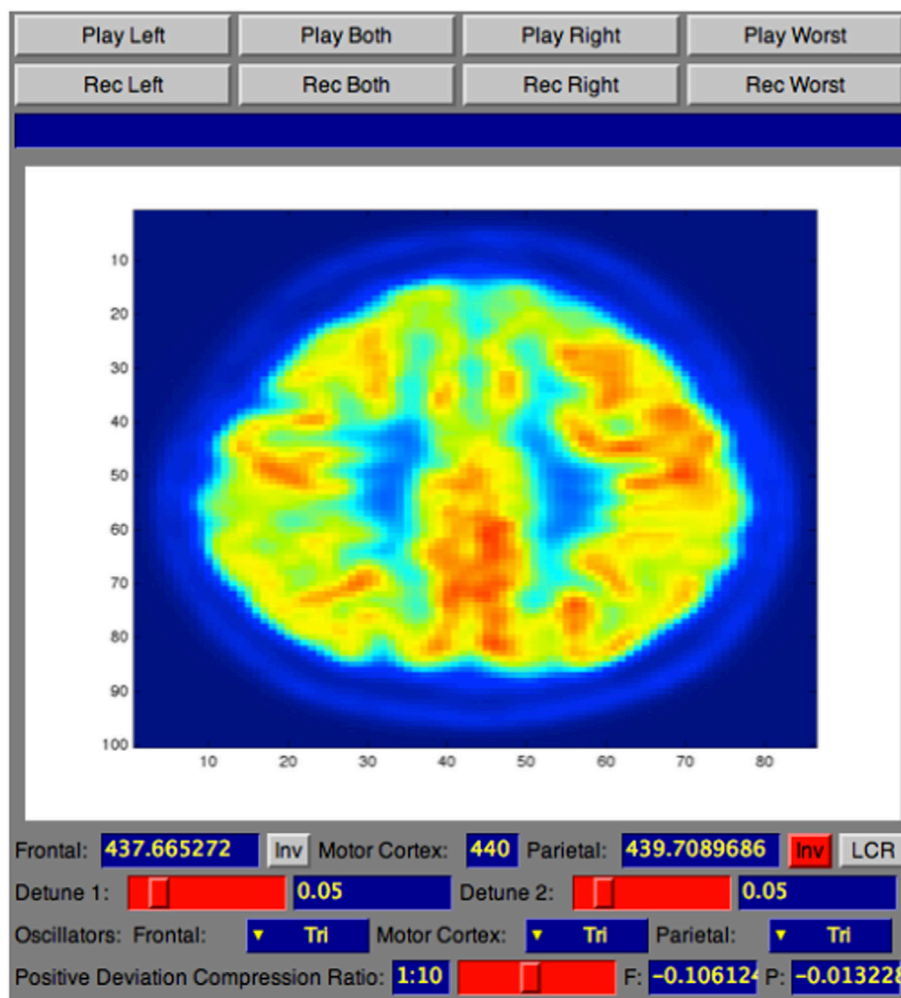


FIGURE 5 | Screenshot of SuperCollider interface for sonification playback and recording.

the difference in metabolic activity between lobes, the higher the level of AD. Thus, the more pronounced the difference between the three frequencies, the faster and more complex the beating pattern. The goal is to find different levels of beatings to indicate varying degrees of AD.

To control the range of frequencies' deviations given the range of voxels' deviation, a scale factor named detune factor was used for different level of AD. The frequencies corresponding to the

lobes under inspection are detuned from the default frequency according to the SD of the voxels' average intensities of frontal and parietal from the voxels' average intensity of the reference lobe (sensorimotor cortex). To ensure that the audible parameter would not collapse to two separate beating tones when there are anomalies of equal deviation in frontal and parietal, from the default frequency, the frequency of the frontal lobe is forced to a positive deviation while the frequency of the parietal lobe is forced to a negative deviation. This difference in tones between lobes creates a beating pattern, which would be perceptible for the listener to be heard. The values for the detune factor are shown in **Table 1**.

Therefore, given the detune factor, the frequency of the frontal (f_{FL}) and parietal (f_{PL}) lobes were determined as follows:

$$f_{FL} = f_{default} * \left(1 + DF * \left| \frac{av_{FL} - av_{SMC}}{av_{SMC}} \right| \right) = f_{default} * (1 + DF * |\Delta_{FL}|)$$

$$f_{PL} = f_{default} * \left(1 + DF * \left| \frac{av_{PL} - av_{SMC}}{av_{SMC}} \right| \right) = f_{default} * (1 + DF * |\Delta_{PL}|).$$

In the equations above, the relative deviation of average intensity (av_{FL} and av_{PL}) with respect to the average intensity of the sensorimotor cortex (av_{SMC}) are linearly mapped to the relative deviation of the oscillator frequency with respect to the oscillator's base frequency ($f_{default}$) through the detune factor (DF) coefficient.

In conclusion, these differences in oscillation between the three oscillators result in beating patterns that will be heard from the two controlled group during the evaluation process. [You can listen to the four sonified sound samples for a subject not affected

```
49,100, 46, 4439
50,100, 46, 4197
51,100, 46, 4063
52,100, 46, 3999
53,100, 46, 3846
54,100, 46, 3560
55,100, 46, 3147
56,100, 46, 2684
57,100, 46, 2279
58,100, 46, 1865
59,100, 46, 1371
60,100, 46, 896
61,100, 46, 534
62,100, 46, 303
63,100, 46, 167
64,100, 46, 92
65,100, 46, 54
66,100, 46, 33
67,100, 46, 21
68,100, 46, 14
```

FIGURE 6 | Excerpt of text format file from the pre-processing step containing x, y, and z voxels' location, and value for the selected slice.

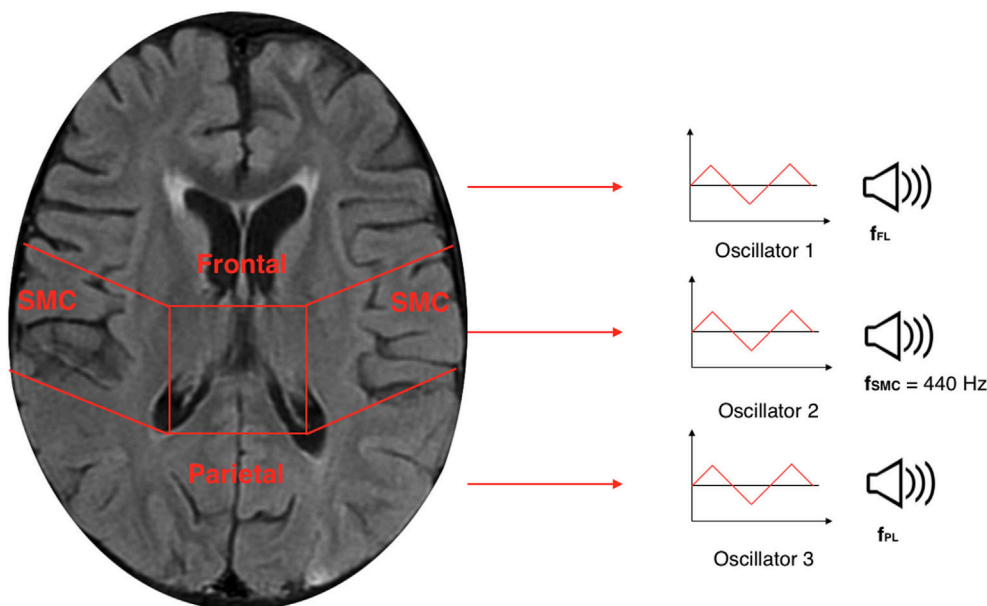


FIGURE 7 | Illustration of a 30th brain slice segmented to frontal, parietal, and sensorimotor cortex to which is assigned a triangular wave oscillator to each of the three lobes.

TABLE 1 | List of detune factor values for normal brain, mild AD, moderate AD, and severe AD.

Brain type	Detune factor
Normal brain	0.05
Mild AD	0.10
Moderate AD	0.15
Severe AD	0.20

by AD (Audio S1 in Supplementary Material), for a patient mildly affected by AD (Audio S2 in Supplementary Material), for a subject moderately affected by AD (Audio S3 in Supplementary Material), and for a subject severely affected by AD (Audio S4 in Supplementary Material).]

3. EVALUATION

The implementation features of TTS developed in the Soniscan++ tool were analyzed in the previous section. The TSS produced differences in oscillators resulting in hearable beating patterns that can tackle differences across the four levels of AD cases presented in this study. The technique was validated among two groups of listeners against the ground truth provided by Dr. Friedman and evaluated on how effectively it allowed those listeners to distinguish between brains of different levels of AD with higher accuracy that using the visualizations only.

The first round of testing was presented to trained musical ears of audio professionals, while the second round of testing to physicians in the field of radiology. The purpose of the first round was to validate the effectiveness of the TTS in accurate categorizations of the beating pattern. The main goal of the second round testing was to determine the accuracy and intra-reader consistency in the diagnosis process. Both testing groups forerun by a training session. The purpose of this warming up session was to let listeners familiarize with the sonification tool before the actual analysis.

During the training session were randomly selected four subjects for each of the four categories of AD from the dataset. These four slices, one for each of the subject selected, were extracted from the Sonification Engine implemented in the Soniscan++ and not used in the testing process. For each subject, it was used the orthogonal 2-dimensional view (basic visual) of the 30th slice and the related 3-dimensional view (advanced visual), and basic and advanced visualizations were played always one after the other for the selected slice. The selected training samples, for basic and advanced visuals, were played for each of the four categories including normal, mild, moderate, and severe. These sonifications were played for 30 s each, with approximately 1 min break one after the other, while participants were looking at the lateral and advanced visualizations related to that particular sonification. The order at which they were presented was always structured with the basic visualization first and the advanced visualization following the basic for each of the four AD level, presented from normal to severe. This training session was performed to introduce them the sound related to each of the four AD categories.

Following the training session, a testing session was presented to the two groups of listeners. Also during this session,

2-dimensional (basic) and 3-dimensional (advanced) visualizations were utilized one after the other for the selected slices in each of the four AD categories. Only this time, the order at which these slices were presented, with respect to the AD level of disorder, was randomized in order to test the sonification tool added to the basic and advanced visualizations. All the sonifications were forerun by a standard visualization without sonification added. This was done to mimic a diagnosis scenario. Overall, the order at which was presented the dataset to the two groups of listeners was divided into four steps. During the first round, it was presented the two visualizations, basic followed by advanced, without any sonification. During the second round for each of the two visualizations was added a sonification related to the particular visualization selected. In the following subsection, the two rounds are presented one after the other, as it was during the evaluation procedure.

3.1. Testing: Round One

During this round, testing occurred over four sessions, including basic and advanced visualizations with and without sonification. Here, the evaluation was performed into two sections, namely “coarse categorization” and “fine categorization.” For the “coarse categorization,” four classes, numbered from one to four, were chosen to correspond to the diagnosis of four different levels of AD (normal, mild, moderate, severe). In the “fine categorization,” participants were instead asked to classify each sonification into a seven step categorization, numbered 1, 1.5, 2, 2.5, 3, 3.5, and 4. In this second case, subjects were instructed to rate both cases that aligned with the training cases to the integer-valued categories (1, 2, 3, and 4), corresponding to AD levels, and cases that may be interpreted as lying between training cases as the fractional-value categories (1.5, 2.5, and 3.5). Each subject during each section had to categorize the presented sonification into one of the four coarse numbers as well as into one of seven fine categories. This was done to investigate whether the TTS provided a finer gradation that improved diagnosis accuracy in distinguishing more accurate or consistent results.

First, a basic visualization of the 30th orthogonal slice of the brain was presented, randomly selected from the dataset. Then, for the same subject, the Sonification Engine implemented in the Soniscan++ selected the relative 3-dimensional visualization. These two sessions were initially shown for the entire dataset without any sonification added and it was asked to the testers to indicate a category one of the four coarse numbers as well as into one of seven fine categories for the entire dataset. After 10 min, they completed the first session, the second section, in which the sonification was added for basic and advanced visual was added. The Sonification Engine randomly presented the same subjects and the same visualizations, only this time 30-s sonification played along with basic and advanced visuals for all the subjects in the dataset.

This subjective listening test was presented to five testers, all graduate students, and faculty members of the Music Technology program at the NYU Steinhardt. This group was chosen to first evaluate the performances of the TTS technique because of their musically trained ears, providing a technical validation before proceeding with the evaluation of radiologists during the second round of testing. Results are reported in the following section.

3.2. Testing: Round Two

As for the first round, also the second round testing took place over two sessions. During these sessions, the Sonification Engine randomly selected a two-dimensional lateral slice and the related three-dimensional projection of the same brain, for each of the 32 de-identified brains in the dataset. In the first session, basic and advanced visualizations were presented without sonification; while in the second session, the sonification was added for both visualizations and played along for each subject in the dataset. Here, the evaluation was divided into four clusters that corresponded to the four different levels of brain in the dataset, including normal, mild, moderate, and severe. All the sessions were set up to mimic the diagnosis process that the radiologist would normally undertake. Therefore, each radiologist during each of the two sessions performed the diagnosis looking at the left hemispheres and right hemispheres of the frontal and the parietal lobes assessing a level of disease with respect to brain presented.

This second round of testing was performed by two radiologists from the NYU Langone Medical Center, one highly experienced radiologist and one less experienced radiologist. The physicians were given no information about the test case, the patient's gender, the age, or the medical history. The first session included the standard visualization of the entire dataset without any sonification. After 10 min from the first session, the second stage tested the TTS performance in enhancing distinguishability among different level of AD. Each session included diagnoses of the same 32 de-identified PET scan cases. The order of presenting the sonification was randomized across diagnoses. Results are shown and discussed in the following session.

4. RESULTS

4.1. Results: Round One

This first group of participants consisting of trained ears of musical professionals aimed to classify the results in “coarse categorization” and “fine categorization” for each of the four categories of AD level, including normal, mild, moderate, and severe. The accuracy of response in the “coarse categorization” responses were matched against the ground truth, while for the “fine categorization” responses were either matched against the ground truth or lied at a distance of 0.5 from it.

Results from **Table 2** present the percentage of accuracies for each participant in discriminating between cases using the “coarse categorization,” and results in **Table 3** illustrate the same accuracy in distinguishing between different AD levels using the “fine categorization.” In the case of coarse categorization, the responses to a pair of duplicate test cases are said to be consistent if both cases were given the same response by the participant. In the case of fine consistency, the responses to a pair of test cases are said to be consistent if both cases were given responses that differ by no more than 0.5. A side-by-side comparison of participant accuracies is given in **Table 4** for coarse and fine sections. In the case of coarse categorization, mimicking the diagnosis procedure of AD, participant displayed an overall accuracy of 87%.

4.2. Results: Round Two

The previous round validated the technique providing its capability to yield accurate categorizations of the beating patterns. However, that testing round was done on the trained musical ears of audio professionals. The goal of the second round was to determine the validity of the technique on radiologists.

At the end of the final testing session, physicians were given a survey with questions. The purpose of the survey was to investigate if they found the sonification helpful in the diagnoses process. The questionnaire is shown in **Table 5** and their answers were

TABLE 2 | Accuracy of participants in differentiating using the “coarse categorization” across four categories of Alzheimer’s dementia.

Participants	Cat.1 (%)	Cat.2 (%)	Cat.3 (%)	Cat.4 (%)
P1	100	88	81	88
P2	88	94	38	94
P3	94	100	81	100
P4	100	94	75	100
P5	94	94	44	100

TABLE 3 | Accuracy of participants in differentiating using the “fine categorization” across four categories of Alzheimer’s dementia.

Participants	Cat.1 (%)	Cat.2 (%)	Cat.3 (%)	Cat.4 (%)
P1	88	88	81	88
P2	100	100	56	100
P3	100	100	100	100
P4	94	100	88	100
P5	100	100	69	100

TABLE 4 | Comparison of participants’ accuracy in discriminating different level of dementia using coarse and fine categorizations.

Participant	Coarse Categ. (%)	Fine Categ. (%)
P1	89	86
P2	78	89
P3	94	100
P4	92	95
P5	83	92

TABLE 5 | Physician questionnaire after the training and testing to evaluate the performances of the “Triple-tone” sonification technique.

Questions	Phys. high exp.	Phys. low exp.
The sonification was tiring to listen to and induced fatigue	2	1
The sonification was pleasant	2	2
The sonification provided additional information that the visual display did not	3	5
The sonification was helpful in discerning between	–	–
(a) normal and mild cases	5	4
(c) mild and moderate cases	5	4
(e) mild and severe cases	4	4
The sonification made me more certain about my diagnosis	4	4

given on a scale from 1 to 5, where 1 = strongly disagree, 2 = disagree, 3 = somewhere in the middle, 4 = agree, and 5 = strongly agree. As shown in **Table 5**, the radiologists found the sonification technique although not pleasant, helpful in the diagnosis process especially when it came to discerning between different levels of AD in solidifying the validity of the overall process.

TABLE 6 | Pearson's correlation for association with the ground truth,

***significant at $p < 0.001$, ** significant at $p < 0.005$, * significant at $p < 0.5$, for Physician 1 (high experienced).

Sess.	LF	LP	RF	RP	W
Physician 1 (high experienced)					
0 Basic Visual	0.64***	0.84***	0.56**	0.8***	0.88***
1 Adv. Visual	0.39*	0.76***	0.37*	0.74***	0.81***
2 Basic Visual + Sonif.	0.69***	0.94***	0.66***	0.94***	0.98***
3 Adv. Visual + Sonif.	0.68***	0.79***	0.67***	0.81***	0.86***
Physician 2 (low experienced)					
0 Basic Visual	0.58***	0.70***	0.54**	0.64***	0.87***
1 Adv. Visual	-0.15	0.00	-0.01	0.12	0.03
2 Basic Visual + Sonif.	0.61***	0.89***	0.51**	0.83***	0.93***
3 Adv. Visual + Sonif.	0.58**	0.85***	0.48*	0.79**	0.95***

Then radiologists underwent to four distinct sessions including basic and advanced visualization with and without the sonification for the 32 de-identified brains. Since AD can be asymmetrical and the aim of this session was to mimic a standard diagnosis procedure, it was necessary to dissect these regions by hemispheres, analyzing the lefts and the rights. So for each one of the 32 unidentified brain scans of patients, physicians analyzed four parts of the brain: left frontal (LF), left parietal (LP), right frontal (RF), and right parietal (RP). The results were then compared to the ground truth provided by Dr. Friedman.

As shown in **Table 6** for each combination of two physicians and four sessions, there was a significant positive correlation between the ground and the frontal, parietal and worst regions. Comparing basic and advanced inspections with and without sonification, a significant increase of the correlation is shown in **Table 6** and also illustrated in **Figure 8**.

Since the highest significant correlation was always between the ground truth and the worst scores, accuracy was assessed in terms of concordance between the truth and worst results and the correlation for the association of the regional scores with worst cores is presented in **Table 7**. In **Table 7**, all correlations were significant ($p < 0.01$). The worst scores were most highly

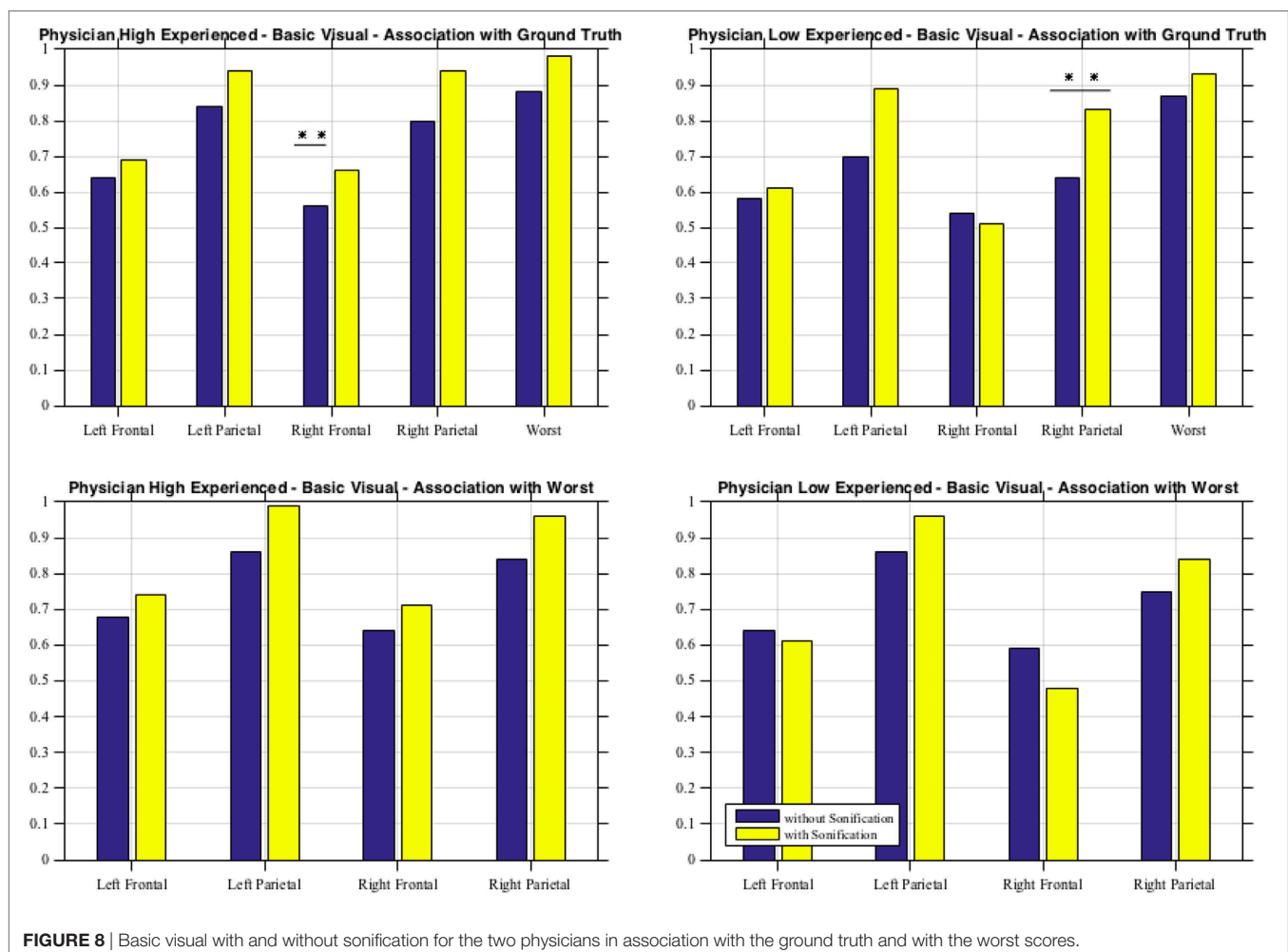


TABLE 7 | Pearson's correlation for association with worst scores.

Sess.	LF	LP	RF	RP
Physician 1 (high experienced)				
Basic Visual	0.68	0.86	0.64	0.84
Advanced Visual	0.57	0.97	0.55	0.91
Basic Visual + Sonification	0.74	0.99	0.71	0.96
Advanced Visual + Sonification	0.87	0.96	0.85	0.92
Physician 2 (low experienced)				
Basic Visual	0.64	0.86	0.59	0.75
Advanced Visual	0.57	0.97	0.55	0.91
Basic Visual + Sonification	0.61	0.96	0.48	0.84
Advanced Visual + Sonification	0.52	0.85	0.44	0.77

correlated with the scores from left parietal region, highlighted in bold. The result from these tables for basic visuals is shown in **Figure 8**.

5. DISCUSSION AND CONCLUSION

Effort has been made in the recent years to explore alternative representation methods to improve the accuracy and consistency in the diagnosis process. Several studies have explored the possibility of aided diagnosis tools using auditory feedbacks. In this work, we present a novel Triple-Tone Sonification technique to analyze PET scans of brains with different levels of Alzheimer's dementia. The model, presented and evaluated using subjective listening test, provides a supportive tool in the diagnosis of Alzheimer's when the metabolic activity is under inspection. Participants involved in this study were professional musician with trained musical ears and physicians in the field of radiology.

Results of the evaluation of the sonification method among the first subjective listening group indicate that participants with musically trained ears were able to categorize the sonifications with an average accuracy of 87% using "coarse categorization." The overall accuracy in diagnostic categorization improved from 87.19 to 92.5% when a finer gradation of categorization was utilized. This indicates that there exist sonifications generated by this technique that place a brain scan "in between" two categories when evaluated by a listener. In the case of "coarse categorization," the in-between scans were perceptually quantized into one of the coarse categories by the participant. In the case of "fine categorization," the participant was able to successfully categorize the sonification as an in-between case.

From the second subjective listening testing, as illustrated in **Tables 6 and 7** and **Figure 8**, results have shown that there is a significant positive correlation between the ground truth and the frontal, parietal, and worst regions. During the basic visual testing, one physician achieved an accuracy of 56–88% while the other performed at an average of 50–55%. Comparing basic and advanced inspections with and without sonification, a significant increase is shown in **Figure 8**. For each session, from an average improvement of about 10%, for the more severe cases, physicians results improved in accuracy up to about 30% with the sonification

added. These results show that, using the TTS method, medical researchers can utilize an additional way of interpreting data that can help detecting something that may be missed with traditional visual inspection.

Additional ways of interpreting and representing data have been largely explored, with a particular attention in the auditory feedback in works, such as Hermann (25), Hermann and Hunt (26), and Hunt and Hermann (27). Sonification provides new tools for recognizing patterns and analyzing data, extending the process of discovery in a number of diverse fields. In this work, a new sonification technique to meliorate the diagnosis of Alzheimer's dementia has been presented. The TTS technique implemented in this study has been evaluated on trained and untrained ears. The promising results acquired in this analysis and discussed in the previous section highlight more possibilities in helping identifying different stages of AD in the diagnosis process. Under the first evaluation of accuracy and consistency, the TTS technique validates its capabilities in allowing a finer gradation of providing accurate diagnosis. In this scenario, intra-reader consistency of categorization with sonification is superior to that of visualization tools alone. During the second round inspection, it was extended the influence of the sonification in the diagnosis of brain scans among two physicians. Here, intra- and inter-reader variability in accurate distinguishability of different levels of AD increased with the sonification added for both experienced and inexperienced physicians. To evaluate this technique, the correlation between each evaluation method has been proposed in this study. The correct diagnosis based on the ground truth has been computed to measure if sonification helped improve accuracy in the diagnosis.

A possible future scenario for the technique would be working with 3D audio and enhancing the analysis and/or experience of diagnosing AD. This sonification technique can be also used to augment existing visualization imaging techniques, such as MRI, in order to bring considerable improvements to medical diagnosis area. Furthermore the detune factor is currently manually changed according to the ground truth of the case. This is acceptable as a validation of the sonification technique, but not for an objective data-driven sonification model. For future investigations, this detune factor will have to be mapped to an appropriate feature of the data. This would ensure that the sonification is purely data-driven and remains completely objective, while still allowing for a variable detune factor to increase the severity of sonification with severity of disease.

The validation of the efficacy of the TTS tool proposed in this article provides evidence through higher accuracy in diagnosis among the two subjective listening groups when this sonification is used. The technique, successfully validated, with future investigations, can become an integral part of a physician's diagnostic toolkit and may open up in helping with AD diagnosis.

ETHICS STATEMENT

This study was carried out in accordance with the recommendations of the Division of Medical Ethics at NYU with

written informed consent from all subjects. All subjects gave written informed consent in accordance with the Declaration of Helsinki. The protocol was approved by the Division of Medical Ethics at NYU.

AUTHOR CONTRIBUTIONS

LG was a visiting student at the NYU Langone Medical Center and researcher involved in the study at the Music Technology Department of NYU Steinhardt. AR was the principal investigator of the project at the Music Technology Department of NYU Steinhardt.

ACKNOWLEDGMENTS

This research is supported in part by the National Center for Advancing Translational Sciences, National Institutes of Health with AR as Principal investigator of the project. The datasets utilized for sonification was obtained from Dr. Kent P. Friedman from the Radiology Department of New York University Langone Medical Center.

FUNDING

This work was supported by the NYU CTSI Award Number NIH/NCATS UL1 TR000038.

REFERENCES

- Berchtold N, Cotman C. Evolution in the conceptualization of dementia and Alzheimer's disease: Greco-roman period to the 1960s. *Neurobiol Aging* (1998) 19(3):173–89.
- McKhann G, Drachman D, Folstein M, Katzman R, Price D, Stadlan EM. Clinical diagnosis of Alzheimer's disease: report of the nincds-adrda work group under the auspices of department of health and human services task force on Alzheimer's disease. *Neurology* (1984) 34(7):939–44. doi:10.1212/WNL.34.7.939
- Meltzer C, Zubietta J, Brandt J, Tune L, Mayberg H, Frost J. Regional hypometabolism in Alzheimer's disease as measured by positron emission tomography after correction for effects of partial volume averaging. *Neurology* (1996) 47(2):454–61. doi:10.1212/WNL.47.2.454
- Frackowiak R, Pozzilli C, Legg N, Boulay GD, Marshall J, Lenzi G, et al. Regional cerebral oxygen supply and utilization in dementia. A clinical and physiological study with oxygen-15 and positron tomography. *Brain* (1981) 104(4):753–78. doi:10.1093/brain/104.4.753
- Miller-Thomas MM, Sipe AL, Benzinger TLS, McConathy J, Connolly S, Schwetye KE. Multimodality review of amyloid-related diseases of the central nervous system. *Radiographics* (2016) 36(4):1147–63. doi:10.1148/rg.2016150172
- Rice SL, Kent FP. Clinical pet-mr imaging in breast cancer and lung cancer. *PET Clin* (2016) 11(4):387–402. doi:10.1016/j.cpet.2016.05.008
- Kok TY, Padhy AK. The conundrum of pet/mr. *World J Nucl Med* (2012) 11(1):1–2. doi:10.4103/1450-1147.98717
- Illán IA, Górriz JM, Ramírez J, Salas-Gonzalez D, López MM, Segovia F, et al. 18f-fdg pet imaging analysis for computer aided Alzheimer's diagnosis. *Inf Sci* (2011) 181(4):903–16. doi:10.1016/j.ins.2010.10.027
- Stockbridge H, Lewis D, Eisenberg B, Lee M, Schacher S, van Belle G, et al. Brain spect: a controlled, blinded assessment of intra-reader and inter-reader agreement. *Nucl Med Commun* (2002) 23(6):537–44. doi:10.1097/00006231-200206000-00005
- Rossiter D, Ng W-Y. A system for the complementary visualization of 3d volume images using 2d and 3d binaurally processed sonification representations. *Proceedings of the 7th Conference on Visualization '96*. Los Alamitos, CA: IEEE Computer Society Press (1996). p. 351–4.
- Hermann T, Hunt A, Neuhoof JG. *The Sonification Handbook*. Berlin: Logos Publishing House (2011).
- Walker BN, Kramer G. Sonification design and metaphors: comments on Walker and Kramer. *ACM Trans Appl Percept* (2005) 2(4):413–7. doi:10.1145/1101530.1101535
- Walker BN, Mauney LM. Universal design of auditory graphs: a comparison of sonification mappings for visually impaired and sighted listeners. *ACM Trans Access Comput* (2010) 2(3):16. doi:10.1145/1714458.1714459
- Polli A. "Atmospherics/weather works": a spatialized meteorological data sonification project. *Leonardo Music J* (2005) 38(1):31–6. doi:10.1162/leon.2005.38.1.31
- Polli A. Heat and the heartbeat of the city: sonifying data describing climate change. *Leonardo Music J* (2006) 16(1):44–5. doi:10.1162/lmj.2006.16.44
- Baier G, Hermann T. The sonification of rhythms in human electroencephalogram. In: Brazil E, editor. *Proceedings of the 10th International Conference on Auditory Display*. Sydney: Springer-Verlag (2004). p. 6–9.
- Kagawa T, Kudo H, Tanoue S, Kiyosue H, Mori H, Nishino H, et al. A supporting method of medical imaging diagnosis with sonification. *Proceedings of the 2002 International Conference on Auditory Display*. Kyoto: IEEE (2012). p. 699–704.
- Ballora M, Pennycook B, Ivanov PC, Glass L, Goldberger AL. Heart rate sonification: a new approach to medical diagnosis. *Leonardo Music J* (2004) 37(1):41–6. doi:10.1162/002409404772828094
- MIM Software Inc. (2014). Available from: <https://www.mimsoftware.com/>
- Piper J. Quantitative comparison of spatial normalization algorithms for 3d pet brain scans. *J Nucl Med* (2007) 48(2):403–8. doi:10.1002/hbm.10047
- Marcus C, Whitworth P, Surasi D, Subramaniam R. Pet/ct in the managements of thyroid cancers. *Am J Roentgenol* (2014) 202(6):1316–29. doi:10.2214/AJR.13.11673
- Nelson A, Piper J, Friedland R, Freeman B. Probabilistic human brain atlas for functional imaging: comparison to single brain atlases. *J Nucl Med* (2007) 48(2):3508–26.
- MATLAB. *Version 7.10.0 (R2010a)* (2010). The MathWorks Inc.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at <http://www.frontiersin.org/article/10.3389/fneur.2017.00647/full#supplementary-material>.

The p-value is the probability that it is found the current result if the correlation coefficient is zero (null hypothesis). If this probability is lower than the conventional 5% ($p < 0.05$), the correlation coefficient is called statistically significant.

Alzheimer's dementia can affect patients' brains in dissimilar and asymmetrical forms. In some cases, the decrease in metabolic activity manifests itself uniformly across lobes, while in others the same decrease is more concentrated in certain regions of the lobes (28). To take into account this asymmetry, the datasets after being segmented into lobes are segmented into left and right halves. Diagnosis of the stage of Alzheimer's is performed according to the worst affected area of the brain (29). In this manner, in case of asymmetries, left and right halves of the brain can be rendered to two separate sonifications, thereby allowing the physician to identify the "worse half" and make a diagnosis that accurately represents the stage of AD. A quantitative estimation is also performed to automatically identify the worse half, as follows:

$$\begin{aligned} \text{if } (|\Delta_{FL}| + |\Delta_{PL}|) \text{ of left} > (|\Delta_{FL}| \\ + |\Delta_{PL}|) \text{ of right} \Rightarrow \text{worst half} = \text{left half} \\ \text{else worst half} = \text{right half}. \end{aligned}$$

24. Wilson S, Cottle D, Collins N. *The SuperCollider Book*. Cambridge: The MIT Press (2011).
25. Hermann T. Taxonomy and definitions for sonification and auditory display. In: Susini P, Warusfel O, editors. *Proceedings of the 14th International Conference on Auditory Display*. Paris, France: IRCAM (2008). p. 24–7.
26. Hermann T, Hunt A. Guest editors' introduction: an introduction to interactive sonification. *IEEE MultiMedia* (2005) 12(2):20–4. doi:10.1109/MMUL.2005.26
27. Hunt A, Hermann T. The importance of interaction in sonification. In: Barrass S, Vickers P, editors. *Proceedings of the 10th International Conference on Auditory Display*. Sydney: ICAD (2004). p. 1–8.
28. Geroldi C, Akkawi NM, Galluzzi S, Ubezio M, Binetti G, Zanetti O, et al. Temporal lobe asymmetry in patients with Alzheimer's disease with delusions. *J Neurol Neurosurg Psychiatry* (2000) 69(2):187–91. doi:10.1136/jnnp.69.2.187
29. Rosen A. An introduction to cognitive aging and dementia: a clinical neuropsychologist's perspective. *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems*. San Jose: ACM (2016). p. 948–51.

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.

Copyright © 2017 Gionfrida and Roginska. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Action Observation Plus Sonification. A Novel Therapeutic Protocol for Parkinson's Patient with Freezing of Gait

Susanna Mezzarobba^{1,2,3*}, Michele Grassi¹, Lorella Pellegrini^{2,3}, Mauro Catalan², Bjorn Kruger⁴, Giovanni Furlanis², Paolo Manganotti^{2,3} and Paolo Bernardis¹

¹ Department of Life Sciences, University of Trieste, Trieste, Italy, ² Azienda Sanitaria Universitaria Integrata di Trieste, Trieste, Italy, ³ Department of Medical, Surgical and Health Sciences, University of Trieste, Trieste, Italy, ⁴ Gokhale Method Institute, Palo Alto, CA, United States

OPEN ACCESS

Edited by:

Diego Minciocchi,
University of Florence, Italy

Reviewed by:

Matthew Rodger,
Queen's University Belfast,
United Kingdom
Elisa Pelosin,
University of Genoa, Italy

*Correspondence:

Susanna Mezzarobba
mezzarob@units.it

Specialty section:

This article was submitted
to Neuroprosthetics,
a section of the journal
Frontiers in Neurology

Received: 20 September 2017

Accepted: 13 December 2017

Published: 04 January 2018

Citation:

Mezzarobba S, Grassi M, Pellegrini L,
Catalan M, Kruger B, Furlanis G,
Manganotti P and Bernardis P (2018)
Action Observation Plus Sonification.
A Novel Therapeutic Protocol for
Parkinson's Patient with Freezing
of Gait.
Front. Neurol. 8:723.
doi: 10.3389/fneur.2017.00723

Freezing of gait (FoG) is a disabling symptom associated with falls, with little or no responsiveness to pharmacological treatment. Current protocols used for rehabilitation are based on the use of external sensory cues. However, cued strategies might generate an important dependence on the environment. Teaching motor strategies without cues [i.e., action observation (AO) plus Sonification] could represent an alternative/innovative approach to rehabilitation that matters most on appropriate allocation of attention and lightening cognitive load. We aimed to test the effects of a novel experimental protocol to treat patients with Parkinson's disease (PD) and FoG, using functional, and clinical scales. The experimental protocol was based on AO plus Sonification. 12 patients were treated with 8 motor gestures. They watched eight videos showing an actor performing the same eight gestures, and then tried to repeat each gesture. Each video was composed by images and sounds of the gestures. By means of the Sonification technique, the sounds of gestures were obtained by transforming kinematic data (velocity) recorded during gesture execution, into pitch variations. The same 8 motor gestures were also used in a second group of 10 patients; which were treated with a standard protocol based on a common sensory stimulation method. All patients were tested with functional and clinical scales before, after, at 1 month, and 3 months after the treatment. Data showed that the experimental protocol have positive effects on functional and clinical tests. In comparison with the baseline evaluations, significant performance improvements were seen in the NFOG questionnaire, and the UPDRS (parts II and III). Importantly, all these improvements were consistently observed at the end, 1 month, and 3 months after treatment. No improvement effects were found in the group of patients treated with the standard protocol. These data suggest that a multisensory approach based on AO plus Sonification, with the two stimuli semantically related, could help PD patients with FoG to relearn gait movements, to reduce freezing episodes, and that these effects could be prolonged over time.

Keywords: freezing of gait, action observation, Sonification, Parkinson's disease, cueing

INTRODUCTION

For decades, motor and gait difficulties have been identified as the main symptoms of Parkinson's disease (PD), and drug therapy—based on dopamine and its agonists—was considered the only feasible solution to ameliorate symptoms. Amid these motor symptoms and gait abnormalities, freezing of gait (FoG) is the most debilitating; a sudden episodic inability to generate an effective stepping, which commonly, leads to falls.

However, PD is a complex neurological disease that comprises also severe psychiatric and cognitive symptoms. Today, the benchmark to treat PD symptoms, especially when they worsen, is the use of specific rehabilitation protocols together with medication and/or surgical therapy.

Drug therapy in PD is a symptomatic therapy, primarily aimed at restoring dopaminergic function in the striatum. Although irreplaceable in the treatment of PD symptoms, several data demonstrate also negative effects, produced by dopamine on certain movements, and cognitive functions. Indeed, while dopaminergic medication clearly enhances certain motor functions, at the same time might negatively affect the learning of movement sequences (1, 2), as well as specific cognitive functions (3, 4). Moreover, the absent or controversial pharmacological responsiveness of FoG has led to an increasing interest in rehabilitation interventions aimed at functional recovery and autonomy, by relearning a physiological gait pattern.

Currently, protocols employed for rehabilitation of PD—with and without FoG—are based on the use of external sensory cues—mainly visual, but also auditory and tactile—because it allows the switch from *automatic* movement (habitual)—controlled by frontostriatal pathways, that PD patients have compromised—to *voluntary controlled* movement [goal directed (5)]. Specifically, Vandenbosche et al. (6) showed that PD patients with FoG exhibit a specific impairment in the acquisition of automaticity—correlated with the working memory functions—and suggested that therapies should focus on training that reduce working memory load, as the cued strategies.

During exposure to visual and auditory cues, patients with FoG, as those without, improve gait kinematics and reduce freezing. Interestingly, visual cues have more powerful effects than auditory cues for reducing FoG (7); proving that the inability to maintain effective scaling of step amplitude could be an important FOG-related deficit. Conversely, auditory cues (metronome) seem to be less effective in the regularization of altered cadence, and disordered coordination of inter-limb movement in patients with FoG. Unfortunately, it has been shown that cueing might generate an important dependence on the environment, particularly the visual ones, considering how important is the exploration of the whole visual field in intentional walk (8).

In the last years, several researchers try to use cues differently. Young et al. (9) asked Parkinson's patients with and without FoG to listen to different auditory cues (i.e., a metronome or ecological footsteps sounds recorded on gravel), and to step in place to each cue, synchronizing their own stepping in time to the sound. Results in patients with FoG showed remarkable improvements in temporal regularity. The authors claim that in PD patients with FoG, the mechanism “action imitation enhances the motor

performance” is supported by their results with action-relevant cues (i.e., footsteps recorded on gravel).

Teaching *motor strategies*, without cues to overcome or avoid freezing episodes can be an alternative/innovative approach to rehabilitation, that matters most on appropriate allocation of attention (10), and lightening cognitive load. One of these strategies—action observation (AO)—is based on the activation/sharing of a common neural substrate, the mirror system (11). The *priming effect* of AO on subsequent motor execution of the observed gesture is well known in neurorehabilitation, although few evidences are available for treatment of patients of PD (12).

Furthermore, one way to reduce cognitive load in the recovery/learning of motor gestures is the use of multisensory approaches that enhance perceptual processes (13), which are known to be reduced in PD patients with FoG (14). The use of multisensory stimuli improves the learning process (13, 15) thanks to a reduced cognitive load, and to an easier storage in short-term memory (16, 17). But, to exert the most efficient facilitatory effect, pairs of stimuli composing the multisensory stimulus should be congruent, and not simply concomitant in space and/or time (18, 19). These findings have stimulated interest toward the use of audiovisual stimuli to facilitate relearning of movements also in the field of neurological rehabilitation.

Evidence on the efficacy of action-related *sonified* sounds (synthesized sounds obtained with a *Sonification* procedure, see the next paragraph) to improve motor performance is well documented [for a review, see Ref. (20)], although in PD patients is still limited. Indeed, Rodger et al. (21) used two different types of sounds (ecological and synthesized) to help guide and improve walking actions of PD patients. One of these techniques was based on Sonification of the ground reaction forces. Both methods showed that PD patients could use rich auditory representations of action to guide and improve the quality of walking, and reducing the risk of falls and injury. Moreover, Schmitz et al. (17) demonstrated that the Sonification of movements enhance the activity in the human AO system including subcortical structures of the motor loop; and therefore, may be an important method to enhance therapy effects in neurological rehabilitation.

The most natural way to use audio–video stimuli is to present images together with *ecological* sounds (i.e., a walker and the sound of his/her feet). Instead of utilizing the real sounds produced during gait, we employed synthesized sounds obtained with the Sonification technique (22). Specifically, in our audiovisual stimuli, the auditory component is obtained by transforming kinematic data of relevant body part movements—visible in the video—into sounds. This process is called *Sonification*. We choose to use sonified sounds—in place of real sounds (i.e., footsteps sound)—because in this way we can convey additional information, important for the understanding and reproduction of a correct movement (i.e., differences in the velocity of the hips rotation during gait), that otherwise will be ignored. This final stimulus is a sort of *augmented* audio–video stimulus. The processing of auditory and visual information together facilitates the recognition of the movement in its spatial and temporal aspects, and the relearning process of the correct pattern of movements. These stimuli could be of particular importance for PD patients

with FoG in which these components are altered, and in which, probably, visuo-perceptive modifications may be present (14).

We hypothesized that AO can be used to facilitate recovery of defective motor control, and given that PD patients with FoG may have major shortages of attention resources, a multisensory approach (i.e., audiovisual stimuli) would help to further reduce the attention load, facilitating learning processes.

The aim of this study was to test the efficacy of a novel protocol based on AO technique and Sonification, and to compare the effect with a standard protocol based on external sensory cues. With this purpose, we designed and realized an experimental study to test the effectiveness of these two protocols in two groups of PD patients with FoG. We hypothesized that gait improvement of the AO plus Sonification protocol would be better than those obtained with the standard protocol, both in the short term and the long term (3-month follow-up).

MATERIALS AND METHODS

Design

The whole pilot RCT was carried out from April 2015 to December 2016. Post-intervention measures were collected at the end, 1 month, and 3 months after the end of the treatment. Patients were randomly assigned to two different training group (experimental and control groups). An investigator, neither involved in the treatment protocol nor in the selection and evaluation of patients, created the computerized randomization procedure (blocked randomization). The same investigator concealed treatment allocation by using small opaque envelopes. Three trained physical therapists with a solid experience in the treatment of PD were involved in the evaluation—one of them—and in the treatment of patients—the other two. Outcome measures were videotaped and also evaluated by a second independent rater blind to the whole experimental study. In case of discrepancies between the two, a third blind rater was used to resolve the evaluation.

Patients were advised to have their medical treatment continued unchanged throughout the study. This study was carried out in accordance with the recommendations of the “Comitato Etico Regionale Unico” guidelines, with written informed consent from all subjects. All subjects gave written informed consent in accordance with the Declaration of Helsinki. The protocol was approved by the Institutional Ethics Committee (Comitato Etico Regionale Unico—Friuli Venezia Giulia. Protocol no. 4456—05.02.2015). Patients who agreed to participate always signed a written informed consent and they were able to leave the experiment at any moment, with no additional explanations. The study has been registered at <http://Clinicaltrials.gov>, NCT03249155.

Participants

Thirty-seven patients with idiopathic PD (see **Figure 1**), according UK Brain Bank (23) were assessed by a neurologist expert in movement disorders, from the outpatient Neurological Clinic, Cattinara Hospital. Eligibility criteria were occurrence of FoG (24) based on patient's verbal account of his/her freezing experience (or recognition of their typical FoG experience when this symptom was described to him/her by a physician); stages 1–3

on the Hoehn and Yahr scale (25); stable medication regimen for at least 8 weeks; no major depressive symptoms as defined by a Beck Depression Inventory score ≤ 16 [BDI (26)]; no signs of dementia as defined by a Mini-Mental Status Examination score > 24 [MMSE (27)]. The exclusion criteria were evidence of any adjunctive orthopedic comorbidities that make it impossible to use physical activities and an independent locomotion; others neurological and psychiatric disease; presence of any implanted stimulating or pacing device in central nervous system. Prior power analysis estimated a sample size group of 10 participants. After the first assessment, we enrolled a total of 24 patients (see **Figure 1**). Two subjects dropped out due to concurrent, unrelated medical events: thus, 22 patients completed the study (see **Table 1**).

Experimental Procedures

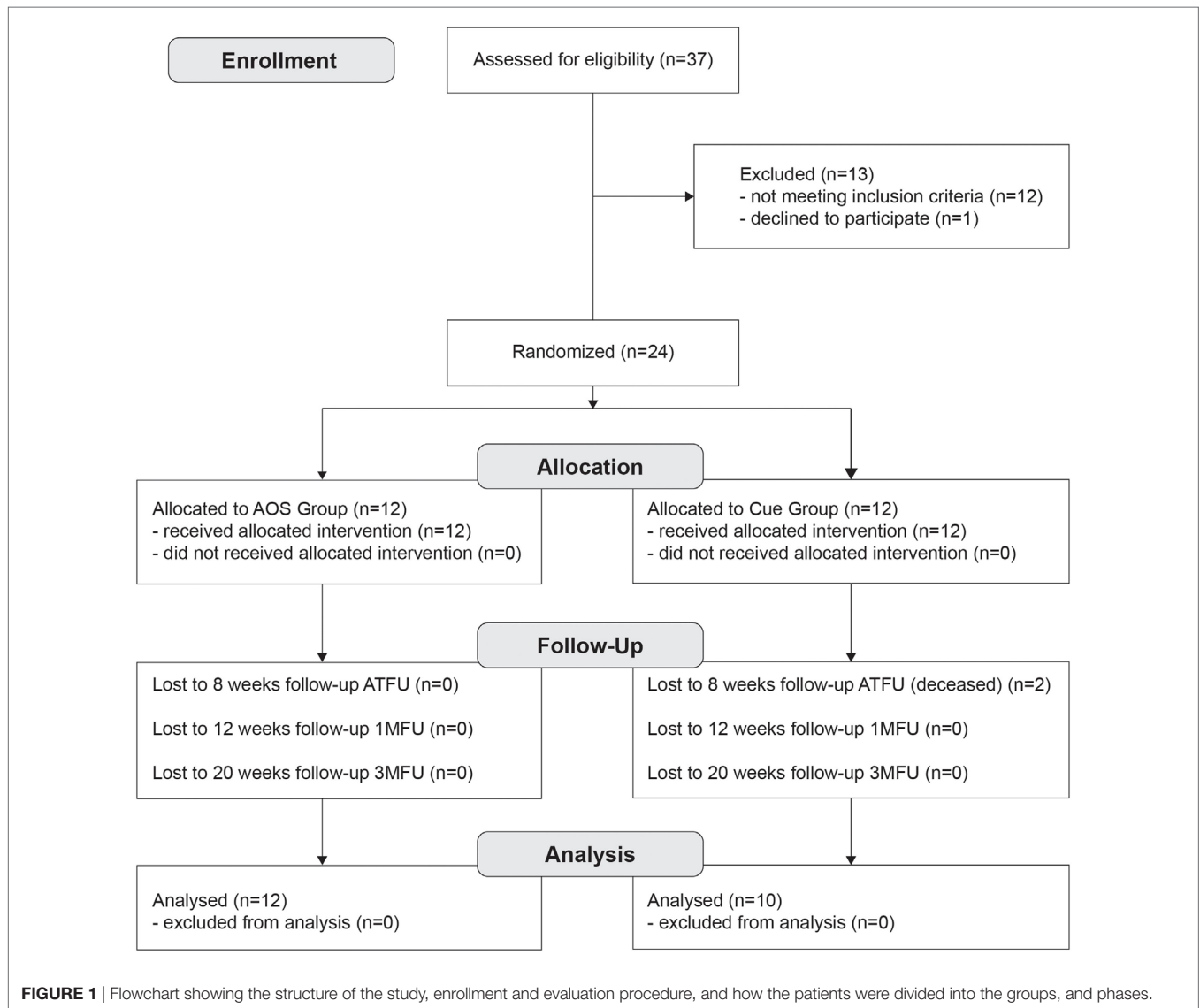
All participants underwent to a 1 h of rehabilitation training during their ON condition (approximately 1 h after the antiparkinsonian medication intake), twice a week, for 8 consecutive weeks, and a total of 16 training sessions.

Experimental Group

The protocol was based on the AO method plus Sonification (AOS). During each training session, eight videos, showing an actor performing eight different motor gestures, were presented to the patient that then tried to repeat, according to the Modeling principles (28). Each video, lasting 1.5 min, was composed by images (from fronto-lateral perspectives) and sounds (obtained with Sonification) of eight specific motor gestures. These gait related gestures were useful for ameliorating weight shifting, step scaling, and bilateral coordination of stepping, known as locomotion features related to FoG. In each session, all videos were presented, from simple to complex motor actions. The contents of the eight videos are reported in Appendix. Each session started with the observation of the audio–video projected on a large sized screen (2.5 m \times 2 m) located in front of the patient at a distance of 2 m. During AO, to increase the accuracy of imitation, patients were asked to attend to the peculiar characteristics of the observed action, and no movements were allowed. At the beginning, after video observation, patients had to practice repetitively the observed actions for the same time (1.5 min). Then, patients performed *on line* the same motor gesture while they were watching the videos. With the aim to facilitate the modeling process, an expert physiotherapist in AO treatment, encouraged and corrected patient's motor execution. Each video was repeated twice.

Control Group

The same eight motor gestures were performed also in the Cue control group with the same order and amount of time, by using attentional strategies. During each training session, patients were asked to practice the motor gesture by means of visual (stripes on the floor) or auditory (metronome) cues, to facilitate the learning of temporal and spatial parameters. As for the experimental group, the expert physiotherapist encouraged and corrected each patient's motor execution to facilitate correct motor learning process. Following physical therapist's instructions, patients progressively learned to perform the eight motor gestures without cues.



Participants of both groups were instructed to not practice further rehabilitation/physiotherapy treatments during the duration of the study. The two therapists involved in the treatments were not dedicated to one group, but equally assigned to both of them.

Clinical Outcomes

The patients who met inclusion/exclusion criteria underwent to a clinical and motor functional evaluation before the treatment (BT), after the treatment (AT), 1 month (1MFU), and 3 months AT (3MFU). The neuropsychological evaluations were only done at the baseline and 1 month (1MFU) AT, since the minimum interval for test administration is 3 months. All clinical evaluations were performed by an experienced neurologist, and a physiotherapist blinded to participants' allocation.

As primary outcome, FOG duration and severity were assessed by using New Freezing of Gait Questionnaire (NFOGQ).

Particularly, we calculate an index of improvement obtained at AT, 1MFU, and 3MFU evaluations, respect to the BT evaluation (see Data Analysis).

A priori power analyses based on a previous experiment that compared the two treatment protocols in individuals with PD and FoG (29), suggested 10 participants per group to achieve a medium effect size ($f = 0.45$, $\alpha p = 0.05$, power = 0.95, critical $F = 4.41$). We recruited 12 participants for each group to account for possible attenuation.

As for the secondary outcomes, disease severity was tested with the Unified Parkinson's Disease Rating Scale (UPDRS II–III), the Hoehn and Yahr scale (25), and quality of life with the 39-item PD Questionnaire (30). Motor functional performance evaluation included Modified Parkinson's Activity Scale (31), Timed Up and Go (32), and 6-min walking test (33). Berg Balance Scale (34) was used to assess static and dynamic balance capabilities. Also for the secondary outcome measures, we calculated an improvement index.

Neuropsychological Evaluation

We assessed patients' most important cognitive functions, useful for learning new motor ability: executive functions, attention, and memory capabilities (Table 2), to exclude that the cognitive

TABLE 1 | Demographic and clinical characteristics of patients with Parkinson's disease at baseline.

	AOS group (n = 12)	Cue group (n = 10)
Patients' characteristics		
Gender female/male (% female)	5/7 (42%)	3/7 (30%)
Age (years)	74.67 ± 5.93	72 ± 5.87
Disease duration (years)	10.75 ± 3.44	9.4 ± 4.86
Mini-Mental State Exam	27.46 ± 1.81	26.58 ± 1.1
Beck Depression Inventory	8 ± 5.98	6.4 ± 5.93
Hoehn and Yahr stage	2.33 ± 0.49	2.3 ± 0.67
Levodopa equivalent	972.5 ± 253.17	983.22 ± 379.58
Primary outcome measure		
NFOGQ	18.17 ± 4.61	16.6 ± 7.86
Secondary outcome measures		
UPDRS II Total	16.42 ± 5.99	17.1 ± 6.47
UPDRS III Total	32.92 ± 8.69	33.2 ± 13.99
MPAS	53.75 ± 6.27	50.2 ± 9.22
BBS	47.75 ± 4.16	47.8 ± 3.46
6MWT (s)	280.75 ± 93.34	296.75 ± 48.32
TUG (ms)	1,299.75 ± 376.74	1,271.3 ± 615.86
PDQ39 mobility	49.38 ± 22.54	35.5 ± 25.87
PDQ39 activities of daily living	33.4 ± 21.99	28.75 ± 17.84
PDQ39 emotional well-being	23.82 ± 21.84	33.33 ± 20.13
PDQ39 stigma	21.04 ± 22.94	18.75 ± 20.41
PDQ39 social support	2.78 ± 6.49	15.83 ± 22.03
PDQ39 cognitions	23.23 ± 15.52	31.88 ± 19.64
PDQ39 communication	24.31 ± 20.86	29.17 ± 21.25
PDQ39 bodily discomfort	31.67 ± 21.92	27.5 ± 19.27
PDQ39 Total	51.67 ± 26.9	50.8 ± 29.43

Data are mean ± SD or as otherwise indicated.

n, number of patients; NFOGQ, New Freezing of Gait Questionnaire; UPDRS, Unified Parkinson's Disease Rating Scale II and III; PDQ39, Parkinson's disease questionnaire 39; MPAS, Modified Parkinson's Activity Scale; BBS, Berg Balance Scale; 6MWT, 6-min walking test; TUG, Time-Up-and-Go; PDQ39, 39-item PD Questionnaire.

profile of the patients of each group was changed after the end of the treatments. This was important to exclude that different levels of efficacy were due to differences in the cognitive profile of the two groups of patients. Global cognitive functioning was tested with the Montreal Cognitive Assessment (35); short-term and long-term memory functions with Digit Span backward (36), Corsi Test (37), Babcock Story Recall Test (38); attention with the Attentive Matrices (39), the Stroop Test (40), and Trail Making Test: parts A and B (41). Executive functions were evaluated with the Frontal Assessment Battery (42) and Tower of London Test (43). Abstract reasoning by Raven matrices (44). All neuropsychological tests scores were corrected on age, sex, and education using normative values. Moreover, patients were always tested in "ON" condition during their optimal antiparkinsonian medication.

Audiovisual Stimuli

The short video used in the experimental group showed two healthy actors' (one male and one female) performing the 8 motor gestures from a lateral and frontal perspective, for a total of 32 different videos (2 gender × 8 gestures × 2 perspectives). Moreover, prior rehabilitation treatment, and to be comfortable with the procedures, each participant practiced the tasks using other videos showing three movements test. The sounds of each video were obtained with the Sonification technique, by transforming kinematic data (i.e., velocity) recorded during the execution of the eight gestures, into audio pitch variations. Actors performed all tasks barefoot, walking along a 10-m walkway surrounded by a seven-camera motion-capture Qualisys System (120 Hz). During the execution of each motor gestures, kinematic data were collected recording four retroreflective markers placed on the left and right anterior superior iliac spine to calculate pelvis movement velocity, and on the left and right lateral malleoli to calculate inferior limbs velocity. All data were recorded and preprocessed by a dedicated software Qualisys Track Manager, and frame by frame instantaneous speed was obtained, and transformed

TABLE 2 | Cognitive profile of patients with Parkinson's disease at baseline and at 1-month follow-up (1MFU).

Cognitive domain	Test	AOS group at baseline	Cue group at baseline	AOS group at 1MFU	Cue group at 1MFU
		Mean ± SD	Mean ± SD	Mean ± SD	Mean ± SD
Short-term memory	Digit Span Backward	4.84 ± 1.27	4.84 ± 1.14	4.59 ± 1.14	4.58 ± 1.02
	Corsi Test	4.85 ± 0.84	4.34 ± 0.84	4.50 ± 0.83	5.01 ± 1.44
Long-term memory	Babcock Story Recall Test				
	Immediate recall	5.78 ± 1.51	4.52 ± 2.48	5.27 ± 1.62	3.83 ± 2.50
	Delayed recall	5.8 ± 1.4	4.41 ± 3.07	5.16 ± 1.88	3.41 ± 3.27
Attention	Attentive Matrices	40.94 ± 8.35	32.46 ± 12.86	36.91 ± 11.27	31.56 ± 13.51
	Trail Making Test:				
	Part A (s)	30.35 ± 21.59	68.51 ± 46.78	52.44 ± 45.98	55.27 ± 38.02
	Part B (s)	147.34 ± 116.51	190.58 ± 130.99	148.35 ± 199.33	166.68 ± 117.38
	Stroop Test				
	Time (s)	28.78 ± 19.19	52.73 ± 42.05	42.59 ± 26.17	47.72 ± 29.42
Executive functions	Errors	3.01 ± 3.55	1.83 ± 1.76	3.64 ± 2.47	3.36 ± 3.39
	Frontal Assessment Battery	16.7 ± 1.22	16.19 ± 1.71	16.55 ± 1.78	16.11 ± 1.80
	Tower of London Test	17.99 ± 1.4	19.56 ± 6.38	18.61 ± 0.29	18.22 ± 1.41
Reasoning	Raven's progressive matrices	29.71 ± 6.18	27.28 ± 5.17	28.85 ± 6.16	27.70 ± 5.31

Data are mean ± SD or as otherwise indicated.

into pitch audio by the open source framework Pd (45) using modules developed by Henkelmann and colleagues (46, 47). The Sonification itself is done in the following steps: first, median filter with a window size of three frames is applied to the kinematic data to suppress sensor noise from the kinematic data. Second, the data stream is linearly scaled to an interval from 0 to 1. Third, the pitch sound itself is generated. Forth, the sound is mapped to the left or right audio channel. The PD module we used to generate our stimuli can be found in the supplemental materials for this publication. 32 audio track were gained (2 actors gender \times 8 different gestures \times 2 perspectives) for each motor gestures, and 2 audio tracks for each movement test. Videos were edited by using Final Cut Pro X software, with a dubbing procedure to merge the sounds with the video part of each gesture. The kinematic-acoustic recording was provided with a visual auditory stimulus congruence.

Data Analysis

Preliminary, applying the Kolmogorov–Smirnov (KS) test we verified the sustainability of a normal distribution for the primary and secondary clinical outcomes. Highly skewed and kurtotic variables were log transformed and then KS tested for the effectiveness of the correction. Outcomes that failed this second test where excluded from the analysis.

For each clinical outcome, the improvement (gain) from pre-training to posttraining (AT, 1MFU, and 3MFU) was computed for each participant by subtracting each person's pretraining score from his/her posttraining score and dividing the difference for the pretraining performance. Formally:

$$\text{gain} = \frac{\text{post-training} - \text{pre-training}}{\text{pre-training}}.$$

Systematic differences in pretraining scores between the two groups of patients were preliminary excluded with *t*-tests on both primary and secondary outcomes measures. As for the cognitive profile, we verified for potential differences in the pretraining (BT) and modification after 3 months (1MFU) with a 2×2 mixed factors ANOVA (Group and Time of evaluation). The results are reported in **Table 3**.

The hypothesis of differences in improvement (gain) between the experimental and control groups was tested by a mixed design ANOVA on the gain scores using *Group* (AOS vs. Cue) as a between subjects factor, and *Time of evaluation* as within-subject factor. Besides main effects, we also considered the interaction terms *Group \times Time* to assess the stability of the effect across evaluation. *Post hoc* Bonferroni's test was employed to assess gain score differences between groups at each time. The significant change threshold was set at $p \leq 0.05$. We interpreted the meaningfulness of the significant changes using the generalized eta-squared (η^2G) statistics calculated following the guidelines by Olejnik and Algina (48) and Bakeman (49).

Linear Discriminant Analysis (LDA)

Besides these statistical criteria, to examine the clinical impact of the AOS and Cue training on outcomes scores, we used also an automated classification rates criterion. This method is commonly used as a technique for pattern classification. In our case,

TABLE 3 | Split plot ANOVA results.

	SS	Mean sq	F	Pr(>F)	η^2G (%)
Main effect of treatment (Group)					
Primary outcome measure					
NFOGQ	2.741	2.741	24.28	0.000	50
Secondary outcome measures					
PDQ39 mobility	12.630	12.633	14.91	0.001	38
UPDRSIII	4.962	4.962	13.36	0.002	33
PDQ39 bodily discomfort	4.962	4.962	13.36	0.002	33
PDQ39 Total	2.031	2.031	9.73	0.006	27
UPDRSII	1.427	1.427	11.14	0.003	24
BBS	0.121	0.121	6.42	0.020	24
6MWT	0.470	0.470	4.11	ns	15
TUG	1.070	1.071	3.95	ns	14
MPAS	0.021	0.021	0.56	ns	2
PDQ39 cognitions	0.471	0.471	0.47	ns	2
Main effect Time (within subjects)					
Primary outcome measure					
NFOGQ	0.091	0.046	3.11	ns	3
Secondary outcome measures					
PDQ39 mobility	0.151	0.075	0.60	ns	1
UPDRSIII	0.334	0.167	2.14	ns	3
PDQ39 bodily discomfort	0.334	0.167	2.14	ns	3
PDQ39 Total	0.158	0.079	2.05	ns	3
UPDRSII	0.049	0.025	0.48	ns	1
BBS	0.005	0.003	3.01	ns	1
6MWT	0.061	0.031	2.91	ns	2
TUG	0.222	0.111	3.66	0.035	3
MPAS	0.009	0.005	0.69	ns	1
PDQ39 cognitions	1.546	0.773	4.17	0.023	6
Interaction effect (Time \times Group)					
Primary outcome measure					
NFOGQ	0.002	0.001	0.07	ns	0
Secondary outcome measures					
PDQ39 mobility	0.386	0.193	1.54	ns	2
UPDRSIII	0.197	0.098	1.26	ns	2
PDQ39 bodily discomfort	0.197	0.098	1.26	ns	2
PDQ39 Total	0.261	0.131	3.41	0.043	5
UPDRSII	0.047	0.023	0.46	ns	1
BBS	0.002	0.001	0.90	ns	0
6MWT	0.022	0.011	1.03	ns	1
TUG	0.034	0.017	0.56	ns	1
MPAS	0.014	0.007	1.02	ns	1
PDQ39 cognitions	0.451	0.225	1.22	ns	2

ns, not significant; NFOGQ, New Freezing of Gait Questionnaire; UPDRS, Unified Parkinson's Disease Rating Scale II and III; PDQ39: Parkinson's disease questionnaire 39; MPAS, Modified Parkinson's Activity Scale; BBS, Berg Balance Scale; 6MWT, 6-min walking test; TUG, Time-Up-and-Go; PDQ39, 39-item PD Questionnaire; SS, sum of squares.

UPDRSIII, TUG, PDQ39 mobility, and PDQ39 bodily discomfort were log transformed for normality. Index η^2G is the generalized eta-squared statistics calculated following the guidelines by Olejnik and Algina (48) and Bakeman (49). For the two main effects and interaction tested we used the formulas $SSA/(SSA + SSs/A + SSPs/A)$, $SSP/(SSP + SSs/A + SSPs/A)$ and $SSPA/(SSPA + SSs/A + SSPs/A)$, where A and P refer to our variables Group and Time, respectively, and s represents the subject factor. Errors related SS are not shown in this table.

we used this method to compare or classify the clinical profiles of the patients in the two groups, and at the different stages of the experimental study. Automated classification problems involve continuous input variables (i.e., our clinical scales), and categorical outcomes (i.e., the rehabilitation protocol or the stage of the study). The algorithm has to learn to predict the category from the input data. We used an LDA algorithm in two differ ways,

to discriminate between groups and within subjects. Finally, we choose to complement standard analysis of variance with this LDA because of recommendations on using simulative approaches to data analysis with small samples (50).

First, the algorithm learned between-groups' discriminative criterion on a fraction of 70% of the data set, then for testing we applied the criterion on the remaining fraction of 30% (see Supplementary Material), measuring the classification accuracy in terms of sensitivity index (51). Average sensitivities were based on a complete random design, simulating all possible combination of 70–30% of the participants (50). Particularly, we trained and tested the LDA four times, over each evaluation time (BT, AT, 1MFU, and 3MFU), and considering the rehabilitation protocol attended by participants as the categorical outcome to predict.

Second, to evaluate stability over time, as in the case of interaction term in the ANOVA, we considered within-subject evaluations, using the LDA algorithm on a subset composed by 70% of pre- and posttraining individual's outcomes, and testing it on the remaining 30% of pre- and posttraining individual's outcomes.

In both cases, we expect that the more effective is the training in transforming the participants clinical profile, the more accurate is the LDA algorithm in (learn to) classify the participants within the training actually practiced. In the first LDA implementation, we expect that the LDA classification would fail only in the comparison between the two groups at the BT time window.

All the analyses were programmed using R statistical language (52).

RESULTS

At the baseline, there were no significant differences between groups with respect to demographics and clinical records, as shown in **Table 1**. Also for the cognitive profile, as reported in **Table 2**, there was no differences except for the interaction Group \times Time in the Corsi test [$F_{(1,20)} = 5.975$, $p = 0.024$, $\eta^2 = 0.225$], but when we compared the two groups in the two moments with a t -test, the difference was not significant [$t(20) = 1.449$, $p = 0.163$; $t(20) = -0.480$, $p = 0.636$].

Primary Outcome Measure

Action observation plus Sonification treatment had a significant positive effect in reducing the primary outcome measure, participant's ratings of FoG severity and duration, as shown at the end of the treatment, and most important, at the second follow-up (**Figure 2**, NFOGQ). Noteworthy, on our sample the standard Cue protocol did not show any relevant gain effect from the baseline evaluation.

Secondary Outcome Measures

Secondary outcome measures that improved in AOS (**Figure 3**) were as follows: severity of motor impairment (UPDRS III); motor problems, and bodily discomfort in activity of daily life (the mobility and bodily discomfort subscales of the PDQ39 questionnaire). For this pool of outcome measures, the positive effect of AOS treatment over Cue training has a great effect size ($\eta^2G > 0.30$) and is stable until the last follow-up (see *post hoc* comparisons reported in **Table 4**). Even in these measures, the

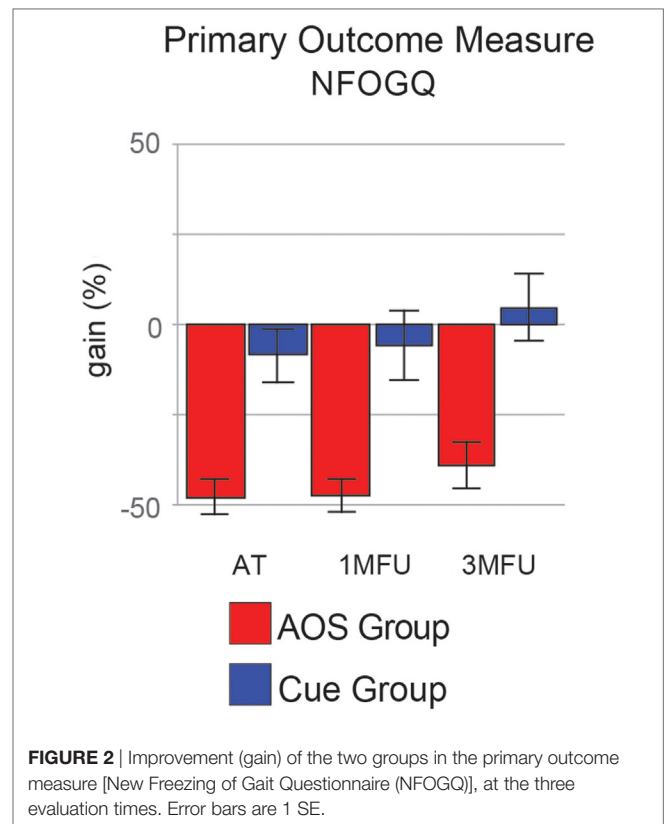


FIGURE 2 | Improvement (gain) of the two groups in the primary outcome measure [New Freezing of Gait Questionnaire (NFOGQ)], at the three evaluation times. Error bars are 1 SE.

standard Cue protocol did not show any relevant gain effect. **Figures 2** and **3** show the gain scores of the main effects.

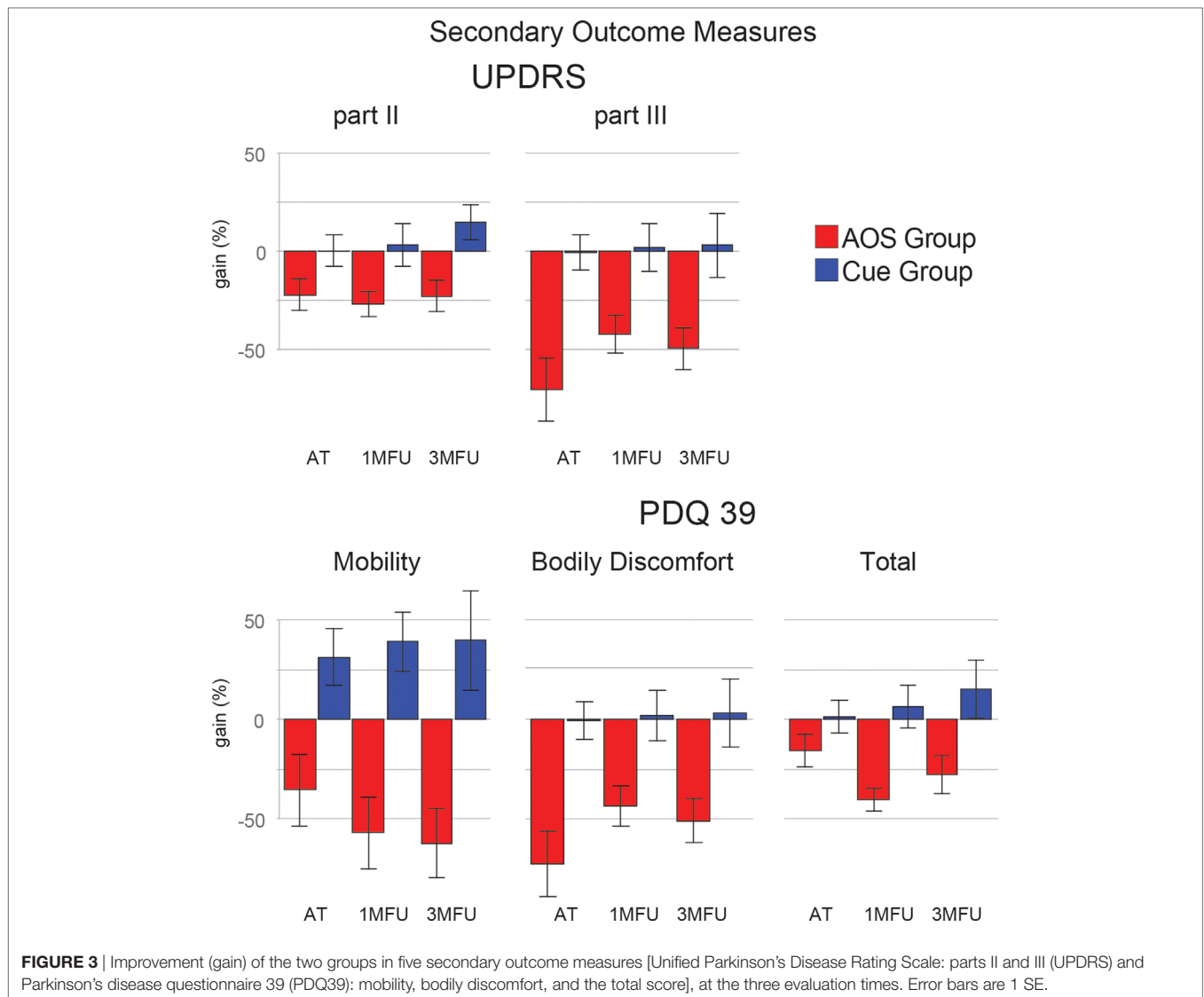
Table 3 reports F -tests for main effects and interaction separately, for all outcomes considered, whereas **Table 4** reports direct comparisons at each evaluation time, between groups.

At first glance, for nearly all secondary outcomes, the group factor (the main effect of rehabilitation protocol) had the greater effect size (η^2G). These effects are stable over time since interaction terms are not significant and/or with negligible amounts of variance explained.

The problems in activity of daily living (PDQ39 total score, UPDRS II) were significantly reduced by AOS training, with stable results also after 3 months. Moreover, AOS training determined also a small improvement on average gain scores of motor balance (BBS, see **Tables 3** and **4**).

Linear Discriminant Analysis

We trained an LDA algorithm to learn to discriminate between rehabilitation protocols attended by participants, using as input variables significant outcomes identified by ANOVA. Particularly, inclusion criteria were the following: (a) significant main effect on group factor and great effect size (>0.30), (b) stable result over evaluation time (no interaction in **Table 1** and significant *post hoc* comparisons in **Table 3**). Input variables were the raw data, not transformed into gain scores. **Figure 4A** shows average discrimination accuracy, expressed in terms of sensitivity—i.e., SDs from chance (51). At the baseline, the algorithm could not



learn a reliable criterion to recognize “AOS” or “Cue” participants since their clinical profiles are homogeneous (Table 2) and hence its performance stops to a chance level. Immediately AT, at 1 month, and after 3 months, experimental protocol differentiates participant's outcomes from the baseline levels and the algorithm can learn a criterion that move the performance (nearly) 1 SD from the chance. Importantly, the effect is far more evident considering the mobility outcome (NFOGQ and UPDRS III), over the improvement in ability of daily activities (PDQ39 mobility and bodily discomfort).

Furthermore, using the same input variables, we trained an LDA algorithm to discriminate each participant's pre- and post-training conditions, within each group; Figure 4B shows average sensitivity index, separately for the AOS and Cue conditions. Participants trained with experimental AOS protocol were discriminable with respect to their baseline condition to an extent of 2 SD from chance, using clinical motor profile, and this result is quite stable over time. The ability of the algorithm to learn

systematically a criterion over the chance was not proven, within participant's undergoing rehabilitation with Cue.

DISCUSSION

Sonification and AO are used together for the first time with the aim to treat motor diseases in PD patients with FoG. The main finding of our study is that this multisensory treatment reduces FoG (number of episodes and duration), and provided positive effects on gait pattern in short- and long-term period.

These results are in agreement with those obtained in two previous studies with the use of AO: Pelosin et al. (29) and Agosta et al. (53). In both these studies, freezing improvements (assessed with FOGQ and NFOGQ, respectively) were evaluated only up to 1 month after the end of the treatment—instead of three—and with mixed and weak results—Pelosin's data showed a significant improvement at the 1-month follow-up, but not at the end of the treatment, and the reverse pattern in the Agosta's. Our data with

TABLE 4 | Bonferroni *post hoc* comparisons AOS group vs. cue group.

	After treatment	1st Follow-up	2nd Follow-up
Primary outcome measure			
NFOGQ	$p \leq 0.001$	$p \leq 0.001$	$p \leq 0.001$
Secondary outcome measures			
PDQ39 mobility	$p \leq 0.05$	$p \leq 0.001$	$p \leq 0.001$
UPDRSIII	$p \leq 0.001$	$p \leq 0.05$	$p \leq 0.05$
PDQ39 bodily discomfort	$p \leq 0.001$	$p \leq 0.05$	$p \leq 0.05$
PDQ39 Total	ns	$p \leq 0.01$	$p \leq 0.01$
UPDRSII	ns	$p \leq 0.05$	$p \leq 0.01$
BBS	ns	$p \leq 0.05$	ns
6MWT	ns	ns	$p \leq 0.05$
TUG	ns	ns	ns
MPAS	ns	ns	ns
PDQ39 cognitions	ns	ns	ns

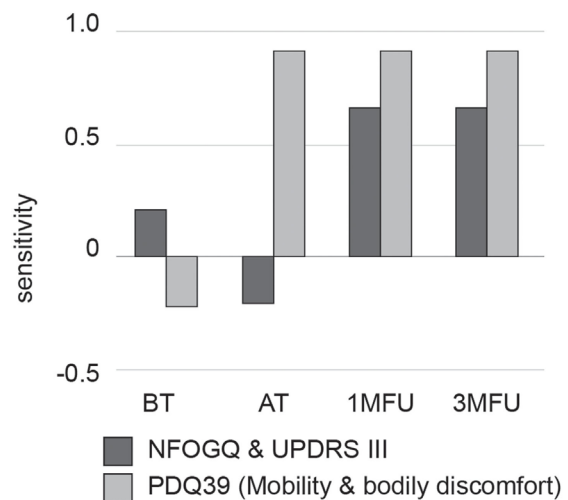
Data are *p*-levels of the direct comparisons at each evaluation time, between groups, after the Bonferroni correction.

NFOGQ, New Freezing of Gait Questionnaire; UPDRS, Unified Parkinson's Disease Rating Scale II and III; PDQ39, Parkinson's disease questionnaire 39; MPAS, Modified Parkinson's Activity Scale; BBS, Berg Balance Scale; 6MWT, 6-min walking test; TUG, Time-Up-and-Go; ns, not significant.

AO plus Sonification showed consistent and significant effects in several of the secondary outcome measures: on motor impairment (UPDRS III), and quality of life (PDQ39 mobility scale), during the entire 3-month period of evaluation; while balance (BBS), gait parameters (6MWT), only at the first and second follow-up, respectively. Overall, our data confirmed the therapeutic potential of a protocol based on AO plus Sonification in treating gait disorders and FoG. In AOS group, patients improved their mobility, acquiring new motor strategies to overcome FoG, and these effects are prolonged over time and generalized to FoG in daily life.

In Cue control group, no enhancements were found for all mobility indices, throughout the three testing times. Only data in PDQ39 questionnaire subitem mobility and activities daily living show trend values toward an improvement, that with a larger sample, could lead to significance. A possible explanation is a residual cue dependence effect, which may have not triggered an effective learning process (54), and pointing out that evidences on the effectiveness of cue trainings in the alleviation of FoG symptoms is still a hot topic. Another potential factor could be the age of our sample. Although the two groups were not statistically different in age and stage of the disease, overall our patients were quite old (73 years). Probably, older patients may require a more specific training to engage a motor consolidation process when a standard protocol based on external sensory cues is used. Indeed, it should be emphasized the lower mean age of participants with PD and FoG in previous studies [66 years—Agosta et al. (53); 66 years—Lu et al. (55); 62 years—Young et al. (9)], when compared with the ones of our research. This age difference could have produced an additional decline in motor learning (56). In a crossover design with old patients with PD and FoG (mean age, 74 years), Bunting-Perry et al. (57)—using a laser beam on a rolling walker as a visual cue—showed no significant effects in diminishing FoG and improving walking.

A LDA: discrimination between Groups



B LDA: discrimination within subjects

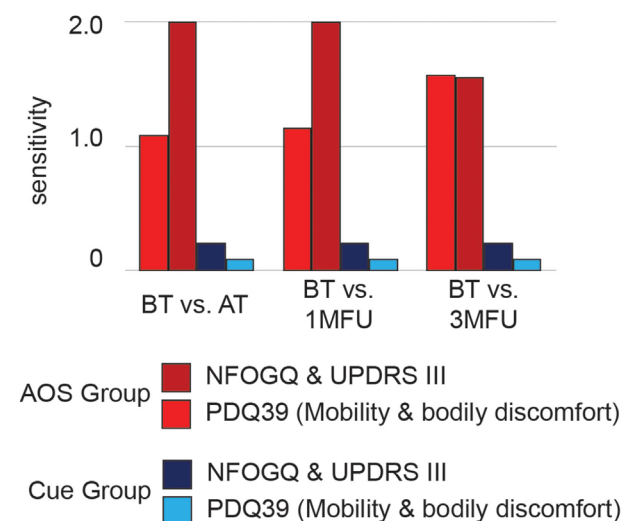


FIGURE 4 | Average discrimination accuracy expressed in terms of sensitivity (d') of the linear discriminant analysis (LDA) algorithm. **(A)** Results of testing between-groups' discrimination. Dark gray bars represent LDA results from the New Freezing of Gait Questionnaire (NFOGQ) and UPDRS III scales. Light gray bars represent LDA results from the PDQ39 mobility and bodily discomfort scales. **(B)** Results of testing within-subject discrimination. Dark and light red bars represent LDA results from the NFOGQ and UPDRS III scales for the Action Observation plus Sonification experimental group. Dark and light blue bars represent LDA results from the PDQ39 mobility and bodily discomfort scales, for the Cue control group.

The peculiar feature and novelty of our approach is the inclusion of a sonified audio track—representing kinematic features of a movement—to the video of the same movement. In other words, we used the Sonification to highlight task-intrinsic (spatial

and temporal) information, otherwise difficult to access. This *augmented* stimulus is very different from those typically employed in AO treatments, since usually the sound part is absent (29), or not related in meaning to the content of the video [not congruent multisensory stimulus as in Ref. (53)]. When a patient attends to a stimulus with a sound cue (i.e., metronome) presented together with a video of an action, the amount of cognitive resources necessary to integrate the information of the two stimuli—not related in meaning—increases [for a review, see Ref. (58)]. In our protocol, the videos are congruent multisensory stimuli, in the sense that sounds and images are related in meaning, and probably bound together at the perceptual level. This conclusion is based on patients' personal reports; they all reported to perceive stimuli as being highly consistent, and treated them as a single audiovisual event [the unity assumption—for a recent review, see Ref. (59)]. In fact, during the training, we did not need to use any particular type of instruction—except that the sound simply derived from the velocity of the movement—to let patient understand the meaning of sonified sounds, and the relation between the two sources of information (sound and image). After the presentation of the examples, the meaning of the stimuli became clear to almost all patients, and for those with some doubts, the presentation of the first stimulus was sufficient to understand.

The observation of action activates in humans the mirror neuron system (MNS) within the premotor cortex, inferior frontal gyrus, and inferior parietal lobule, that maps sensory signal onto the same neural circuits involved in motor planning and execution of the observed motor gesture. Congruent Sonification may have improved AO priming effect on movement. Indeed, during congruent audiovisual stimuli observation, Schmitz et al. (17) demonstrated an amplified activation of some of the major MNS areas, particularly frontal operculum, inferior parietal lobule and the superior temporal areas. Thanks to an enhanced perceptual analysis of the movement, congruent Sonification could lead to an improved neural representation of the observed motor action, and to an easier learning process thanks to a lightened cognitive load.

Motor learning involves the interaction of several components (60): extraction and processing of task-relevant sensory information, making decision aimed to define which movements to perform (and in which order), activating control processes, and finally a reactive and biomechanical control. A multisensory AO plus Sonification protocol might have aided the first phase of motor learning, facilitating the extraction and integration of visual and coherent auditory inputs, for a better understanding of spatial and temporal features of motor action.

Moreover, we hypothesize that our multisensory protocol—based on congruent and unitary stimuli—could have produced positive effects on memory, and specifically working memory processes. In fact, Lehmann and Murray (15) showed that semantically congruent multisensory stimuli can enhance subsequent processing and memory performance, and more recently Brunetti et al. (61) demonstrated that crossmodal correspondence (i.e., audiovisual congruent stimuli) produced faster reaction times and higher accuracy in a classical working memory task (*n*-back task). Given these findings, and given that PD patients are known to be impaired in working memory processes [see, for example, Ref. (62)],

the use of multisensory stimuli could have facilitated the processing and consequently the production of more effective gait patterns.

Sonification, as alternative, can be used in a rehabilitation program for patient with PD by generating additional real-time movement information, being suitable for integration with visual and proprioceptive perceptual feedback, while the patient is performing physical exercises. With ongoing training activity, synchronously processed auditory information should be initially integrated into the emerging internal models, enhancing the efficacy of motor learning. This is achieved by a direct mapping of kinematic and dynamic motion parameters to electronic sounds, resulting in continuous auditory and convergent audiovisual or audio-proprioceptive stimulus arrays.

A critical analysis of protocols' features emphasize that learning strategies used in the two groups could be also partially different in terms of learning mechanisms. Indeed, they are more related to a modeling process—with movement-related analogic representations—in the experimental group, while in the control group followed a cueing approach—with abstract and propositional representations. Agosta et al. (53) used a similar experimental design with a control group that underwent to a motor learning process by instructions and an experimental group that improved motor action by AO. In our control group, we used visual and auditory cue whose effectiveness had already been stressed by several studies in PD patients with FoG, and these results could be also considered as a further confirmation of the effectiveness of the AO therapy.

The combination in our AO plus Sonification protocol of a multisensory and analogic approach instead of a unisensory and abstract approach, produced promising positive effects, although we cannot evaluate nor the relative impact of each component, neither the effect of their interaction. However, this matter remains to be fully addressed.

Finally, as for the long-lasting effects, our protocol was a not intensive 8-week training program, which is not a long rehabilitative period from a motor learning and physical exercise perspective. Nevertheless, given that our results showed both immediate (upon the end of treatment), and long-term retention (3 months following cessation of treatment) of gait improvement, we may suppose that these benefits can be explained with a neuroplasticity process induced by goal-based exercises (63). As reported in previous studies (63, 64), goal-based exercise can promote neuroplasticity effects, which have been demonstrated in several neurological conditions, and also in PD, through changes in cortical excitability and cortical representation. Recently, using fMRI, Agosta et al. (53) demonstrated AO-related performance enhancement in patients with PD and FoG was possible with an intensive 4-week training program (12 sessions) and was associated with an increased activation of motor cortical areas and fronto parietal regions of the MNS.

ETHICS STATEMENT

This study was carried out in accordance with the recommendations of the "Comitato Etico Regionale Unico" guidelines, with written informed consent from all subjects. All subjects gave

written informed consent in accordance with the Declaration of Helsinki. The protocol was approved by the Institutional Ethics Committee (Comitato Etico Regionale Unico—Friuli Venezia Giulia. Protocol no. 4456—05.02.2015). Patients who agreed to participate always signed a written informed consent and they were able to leave the experiment at any moment, with no additional explanations. The study has been registered at <http://Clinicaltrials.gov>, NCT03249155.

AUTHOR CONTRIBUTIONS

Conception, and design of the research project: SM and PB. Organization of the research project: SM, PB, and PM. Execution of the research project: SM, LP, MC, BK, GF, and PB. Treatment of the patients: SM and LP. Statistical analysis and interpretation of data; writing of the manuscript first draft: SM, MG, and PB. Manuscript review and critique: SM, MG, LP, MC, BK, PM, and PB.

REFERENCES

- Kwak Y, Müller ML, Bohnen NI, Dayalu P, Seidler RD. L-DOPA changes ventral striatum recruitment during motor sequence learning in Parkinson's disease. *Behav Brain Res* (2012) 230:116–24. doi:10.1016/j.bbr.2012.02.006
- Ruitenberg MFL, Duthoo W, Santens P, Seidler RD, Notebaert W, Abrahamse EL. Sequence learning in Parkinson's disease: focusing on action dynamics and the role of dopaminergic medication. *Neuropsychologia* (2016) 93(Pt A):30–9. doi:10.1016/j.neuropsychologia.2016.09.027
- Cools R, Barker RA, Sahakian BJ, Robbins TW. Enhanced or impaired cognitive function in Parkinson's disease as a function of dopaminergic medication and task demands. *Cereb Cortex* (2001) 11(12):1136–43. doi:10.1093/cercor/11.12.1136
- Duthoo W, Braem S, Houtman F, Schoupe N, Santens P, Notebaert W. Dopaminergic medication counteracts conflict adaptation in patients with Parkinson's disease. *Neuropsychology* (2013) 27(5):556–61. doi:10.1037/a0033377
- Redgrave P, Rodriguez M, Smith Y, Rodriguez-Oroz MC, Lehericy S, Bergman H, et al. Goal-directed and habitual control in the basal ganglia: implications for Parkinson's disease. *Nat Rev Neurosci* (2010) 11(11):760–72. doi:10.1038/nrn2915
- Vandenbosche J, Deroost N, Soetens E, Coomans D, Spilidooren J, Vercruysse S, et al. Impaired implicit sequence learning in Parkinson's disease patients with freezing of gait. *Neuropsychology* (2013) 27(1):28–36. doi:10.1037/a0031278
- Ginis P, Nackaerts E, Nieuwboer A, Heremans E. Cueing for people with Parkinson's disease with freezing of gait: a narrative review of the state-of-the-art and novel perspectives. *Ann Phys Rehabil Med* (2017). doi:10.1016/j.rehab.2017.08.002
- Patla AE, Vickers JN. Where and when do we look as we approach and step over an obstacle in the travel path? *Neuroreport* (1997) 8(17):3661–5. doi:10.1097/00001756-199712010-00002
- Young WR, Shreve L, Quinn EJ, Craig C, Bronte-Stewart H. Auditory cueing in Parkinson's patients with freezing of gait. What matters most: action-relevance or cue-continuity? *Neuropsychologia* (2016) 87:54–62. doi:10.1016/j.neuropsychologia.2016.04.034
- Stefan K, Cohen LG, Duque J, Mazzocchio R, Celnik P, Sawaki L, et al. Formation of a motor memory by action observation. *J Neurosci* (2005) 25(41):9339–46. doi:10.1523/JNEUROSCI.2282-05.2005
- Gatti R, Tettamanti A, Gough PM, Riboldi E, Marinoni L, Buccino G. Action observation versus motor imagery in learning a complex motor task: a short review of literature and a kinematics study. *Neurosci Lett* (2013) 540:37–42. doi:10.1016/j.neulet.2012.11.039
- Abbruzzese G, Avanzino L, Marchese R, Pelosin E. Action observation and motor imagery: innovative cognitive tools in the rehabilitation of Parkinson's disease. *Parkinsons Dis* (2015) 2015:124214. doi:10.1155/2015/124214

ACKNOWLEDGMENTS

The authors wish to thank the patients of the Neurology Clinic—Cattinara Hospital (Trieste) for being so generous with their time and efforts.

FUNDING

This work was not supported by any third party funding or research grant.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at <http://www.frontiersin.org/articles/10.3389/fneur.2017.00723/full#supplementary-material>.

- Shams L, Seitz AR. Benefits of multisensory learning. *Trends Cogn Sci* (2008) 12(11):411–7. doi:10.1016/j.tics.2008.07.006
- Davidson S, Cronin-Golomb A, Lee A. Visual and spatial symptoms in Parkinson's disease. *Vision Res* (2005) 45(10):1285–96. doi:10.1016/j.visres.2004.11.006
- Lehmann S, Murray MM. The role of multisensory memories in unisensory object discrimination. *Brain Res Cogn Brain Res* (2005) 24(2):326–34. doi:10.1016/j.cogbrainres.2005.02.005
- Von Kriegstein K, Giraud AL. Implicit multisensory associations influence voice recognition. *PLoS Biol* (2006) 4(10):e326. doi:10.1371/journal.pbio.0040326
- Schmitz G, Mohammadi B, Hammer A, Heldmann M, Samii A, Münte TF, et al. Observation of sonified movements engages a basal ganglia frontocortical network. *BMC Neurosci* (2013) 14(1):32. doi:10.1186/1471-2202-14-32
- Kim RS, Seitz AR, Shams L. Benefits of stimulus congruency for multisensory facilitation of visual learning. *PLoS One* (2008) 3(1):e1532. doi:10.1371/journal.pone.0001532
- Silva AE, Barakat BK, Jimenez LO, Shams L. Multisensory congruency enhances explicit awareness in a sequence learning task. *Multisensor Res* (2017) 30(7–8):681–9. doi:10.1163/22134808-00002587
- Sigrist R, Rauter G, Rieger R, Wolf P. Augmented visual, auditory, haptic, and multimodal feedback in motor learning: a review. *Psychon Bull Rev* (2013) 20(1):21–53. doi:10.3758/s13423-012-0333-8
- Rodger MWM, Young WR, Craig CM. Synthesis of walking sounds for alleviating gait disturbances in Parkinson's disease. *IEEE Trans Neural Syst Rehabil Eng* (2014) 22(3):543–8. doi:10.1109/TNSRE.2013.2285410
- Effenberg AO. Movement sonification: effects on perception and action. *IEEE Multimedia* (2005) 12(2):53–9. doi:10.1109/MMUL.2005.31
- Hughes AJ, Daniel SE, Kilford L, Lees AJ. Accuracy of clinical diagnosis of idiopathic Parkinson's disease: a clinico-pathological study of 100 cases. *J Neurol Neurosurg Psychiatry* (1992) 55(3):181–4. doi:10.1136/jnnp.55.3.181
- Nieuwboer A, Rochester L, Herman T, Vandenberghe W, Emil GE, Thomas T, et al. Reliability of the new freezing of gait questionnaire: agreement between patients with Parkinson's disease and their carers. *Gait Posture* (2009) 30(4):459–63. doi:10.1016/j.gaitpost.2009.07.108
- Goetz CG, Poewe W, Rascol O, Sampaio C, Stebbins GT, Counsell C, et al. Movement Disorder Society Task Force report on the Hoehn and Yahr staging scale: status and recommendations. *Mov Disord* (2004) 19:1020–8. doi:10.1002/mds.20213
- Beck AT, Steer RA, Brown GK. *Beck Depression Inventory: Manual BDI-II*. New York: Psychological Corporation (1996).
- Arevalo-Rodriguez I, Smailagic N, Roqué I, Figuls M, Ciapponi A, Sanchez-Perez E, et al. Mini-Mental State Examination (MMSE) for the detection of Alzheimer's disease and other dementias in people with mild cognitive impairment (MCI). *Cochrane Database Syst Rev* (2015) 25(3):CD010783. doi:10.1002/14651858.CD010783.pub2

28. Bandura A. *Social Learning Theory*. Toronto, Canada: Prentice-Hall (1977).
29. Pelosin E, Avanzino L, Bove M, Stramesi P, Nieuwboer A, Abbruzzese G. Action observation improves freezing of gait in patients with Parkinson's disease. *Neurorehabil Neural Repair* (2010) 24(8):746–52. doi:10.1177/1545968310368685
30. Peto V, Jenkinson C, Fitzpatrick R. PDQ-39: a review of the development, validation and application of a Parkinson's disease quality of life questionnaire and its associated measures. *J Neurol* (1998) 245(1):S10–4. doi:10.1007/PL00007730
31. Keus SHJ, Nieuwboer A, Bloem BR, Borm GF, Munneke M. Clinimetric analyses of the Modified Parkinson Activity Scale. *Parkinsonism Relat Disord* (2009) 15(4):263–9. doi:10.1016/j.parkreldis.2008.06.003
32. Podsiadlo D, Richardson S. The timed "Up & Go": a test of basic functional mobility for frail elderly persons. *J Am Geriatr Soc* (1991) 39(2):142–8. doi:10.1111/j.1532-5415.1991.tb01616.x
33. Falvo MJ, Earhart GM. Six-minute walk distance in persons with Parkinson disease: a hierarchical regression model. *Arch Phys Med Rehabil* (2009) 90(6):1004–8. doi:10.1016/j.apmr.2008.12.018
34. Berg KO, Wood-Dauphinee SL, Williams JL, Maki B. Measuring balance in the elderly: validation of an instrument. *Can J Public Health* (1992) 83:S7–11.
35. Nasreddine ZS, Phillips NA, Bédirian V, Charbonneau S, Whitehead V, Collin I, et al. The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment. *J Am Geriatr Soc* (2005) 53(4):695–9. doi:10.1111/j.1532-5415.2005.53221.x
36. Orsini A, Grossi D, Capitani E, Laiacina M, Papagno C, Vallar G. Verbal and spatial immediate memory span: normative data from 1355 adults and 1112 children. *Ital J Neurol Sci* (1987) 8(6):537–48. doi:10.1007/BF02333660
37. De Renzi E, Nichelli P. Verbal and non-verbal short-term memory impairment following hemispheric damage. *Cortex* (1975) 11(4):341–54. doi:10.1016/S0010-9452(75)80026-8
38. Carlesimo GA, Buccione I, Fadda L. Normative data of two memory tasks: short-story recall and Rey's figure. *Nuova Rivista Di Neurologia* (2002) 12(1):1–13.
39. Della Sala S, Laiacina M, Spinnler H, Ubezio C. A cancellation test: its reliability in assessing attentional deficits in Alzheimer's disease. *Psychol Med* (1992) 22(4):885–901. doi:10.1017/S0033291700038460
40. Barbarotto R, Laiacina M, Frosio R, Vecchio M, Farinato A, Capitani E. A normative study on visual reaction times and two Stroop colour-word tests. *Ital J Neurol Sci* (1998) 19(3):161–70. doi:10.1007/BF00831566
41. Giovagnoli AR, Del Pesce M, Mascheroni S, Simoncelli M, Laiacina M, Capitani E. Trail making test: normative values from 287 normal adult controls. *Ital J Neurol Sci* (1996) 17(4):305–9. doi:10.1007/BF01997792
42. Iavarone A, Ronga B, Pellegrino L, Loré E, Vitaliano S, Galeone F, et al. The Frontal Assessment Battery (FAB): normative data from an Italian sample and performances of patients with Alzheimer's disease and frontotemporal dementia. *Funct Neurol* (2004) 19(3):191–5.
43. Krikorian R, Bartok J, Gay N. Tower of London procedure: a standard method and developmental data. *J Clin Exp Neuropsychol* (1994) 16(6):840–50. doi:10.1080/01688639408402697
44. Belacchi C, Scalisi TG, Cannoni E, Cornoldi C. *CPM-Coloured Progressive Matrices. Standardizzazione Italiana*. Firenze: Giunti – Organizzazioni Speciali (2008).
45. Puckette M. *Pd Documentation*. (2017). Available from: http://msp.ucsd.edu/Pd_documentation/
46. Henkelmann C. *Improving the Aesthetic Quality of Realtime Motion Data Sonification* (No. CG-2007-4). Bonn, Germany: Universität Bonn (2007).
47. Henkelmann C, Zinke A, Weber A. Towards aesthetic realtime sonification of motion data. Presented at the 4. Workshop Virtuelle Realität und Augmented Reality der GI-Fachgruppe VR/AR 2007. Bonn, Germany: Weimar (2007).
48. Olejnik S, Algina J. Generalized eta and omega squared statistics: measures of effect size for some common research designs. *Psychol Methods* (2003) 8(4):434–47. doi:10.1037/1082-989X.8.4.434
49. Bakeman R. Recommended effect size statistics for repeated measures designs. *Behav Res Methods* (2005) 37(3):379–84. doi:10.3758/BF03192707
50. Moreau D, Kirk IJ, Waldie KE. Seven pervasive statistical flaws in cognitive training interventions. *Front Hum Neurosci* (2016) 10:153. doi:10.3389/fnhum.2016.00153
51. Macmillan NA, Creelman CD. *Detection Theory: A User's Guide*. Mahwah, New Jersey: Laurence Erlbaum Associates (2005).
52. R Development Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing (2017). Available from: <http://www.R-project.org>
53. Agosta F, Gatti R, Sarasso E, Volonté MA, Canu E, Meani A, et al. Brain plasticity in Parkinson's disease with freezing of gait induced by action observation training. *J Neurol* (2017) 264(1):88–101. doi:10.1007/s00415-016-8309-7
54. Nieuwboer A, Rochester L, Müncks L, Swinnen SP. Motor learning in Parkinson's disease: limitations and potential for rehabilitation. *Parkinsonism Relat Disord* (2009) 15(Suppl 3):S53–8. doi:10.1016/S1353-8020(09)70781-3
55. Lu C, Amundsen Huffmaster SL, Tuite PJ, Vachon JM, MacKinnon CD. Effect of cue timing and modality on gait initiation in Parkinson disease with freezing of gait. *Arch Phys Med Rehabil* (2017) 98(7):1291–9.e1. doi:10.1016/j.apmr.2017.01.009
56. Seidler RD, Bernard JA, Burutolu TB, Fling BW, Gordon MT, Gwin JT, et al. Motor control and aging: links to age-related brain structural, functional, and biochemical effects. *Neurosci Biobehav Rev* (2010) 34(5):721–33. doi:10.1016/j.neubiorev.2009.10.005
57. Bunting-Perry L, Spindler M, Robinson KM, Noorigian J, Cianci HJ, Duda JE. Laser light visual cueing for freezing of gait in Parkinson disease: a pilot study with male participants. *J Rehabil Res Dev* (2013) 50(2):223–30. doi:10.1682/JRRD.2011.12.0255
58. Spence C. Audiovisual multisensory integration. *Acoust Sci Technol* (2007) 28(2):61–70. doi:10.1250/ast.28.61
59. Vatakis A, Spence C. Crossmodal binding: evaluating the "unity assumption" using audiovisual speech stimuli. *Percept Psychophys* (2007) 69(5):744–56. doi:10.3758/BF03193776
60. Wolpert DM, Diedrichsen J, Flanagan JR. Principles of sensorimotor learning. *Nat Rev Neurosci* (2011) 12:739–51. doi:10.1038/nrn3112
61. Brunetti R, Indraccolo A, Mastroberardino S, Spence C, Santangelo V. The impact of cross-modal correspondences on working memory performance. *J Exp Psychol Hum Percept Perform* (2017) 43(4):819–31. doi:10.1037/xhp0000348
62. Hall JM, Shine JM, Walton CC, Gilat M, Kamsma YPT, Naismith SL, et al. Early phenotypic differences between Parkinson's disease patients with and without freezing of gait. *Parkinsonism Relat Disord* (2014) 20(6):604–7. doi:10.1016/j.parkreldis.2014.02.028
63. Petzinger GM, Fisher BE, McEwen S, Beeler JA, Walsh JP, Jakowec MW. Exercise-enhanced neuroplasticity targeting motor and cognitive circuitry in Parkinson's disease. *Lancet Neurol* (2013) 12(7):716–26. doi:10.1016/S1474-4422(13)70123-6
64. Forrester LW, Hanley DF, Macko RF. Effects of treadmill exercise on transcranial magnetic stimulation-induced excitability to quadriceps after stroke. *Arch Phys Med Rehabil* (2006) 87(2):229–34. doi:10.1016/j.apmr.2005.10.016

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.

Copyright © 2018 Mezzarobba, Grassi, Pellegrini, Catalan, Kruger, Furlanis, Manganotti and Bernardis. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

APPENDIX

Exercise 1

Shifting the body weight in the frontal plane and taking a step—Actor stands straight up, with both feet on the floor, shifting the body weight to the right (or to the left), to the left (or to the right), and then raise and move forward the right (or the left) leg and the body to take the first step.

Exercise 2

Shifting the body weight in the sagittal plane and taking a step—Actor stands straight up with both feet on the floor. One foot placed in front of the other, with the heel ahead of the other foot's toes. The actor shifts weight from one foot to the other, always keeping the feet on the floor; afterwards he takes a step forward.

Exercise 3

Gait initiation—Actor starts to walk with the preferred leg.

Exercise 4

Turning around—Actor walks two steps with a straight trajectory, and then made a 180° turn in a narrow quarter (U-turn).

Exercise 5

Stepping over an obstacle—Actor walks three steps with a straight trajectory, and then steps over the obstacle (obstacle's height: 10% of patient's height).

Exercise 6

Sit-to-walk—Actor is seated on a backless and armless stool (knee angle 100°), and then raises and walks three steps forward.

Exercise 7

Walking straight with long steps—Actor walks about 10 long steps with a straight trajectory trying to maintain a steady pace and to take long steps.

Exercise 8

Walking through a doorway—Actor walks three steps with a straight trajectory, moves through a real doorway without stopping and then continue to walk two more steps.



Interactive Sonification Exploring Emergent Behavior Applying Models for Biological Information and Listening

Insook Choi*

Studio for International Media & Technology, MediaCityUK, School of Arts & Media, University of Salford, Manchester, United Kingdom

OPEN ACCESS

Edited by:

Diego Minicacchi,
Università degli Studi di Firenze, Italy

Reviewed by:

Waldemar Karwowski,
University of Central Florida,
United States
Hiroaki Wagatsuma,
Kyushu Institute of Technology, Japan

*Correspondence:

Insook Choi
insook@insookchoi.com

Specialty section:

This article was submitted to
Neuroprosthetics,
a section of the journal
Frontiers in Neuroscience

Received: 04 October 2017

Accepted: 12 March 2018

Published: 27 April 2018

Citation:

Choi I (2018) Interactive Sonification Exploring Emergent Behavior Applying Models for Biological Information and Listening. *Front. Neurosci.* 12:197. doi: 10.3389/fnins.2018.00197

Sonification is an open-ended design task to construct sound informing a listener of data. Understanding application context is critical for shaping design requirements for data translation into sound. Sonification requires methodology to maintain reproducibility when data sources exhibit non-linear properties of self-organization and emergent behavior. This research formalizes interactive sonification in an extensible model to support reproducibility when data exhibits emergent behavior. In the absence of sonification theory, extensibility demonstrates relevant methods across case studies. The interactive sonification framework foregrounds three factors: reproducible system implementation for generating sonification; interactive mechanisms enhancing a listener's multisensory observations; and reproducible data from models that characterize emergent behavior. Supramodal attention research suggests interactive exploration with auditory feedback can generate context for recognizing irregular patterns and transient dynamics. The sonification framework provides circular causality as a signal pathway for modeling a listener interacting with emergent behavior. The extensible sonification model adopts a data acquisition pathway to formalize functional symmetry across three subsystems: Experimental Data Source, Sound Generation, and Guided Exploration. To differentiate time criticality and dimensionality of emerging dynamics, *tuning functions* are applied between subsystems to maintain scale and symmetry of concurrent processes and temporal dynamics. Tuning functions accommodate sonification design strategies that yield order parameter values to render emerging patterns discoverable as well as *rehearsable*, to reproduce desired instances for clinical listeners. Case studies are implemented with two computational models, Chua's circuit and Swarm Chemistry social agent simulation, generating data in real-time that exhibits emergent behavior. *Heuristic Listening* is introduced as an informal model of a listener's clinical attention to data sonification through multisensory interaction in a context of structured inquiry. Three methods are introduced to assess the proposed sonification framework: Listening Scenario classification, data flow Attunement, and Sonification Design Patterns to classify sound control. Case study implementations are assessed against these methods comparing levels of abstraction between experimental data and sound generation. Outcomes demonstrate the framework performance as a reference model

for representing experimental implementations, also for identifying common sonification structures having different experimental implementations, identifying common functions implemented in different subsystems, and comparing impact of affordances across multiple implementations of listening scenarios.

Keywords: sonification, listening, emergent behavior, interaction design, cognitive cycle, supramodal attention, biological information, media psychology

INTRODUCTION: WHAT DO WE LISTEN TO WHEN WE LISTEN TO DATA?

Sonification is an open-ended design task. Methods differ based on applications. Understanding the application context is critical for shaping listening scenarios and design requirements and subsequent choice of data translation strategies and sound production. In cases where the experimental data source is predictable in terms of well-defined data dimensions and boundaries, the relationship between system parameters and sound parameters can be relatively linear. However, when the experimental data source is unpredictable and the data exhibits emergent behavior, sonification requires a methodology to establish reliable rendition of the dynamics.

In the absence of established sonification theory, the field is challenged by *ad-hoc* variance in instruments, implementations and interpretations, limiting the scalability of case study results. This research examines the rationale and feasibility to formalize an extensible model for interactive sonification, applied to data that exhibits emergent behavior. An extensible model is proposed as an interactive sonification framework foregrounding three factors: reproducible system implementation for generating sonification, reproducible data from models that characterize emergent behavior, and interactive mechanisms enhancing a listener's multisensory observations. Using auditory sensation to seek biological information dates back to ancient Greek times when physicians monitored pulses for a clinical diagnosis (Kaniusas, 2012). The importance of engaging all senses—to see, feel, and hear—was emphasized in order to recognize patterns and put them to use (Castiglioni, 1947). Diagnoses by listening, termed *auscultation* from the nineteenth century (Laennec and Forbes, 1835), is limited to natural signals exhibiting amplitude and frequency accessible to human hearing. Patients are observed through multisensory interaction including touch and visual inspection (see Appendix 1 in Supplementary Material). The work presented here maintains a multisensory and multimodal approach in configuring sound to convey signatures of non-linear behavior that are characteristic of biological information. Today the auditory monitoring of physiological states is extended with modern equipment and digital signal processing inevitably introducing layers of artifacts. The very focus, or rather the intent of the motivation to listen to biological information when working with extended instrumentation and digital abstraction, is what this presentation aims to be in service of.

Experimental observation uses various methodologies to obtain information from a data source external to the observing system. To make sense of information the observing system

performs measurements in order to gain insights about the states of the observed data source. Perceptualization implies transformation of data to yield observable information. To interpret data it is necessary to understand the transformations that bring the data into an observable form. “The algorithms that transform data are at the heart of data visualization.” (Hansen and Johnson, 2005, p. 1). Observers acquire the ability to interpret data by becoming familiar with the instruments that transform data, for example a doctor performs auscultation by understanding the functional attributes of the stethoscope, which informs interpretation of the transmitted sounds. **Figure 1A** illustrates the auscultation functional pathway. Sonification applies transformations requiring digital signal processing of machine-readable information, a functional pathway illustrated in **Figure 1B**, generalizing and digitizing the auscultation pathway. Since no pathway from information to data and back to information is immune to idiosyncratic complexity, the simplest definition is: *Sonification is a construction of sound that informs a listener of data.* In concept, sonification informed by biological data returns sounds that carry information about that biological system.

The present research aims to develop and extend sonification methods for data that exhibits emergent behaviors, addressing cases where reproducibility of covariance is quasi-deterministic for sounds and a corresponding data source. The research examines the viability of using models of emergent behavior to develop sonification methods that may be applied to multiple cases of biological information. Results presented here provide an example of using data models to formalize a sonification method to enable application with more than one type of data and more than one type of sound production. Following are assumptions based on sonification methods adopted in this research.

1. Sounds are audible because acoustic energy changes over time. Sound energy varies in multiple audible attributes, described in terms such as pitch, loudness, tone quality, duration, silence, repetition, and pattern. Depending upon the type of sound production, the number of audible attributes may be different than the number of sound control parameters required for those attributes.
2. Sonification aims for variance in sound to correspond to variance in the data represented by the sound. To sonify variance of data values requires a coherent model for covariance of sound control parameters. The goal is to enable a listener to identify salient features in sounds and relate these to salient features in the data.
3. The coupling of data to sound production requires four multidimensional component signals: the sound, having m

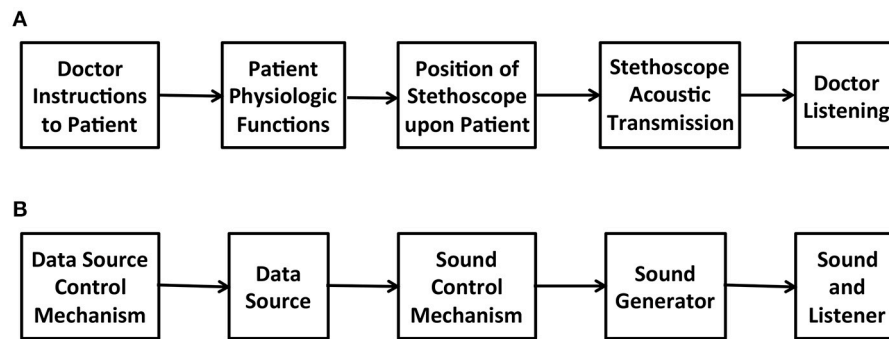


FIGURE 1 | (A) Auscultation functional pathway including physician interaction with patient to induce physiological changes that generate audible differences. **(B)** Interactive sonification functional pathway comprised of requisite system components. The illustration shows parallel structure to auscultation **(A)**.

audible attributes; the sound control signal, having n control parameters; the data to be sonified, having p dimensions; and the control signal applied to the data source, having q control parameters. These component signals have independent numbers of dimensions determined by the types of data, sound production, and observer interaction.

To establish a method for extensibility of results the coupling of sonification component signals is defined as a sonification model comprised of a series of functions that meet the following requirements:

- the sonification model is functional for multiple data sources, types of sound production, and modes of observer interaction;
- the model maintains its structure while the dimensionality of the component signals may change;
- the model is extensible to many applications as a sonification framework.

The present research applies a candidate framework to sonify two models of data sources that exhibit emergent behavior, Chua's circuit and Swarm Chemistry.

Biological systems are complex dynamical systems that may exhibit emergent behavior. Emergent behavior produces salient features in data that can vary independently of the control state of the data source. When applying sonification to a data source that exhibits emergent behavior, the aggregate coupling of sonification components may produce inconsistent correspondence of data features and sound features. To develop robust correspondence of sound to data, this research adopts data sources comprised of models that exhibit emergent behavior. Two sonification methods are presented here, one using stable regions in the data source to generate bounding reference sounds for unstable emergent regions, the other using automated feature recognition.

The use of models of emergent behavior for sonification test cases is adopted from research practices for measuring biological signals. When biological information is acquired experimentally, computational models are often used to ensure the relevance of the data and provide quality assurance for unstable and transient experimental conditions (James and Hesse, 2005). Simulations aid discernment and interpretation of transduced data, providing stable reference measurements for developing models of

experimental physiological states. Interpreting neurological impulse patterns, Faure and Korn report, "The methods used in each of these studies have almost invariably combined the analysis of experimental data with simulations using formal models." (Faure and Korn, 2003, p. 787). Biosignals can be modeled in terms of signature dynamic properties, for example simulation of spiking and bursting of cortical neurons (Izhikevich, 2003). Liljenström applies simulation of non-linear circuit oscillations to reproduce non-linear dynamics measured in brain signals, demonstrating chaotic oscillations as highly efficient for neural information processing. Simulations can be measured to demonstrate high sensitivity to input stimulus and rapid convergence on stable oscillations that may represent learned patterns (Liljenström, 2010, 2012).

In line with the use of computational models in experimental observation, the work reported here was performed with simulations recognized as paradigms for modeling biological signals. The rationale for selecting test cases is to identify models with properties that represent a broad range of applications. Emergent behaviors create non-deterministic conditions for sonification covariance with data pattern formation. Unpredictability limits the reliability of salient features' correspondence in data and sound. To address this the research applies models of data sources that exhibit emergent pattern formation. Wiener characterizes a pattern as an arrangement of elements, where the distinctness of the pattern is characterized by the order among the elements rather than the intrinsic qualities of individual elements (Wiener, 1954). This relationship can be quantitatively expressed as a set of higher-level order parameters, the concept introduced by Haken, which describes enslaving the behaviors of ensembles of elements by which patterns are formed (Haken, 1983). Biological systems in diverse areas of study have been observed to exhibit such ensembles' emergent properties. Examples are interaction patterns of groups of neurons expressed in the patterns of bursting (Wang and Rinzl, 1998), voltage oscillations in muscle fibers (Morris and Lecar, 1981), the patterns of clusters of autonomous agents at multicellular level such as behaviors demonstrated in Globally Coupled Maps (Kaneko, 2015) and multi-organism levels such as swarming and flocking behaviors of insects and birds

(Charbonneau et al., 2013). These classes of examples suggest two paradigmatic models: oscillation and agents. These two levels of abstraction cover a wide range of cases for sonification of emergent behavior. The case of Chua's circuit is selected to demonstrate the oscillation model. Chua's circuit is a well-studied paradigm of non-linear dynamics (Chua, 2005), as it exhibits signal properties from periodic to chaotic as observed in many organisms including the brains of vertebrates and humans. The case of Swarm Chemistry is selected to demonstrate the agent model. Swarm Chemistry is an interactive evolutionary computing (IEC) framework for studying collective behaviors of self-organizational agents implemented as heterogeneous swarm simulation (Sayama, 2007). Chua's circuit and Swarm Chemistry exhibit dominant multi-paradigms of non-linear behaviors and yield emergent characteristics representative of biological information. In neurosciences, "Overall, both theoretical and experimental works in the field seem to demonstrate that the advanced tools of non-linear analysis can much more accurately describe and represent the complexity of brain dynamics than traditional mathematical and computational methods based on linear and deterministic analysis (Mattei, 2014, p. 1)." Incorporating the breadth of these paradigms, an adaptable sonification framework can be assessed for capacity to generate sounds that represent non-deterministic properties such as emergent features and patterns.

Perceptualization of data requires assumptions about observers that may be formalized as a model of an observer. Listener interaction with the data source provides a context for an informal model of an observer in a sonification framework. Interaction provides a frame of reference for a listener to identify emergent behavior by comparing changes in the state of the data source and behaviors exhibited in the data. A listener interacting with a data source can determine whether observed behaviors are emergent properties or are controllable by direct manipulation of the data source. Emergent behavior is more difficult to disambiguate if a listener is not interacting with the data source during observation. Models of observer interaction are required to support extensible outcomes of user assessment of sonification test cases. The hypothesized sonification framework includes a normalized representation of observer interaction assessed across multiple applications.

MATERIALS AND MODELS FOR INTERACTIVE SONIFICATION OF EMERGENT BEHAVIOR

This research applies an experimental configuration for user interaction and models implemented as computational simulations. Section Two Dynamical Systems Models: Commonalities and Differences introduces Chua's circuit and Swarm Chemistry, two computational models implemented as real-time applications with interactive control, to generate sample data that exhibits emergent behavior. A model of a data acquisition pathway is introduced in section Data Acquisition Model for Sonification Signal Processing, and applied in section Extended Model for a Sonification Framework for testing

the hypothesis of extensible sonification modeling. In section Heuristic Listening and the LIDA Model the LIDA model is consulted to develop criteria accounting for a listener's disposition toward interactive sonification.

Physical materials required for this research include an instrument configuration for use case trials. The experimental configuration provides sound synthesis, two-channel stereo audio display, interface devices including computer mouse and large-format touch screen, computer graphic display of graphical user interfaces, and large-format display of data visualization. Case studies apply real-time interactive simulations in multisensory configurations. Sonification and visualization are synchronized with user generated interactive signals.

Sonification components are implemented as three subsystems for concurrent asynchronous processing: a data source to be rendered in sound, a sound production subsystem, and an observer interface for interactive exploration. Data transmission occurs at 10–20 Hz for user generated interactive signals and system control signals; visual display is refreshed at 24 Hz and sound is generated at 44.1 kHz per stereo channel. A scheduler ensures concurrent real-time responses across asynchronous processes.

Two Dynamical Systems Models: Commonalities and Differences

A dynamical system is iterative based on a numerical model that defines state, initial conditions and system control parameters. The simulation outputs data as a time series signal exhibiting dynamical system behavior defined as trajectories in a multidimensional phase space. The phase space represents all possible signal states. Control parameters define all possible states in control space. Control parameter values determine system conditions but may not directly determine system outputs. Complex dynamical systems exhibit self-organizing and emergent behaviors that are not predictable from previous states or from control parameter values (Silva, 1993).

Webber and Zbilut state that physiological systems can be best characterized as complex dynamical processes (Webber and Zbilut, 1994). They apply the lessons learned from complex systems theory that simple structure from a low dimensional network may generate a wide range of patterns with little experimental preparation. Chua's circuit and Swarm Chemistry exhibit this property. Chua's circuit has seven control parameters and Swarm Chemistry has six behavioral parameters. Both models exhibit emergent properties at multiple time scales. Salient features recur at periodic and aperiodic intervals, with short patterns sometimes embedded in long patterns. The Chua's circuit is a non-linear oscillator that generates a continuous-time signal from seven circuit components. Swarm Chemistry animates movements of hundreds of self-propelled agents along individual paths, defined in a bounded plane. For sound production Chua's circuit signal can be scaled to a human audible range whereas the Swarm Chemistry data comprises points in space that cannot directly produce sound. This contrast differentiates requirements for two approaches to sonification and offers insights to establishing a common framework. In both

cases the sonification couples simulation dynamics with sound generation for listeners' exploratory interaction. Section Results: Applying the Proposed Sonification Framework to Multiple Experimental Systems presents differences in sonification design corresponding to structural differences of the simulations.

Data Acquisition Model for Sonification Signal Processing

To test the hypothesis of extensible sonification modeling, a common model of data acquisition pathway is adopted for all sonification component signals and applied to all use cases tested. The model is based on established practice and given rigorous uniform application to test the hypothesis that *a model can identify symmetry of functional requirements to represent all sonification components*. The data acquisition model, illustrated in **Figure 2** provides five processes that connect control data to a signal generator and then to sample data, and perform data stream conversion from control rate iteration, to signal frequency, to sample rate iteration. The model in **Figure 2** provides a basic and extensible organization.

- At the center of the data acquisition model a signal generator provides a source of sample data.
- Control data enters the model through a multidimensional transfer function TF_{input} and is applied to a set of parameters that control the signal generator.
- The resulting signal is sampled to acquire data that characterizes the signal, and this data is passed to a multidimensional transfer function TF_{output} .

This design exposes and formalizes the functional requirements for managing differences in data dimensions and in temporal definition, which may occur between control signals and sampling processes. Frequency differences may result in serial oversampling or undersampling between processes, requiring adjustment to eliminate aliasing. Collectively these frequencies determine the overall frequency required for input data to affect output data in an implementation of the model. Sonification modeling formalizes over- and under-sampling differences among its temporal dynamics. Formal definition and systematic management of data dimensions and temporal differences are required to establish extensibility of sonification models to multiple application domains.

Extended Model for a Sonification Framework

A key research hypothesis is to establish an extensible sonification model. To implement this hypothesis the data acquisition model introduced above is adopted threefold to formalize each of three sonification subsystems: data source, sound production, and observer interface for interactive exploration. The model represents concurrent processes within a subsystem and by extension concurrent processes across the full sonification implementation. The three subsystems are referred to as Experimental Data Source, Sound Generation, and Guided Exploration. In the proposed sonification model the subsystems are arranged in series with the output of one applied to the input

of the next: output of Experimental Data Source is input to Sound Generation, the output of which is input to Guided Exploration, the output of which is input to control the Experimental Data Source. The iterative and concurrent nature of all processes can be represented as circularity of inputs and outputs, illustrated in **Figure 3**. Using the data acquisition reference model, a transfer function TF_n applies to each subsystem input and output. In circularity, each subsystem output TF_n performs a duplicate transfer function as the input of the next subsystem in the series.

Figures 4A–C apply the data acquisition model to the particulars of each subsystem. In the Experimental Data Source subsystem (**Figure 4A**) data entering from TF_3 is applied to the control space of Chua's circuit or Swarm Chemistry. The phase space exhibits stable states, transition states and emergent behaviors, which are sampled to acquire relevant features. The sample space is output to TF_1 . In the Sound Generation subsystem (**Figure 4B**) data entering from TF_1 is applied to the sound control space, generating digital audio signals that are sampled to generate audible information output to TF_2 . In the Guided Exploration subsystem (**Figure 4C**) a listener acquires sound represented at TF_2 , and responds by manipulating an interface to explore the system through listening. The exploration signal is sampled and output to TF_3 , to be applied to the control space of the Experimental Data Source (**Figure 4A**).

To include a listener model, TF_2 represents heuristic analysis rather than a formal computational model of listening. A listener performs heuristic analysis of audible signals received at TF_2 to identify expected features that represent known data states, transformations and deviations. The term *heuristic listening* is introduced to describe a cognitive process that connects a listener's auditory percept to expectations and action planning (section Heuristic Listening and the LIDA Model).

To summarize the extended sonification model: circular causality is provided through three component subsystems that collectively form a round trip of data elicitation, audible interpretation, and listener response (**Figure 3**). Each subsystem is defined as a data acquisition pathway consisting of control space, phase space, and sample space. The data acquisition reference model recognizes functional tripartite symmetry of the subsystems. The data types passing through each subsystem are different but their data pathways share common functional relationships for generating, acquiring and interpreting data. This extended model is proposed as an Interactive Sonification Framework.

Heuristic Listening and the LIDA Model

A model of listeners' engagement with sound provides essential context in the production of sound to represent data. The model considered here addresses both preattentive and attentional hearing as well as multisensory affects upon listening. Neurophysiological study of auditory attention identifies mechanisms underlying interactive listening experiences, including attentive and pre-attentive processing, and top-down vs. bottom-up interplay of attentional mechanisms (Fritz et al., 2007). Study of neuronal signals indicates that auditory processing is influenced by conscious focus of attention

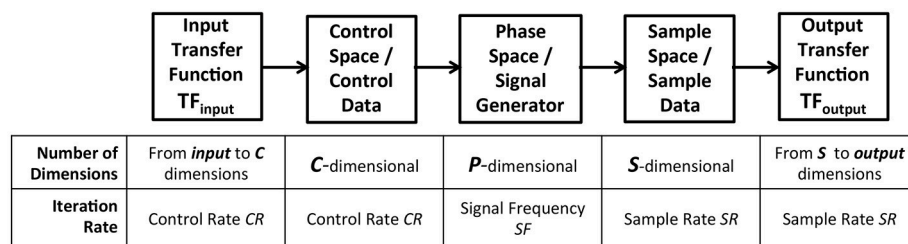


FIGURE 2 | Data-elicitation pathway applied as a model to formalize the functions of subsystems required for sonification. Each component in the pathway is dynamic, receives input data, and generates output data. The data of each component may be represented by a unique number of dimensions. Each component is dynamic with a periodic iteration at one of three frequencies: a control rate, a signal frequency and a sample rate. TF_{input} receives data in *input* dimensions and generates *c*-dimensional data required for the control signal. Control space defines *c* dimensions for system control with minimum and maximum bounding values on each dimension, defining all possible states of control for signal generation. Phase space is multidimensional with *p* dimensions for system variables specifying the instantaneous state of the system output. Phase space encompasses all possible states of the output signal. Sample space represents a parameterized multivariate stream of digital data that discretizes the phase space signal. Sample space represents the signal in discrete time steps with a set of *s* values at each step; the sampling method and data format vary with each subsystem. TF_{output} receives the sample data in *s* dimensions and generates *output* dimensional data required downstream. Control space, phase space, and sample space define periodic iterations, indicated as control rate *CR*, signal frequency *SF*, and sample rate *SR*. Frequencies of these three periods may vary independently and are concurrent within the model. For example Chua's circuit has a control rate of 15–20 Hz, signal frequency rate of 20 kHz, and data sample rate of 44.1 kHz, while Swarm Chemistry has a control rate of 10 Hz, signal frequency rate of 20 Hz, and data sample rate of 12–15 Hz.

and expectation (Brechmann and Scheich, 2005; Voisin et al., 2006; Sussman et al., 2007). Expectation is a temporal process that may be largely supramodal (Nagarajan et al., 1998; Ivry and Spencer, 2004), meaning attention is mutually reinforced across vision, sound and touch. Multisensory context enhances a listener's attention to and discernment of sounds (Pastor et al., 2006; Best et al., 2007) including visual modulation of the audio cortex (Kayser et al., 2007). Teng et al. (2016) and Holcombe (2009) that visual cognition seems to have a timescale similar to audio for dynamic event perception. A listener's attention and expectation can influence physiological changes in the brain's plasticity of neuronal dispositions and responsiveness (Hillyard et al., 1973; Woldorff and Hillyard, 1991; Fritz et al., 2005). The resulting modifications in neural signal processing improve temporal performance and acuity of conscious recognition and identification of sounds (Spitzer et al., 1988; Cusack et al., 2004; Alain et al., 2007). Fritz et al. (2005) describes MMN, a supramodal neural mechanism for "oddball sensing" that detects unusual changes in surroundings. MMN indicates a listener's capacity for maintaining an heuristic focus on the transition from pre-attentive to new or recognized sounds.

It is important recall that Sonification does not interpret itself; it requires informed skill and learning how to listen. Heuristic Listening is introduced as an informal model describing multisensory cycles of action and observation that contribute to a listener's attentive process. Heuristic Listening defined here as a *clinically informed skill of multisensory enhanced attention to sounds that may be meaningful in a context of exploration and structured inquiry*. This skilled listening practice is similar to heightened everyday situations where a listener has an expectation a sound will occur but is uncertain of when the sound may occur. Heuristic Listening involves a listener's affective presence in an environment, with context awareness, attention, prediction, possible responses to false cues, and a response performed when an awaited sound occurs.

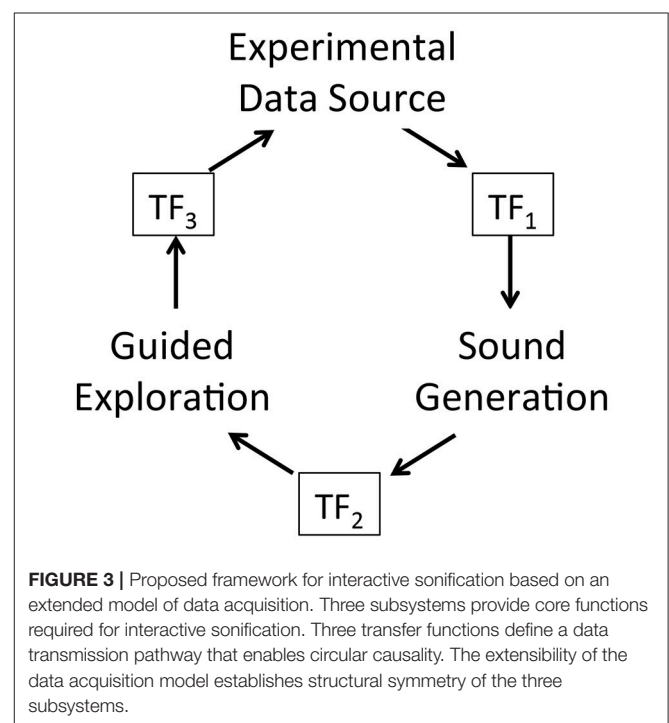


FIGURE 3 | Proposed framework for interactive sonification based on an extended model of data acquisition. Three subsystems provide core functions required for interactive sonification. Three transfer functions define a data transmission pathway that enables circular causality. The extensibility of the data acquisition model establishes structural symmetry of the three subsystems.

A model of Heuristic Listening requires representation of an on-going multi-temporal cognitive cycle, where sound events are disambiguated and articulated by the listener's actions, and where the listener's actions may also set expectations for sound events. Recognition of sounds and events occurs across a multisensory and supramodal cognitive cycle that continuously integrates multiple time layers, where multiple event recognition and observer actions overlap in multiple onsets and terminations. The reference example for this research is

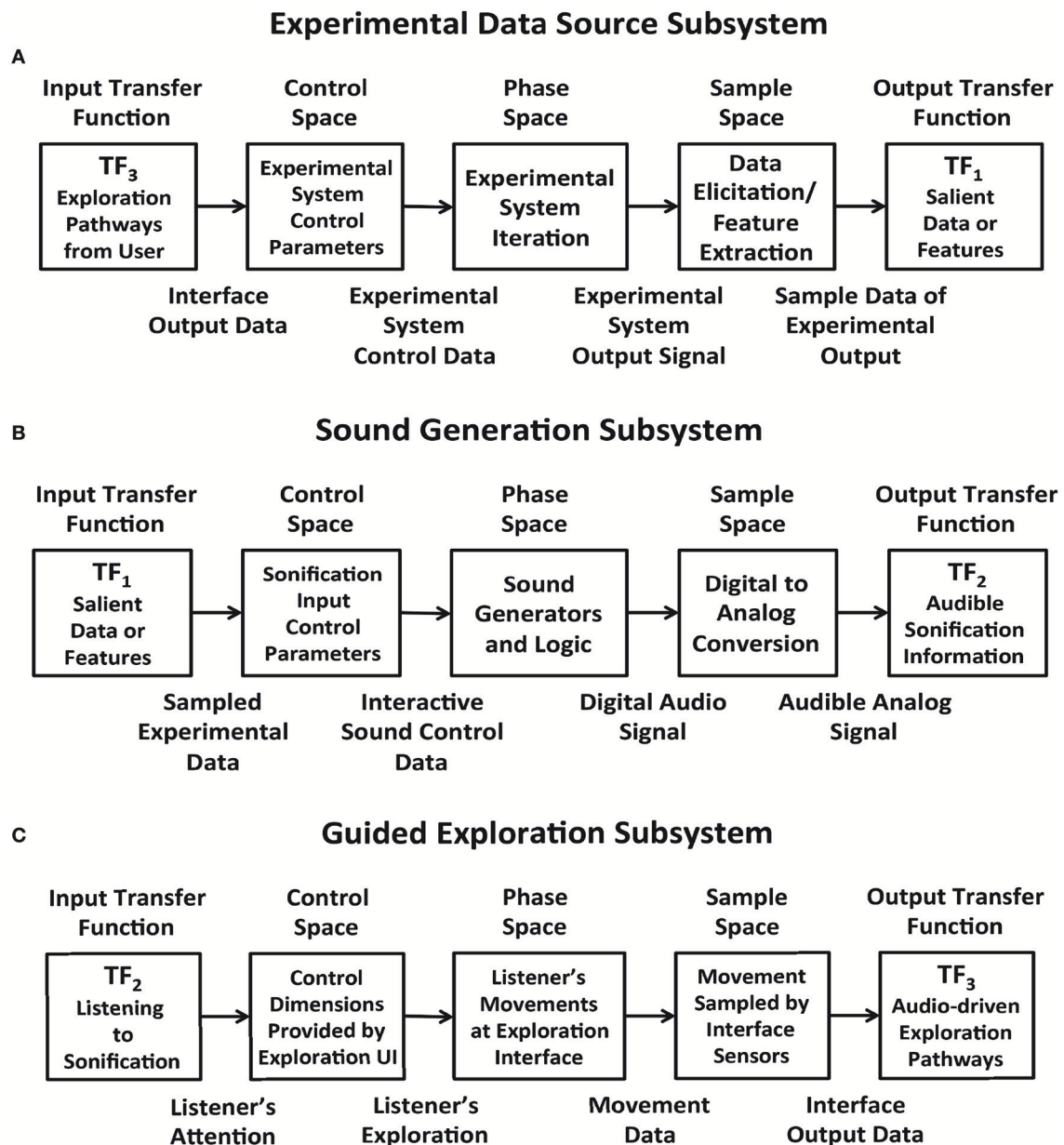


FIGURE 4 | (A) Experimental Data Source subsystem represented using the data elicitation model. Data from a user interface enters through transfer function TF_3 and is applied to the control parameters of a dynamic experimental system, which generates a stream of output data that exhibits emergent behavior. Characteristics of the signal are sampled to acquire relevant features and the sample stream is output to TF_1 . (B) Sound Generation subsystem represented using the data elicitation model. Data entering from TF_1 is applied to the sound control parameters for sound generators. Associated sound design logic encodes procedures that generate sonification. The resulting digital audio signals are sampled to generate audible information output to TF_2 . (C) Guided Exploration subsystem represented using the data elicitation model. A listener positioned at a control interface attends to sonification input at TF_2 . The listener performs movements in response to the sonification, by manipulating the interface. In sample space the interface states from listener's actions are digitally acquired. At TF_3 the interface samples are output to the control space of the experimental data source (see **panel A**). This enables the listener to explore the relationship between actions and audible changes that may be generated.

the LIDA model (Madl et al., 2011) describing a cognitive cycle comprised of Perception, Understanding, and Action Selection. LIDA identifies a 260–390 ms cognitive cycle of 200–280 ms unconscious processing, subdivided into Perception and Understanding, followed by 60–110 ms conscious Action Selection. These phases characterize the perception of audible

attributes that occur within corresponding time windows, such as pitch, loudness, duration, and pulse.

Listening With Multisensory Interaction

Heuristic Listening describes enhanced time-sensitive expectation as a context for developing interactive sonification.

Previously this author has studied listening as a function of physical interactions with emergent systems to generate sound, and introduced a kinaesthetic framework based upon multi-temporal cognitive cycles of multisensory attention and action (Choi, 2017b). Interactive sonification is designed to engage this dynamic temporal acuity as sound events are generated by a listener's actions. An action creates a time focus of attention that may elevate or suppress neuronal responses depending on whether the sound is highly relevant or irrelevant to the conscious listening task (Martikainen et al., 2005). (Lange and Roder, 2006) reports that listeners who receive cues to aid prediction of audible event timing will experience temporally heightened neuronal attention. These findings suggest that a listener can elevate her level of attention to sonification by performing an active inquiry and having interaction with the experimental system being observed. Further, timing and intensity of a listener's exploratory actions will elevate expectations for corresponding changes in sounds. A listener's performance in terms of recognizing sounds may improve if the system provides multisensory attentional engagements. The present research provides three types of attentional engagements: visual cues from dynamic visualization of the experimental data, somatosensory cues from spatial orientation of physical movements within an interface (Hotting et al., 2003), and semantic cues representing users' actions in graphical user interfaces.

The neurophysiological basis of heuristic listening contextualizes a listener's experience of transition from expectation to recognition, reflecting the temporal dynamics of pre-attentive to attentive auditory cognition. A listener may be thought of as having a pre-attentive streaming segregation "buffer." Incoming auditory signals accumulate in that buffer for durations that may be as much as several seconds. During the buffer period a listener's expectation can impact the rate of transition from pre-attentive to attentive state (Bregman, 1978; Molholm et al., 2005). A transition from pre-attention to attention is a moment of critical phase transition; it indicates the listener either recognizing a familiar sound pattern or learning an unexpected sound pattern. This phase transition is the attentive focus of heuristic listening; sounds may be unfamiliar and still be recognized to represent states of an underlying order. Identifying emergent properties in sounds engages heuristic listening in exploring regions of an experimental system by recognizing combinations of familiar and unfamiliar states in sounds. Unfamiliar sounds may be unstable or may exhibit unexpected transitions to new stable patterns. Listening memory plays a temporal role in uncertainty and recognition, exhibiting attributes of *ensemble coding* (Albrecht and Scholl, 2010), a mechanism of temporal statistical summary of information in a perceived scene. Ensemble coding is well-documented in visual summary of complex scenic features, and has been experimentally demonstrated in audible tone patterns (Piazza et al., 2013). The model of ensemble coding provides a foundation to account for a listener experiencing multi-temporal dynamic layers of sounds unfolding in time. In this example a listener simultaneously reflects on sounds previously heard, acquires sounds newly heard, and anticipates sounds yet to

be heard (Ulanovsky et al., 2004). A moment of heuristic listening collocates the anticipatory, immediate, and predictive neurological processes of listening. Finally, heuristic listening implies a skill requirement supported by everyday listening experience, and a listener's capacity to become more acute by training and performing multisensory observation. These models are considered in design of the research methods applied in this work. Appendix 1 in Supplementary Material presents clinical examples of heuristic listening.

RESEARCH METHODS APPLIED TO INTERACTIVE SONIFICATION OF EMERGENT BEHAVIOR

The goal of this research is to assess extensibility of a model as an interactive sonification framework, and to demonstrate its application for emergent behaviors. To perform assessment three methods are combined: *listening scenario classification*, *attunement*, and *control classification using sonification design patterns*. The combined methods are applied to sonify the Chua's circuit and Swarm Chemistry, and the research compares each application to show how these methods work together. The study aims to demonstrate the extensibility of a framework for interactive sonification by comparing variation and consistency in each method across multiple applications.

Listening Scenario Classification

A sonification application includes a context whereby a listener acquires sounds in relation to other modalities of observation. This research introduces a sonification listening scenario as *a system design that prepares and enables listeners expectations of how sound generation is coupled to an experimental system*. Coupling requires that transformations of sounds correspond to salient emergent properties or state changes of the experimental system. A properly engineered sonification listening scenario yields an interactive learning pathway for listening skill acquisition, supported by a listener's understanding of the experimental apparatus.

A listening scenario is a construct designed with multimodal attributes that become part of a local listening environment; the listening experience is not determined solely by the audible sonification output. Sounds are perceived in a highly subjective environment often fused in multisensory percepts. (Bregman, 1990) describes an *auditory scene* as a temporal superposition of "component" sounds comprised of multiple sources, some of which are not controllable, even in an isolated listening environment such as headphones or an anechoic chamber. Environmental conditions generate component sounds that are attended at different levels of awareness. Neurophysiological pre-attentive mechanisms for audio stream segregation play an important role for differentiating and keeping track of multiple sounds from different sources in a complex auditory scene. Some but not all audible sounds rise to conscious awareness, a subset of audible sounds is noted as distinct events, a subset of these may draw a listener's attention. According to directed attention hypothesis (Welch and Warren, 1980; Andersen et al.,

TABLE 1 | Classification of sonification listening scenarios organized by type of affordance.

Affordance and means	Indicative classes of user experience
1. Affordance: Sounds for a domain investigation Means: Sound generation and sound design	<ul style="list-style-type: none"> • Identify sound qualities - what is heard • Range of change exhibited by a sound • The meaning of a sound in relation to experimental data; how a sound can be described with respect to the state of the simulation
2. Affordance: Interfaces for a domain investigation Means: Control paradigm	<ul style="list-style-type: none"> • Sensory properties of a user interface including appearance, function, orientation, and ease of operation; • Degrees of freedom provided in an interface • Multidimensional control of an experimental system; the number of system control parameters accessed
3. Affordance: Perceptually compatible temporality of events Means: Scheduling of data	<ul style="list-style-type: none"> • Synchronization or latency between the introduction of changes in control space, and the experimental system's response in phase space • Synchronization or latency between phase space output and corresponding transformations of sounds • Temporal resolution in a signal output from an experimental system and how this is reflected in audible temporal resolution
4. Affordance: Multisensory information Means: Engaging two or more modalities of an observer	<ul style="list-style-type: none"> • Visual cues provided by an experimental control interface while sound is generated • Visualization of data alongside sonification • Enactive control of the domain investigation while listening; physical responses of interface
5. Affordance: Complexification Means: Generating and controlling levels of detail	<ul style="list-style-type: none"> • Varying the ratio of experimental data to sound control data, with the purpose of modifying the level of detail of data mapped into sound • Reduction of experimental data: Filtering the level of detail in data output from the domain investigation
6. Affordance: Learning Provision Means: Rehearsal for developing competence	<ul style="list-style-type: none"> • Access to example scenarios of experimental states and associated sonification outputs • Access to practice scenario interaction with rehearsal time to become acquainted with the domain investigation and sound generation

2004), multisensory percept plays a role in determining what sounds in the auditory scene are identified or disregarded based on what modality is dominant at any given moment. Both multisensory fusion and modality dominance can contribute to highly subjective listening. Supramodal auditory attention hypothesis states that stimulus driven shifts of auditory attention are controlled by a supramodal mechanism (Ward, 1994). A sonification listening scenario enabled with a multimodal interface engages multiple senses to inform the interpretation of sounds.

The auscultation training examples surveyed in Appendix 1 in Supplementary Material demonstrate how sounds representing known experimental states may be learned through observation of established cases. Classification of a sonification listening scenario can be formalized as a guiding template with examples of multimodal system norms and corresponding sound qualities. Sounds that are talismans of unfamiliar states or properties may thereafter be established through empirical observation. Deviance detection is highly sensitive in auditory perception (Fritz et al., 2007) and is coupled to neural responses in other sensory regions (Downar et al., 2000; Huang et al., 2005). Once a listener learns to recognize sounds that are a norm, exceptional sound events can be recognized.

In this research the classification of listening scenario is organized by classifying the affordances of the sonification system design: (1) type of affordance, (2) means to realize the affordance,

and (3) indicative listener experiences related to the affordance. **Table 1** presents a classification using on six types of affordance:

- audible quality,
- user interface,
- temporality,
- multisensory qualities,
- complexification, and
- learning provision.

These attributes are device-agnostic and data source-agnostic. Examples of indicative user experiences provided in **Table 1** represent potential criteria for qualitative and quantitative measurement. Appendix 2 in Supplementary Material presents a test case of the listening scenario classification method applied to independent published research involving personalized sonification of EEG feedback.

Attunement of Explorable Space

Attunement is an *a priori* process for conditioning a playable space for auditory display (Choi, 2014a). The modeling of playable space is introduced from other applications of model-based interaction (Choi and Bargar, 2011; Choi, 2014b). Playable space is not a user interface; it is an enabling design for the development of interfaces. A model of a playable space may be formalized as a set of canonical relationships that enable the development of auditory interfaces for observing

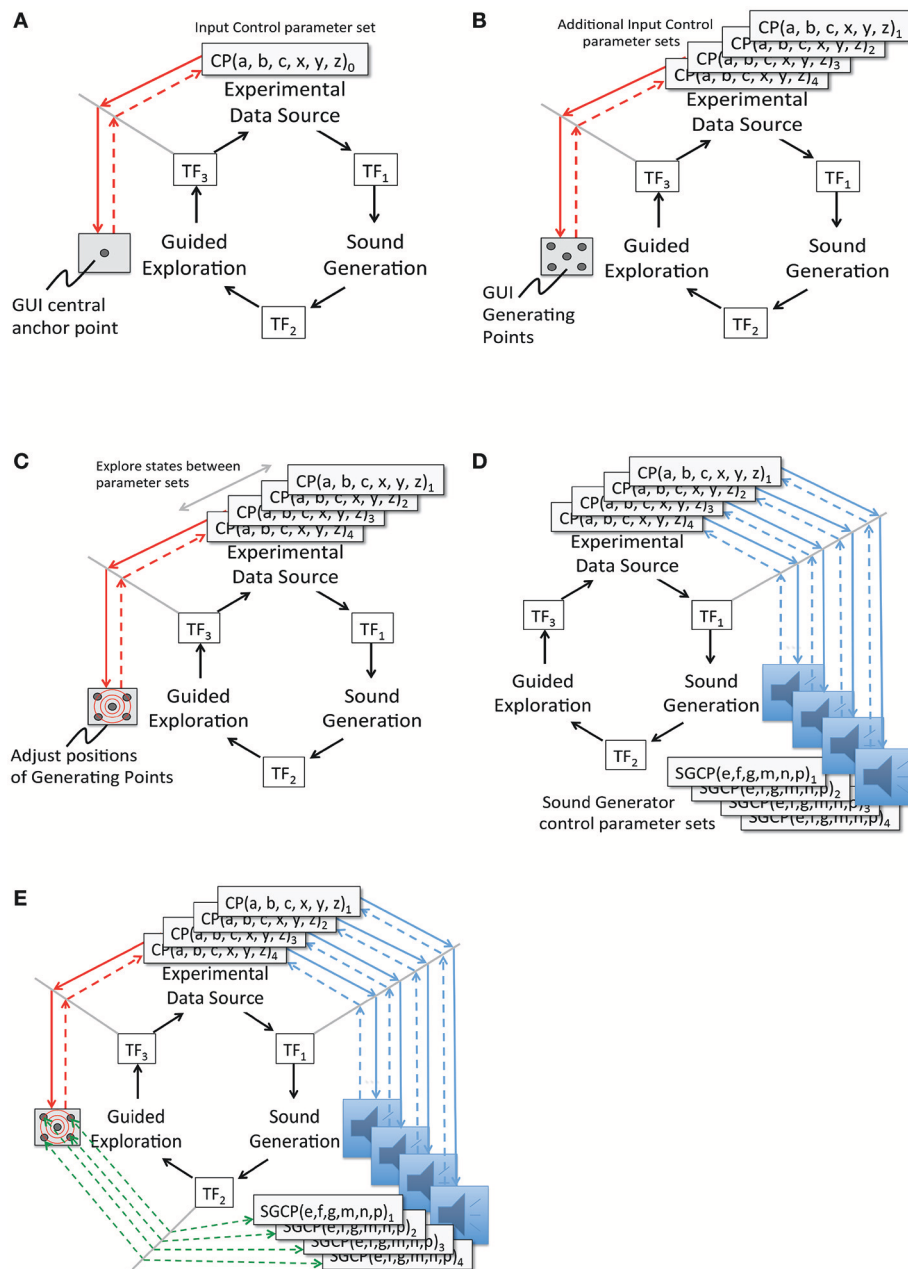


FIGURE 5 | (A) Attunement Step 1: Between two subsystems, Guided Exploration and Experimental Data Source, define an anchor point in the interface to control the simulation. Process: (a) For Experimental Data Source: Initialize control parameter values of a known stable state. (b) For Guided Exploration: Select a central position in the interface and assign the control parameter values from Step A. (c) For TF_3 : Encode the position assigned in Step B as the initial anchor point for attunement. (This is a multidimensional mapping from the set of control values A to the interface state B). Requirement: The initial mapping and interface anchor position are selected by human inference based on prior experience with the subsystems. **(B)** Attunement Step 2: Define *generating points* connecting the Experimental Simulation and the Guided Exploration interface. Process: (a) For Experimental Data Source: Identify control parameter sets of further stable simulation states. (b) For Guided Exploration: Select a set of new interface positions to correspond to each set of simulation state control parameter values from Step E. (Each set of control values E_n requires a multidimensional mapping to an interface position F_n). For TF_3 : Encode each interface position assigned in Step B as a generating point, GP_n . Requirement: Interface positions for stable states are readily accessible and evenly positioned in the control interface. **(C)** Attunement Step 3: Establish a Scale of Transformation between generating points. Process: (a) For Guided Exploration: Explore interface positions between the Generating Points. (b) For Experimental Data Source: Observe changes in simulation Phase Space induced by intermediate positions in the exploration interface. (c) For TF_3 : Modify GP interface positions or GP control values to optimize symmetry of the interface, in terms of state transitions actuated when the interface moves between GPs. Requirement: Relative degree of change in the interface should correspond to relative degree of change in the simulation phase space. **(D)** Attunement Step 4: Between two subsystems, Experimental Data Source and Sound Generation. Process: (a) For Experimental Data Source: For each previously selected control state, observe characteristics of

(Continued)

FIGURE 5 | the data output. (b) For Sound Generation: Working with a selection of procedurally generated sounds, identify a set of sound control parameter values corresponding to each Data Source control state. (c) For TF_1 : For each Data Source control state, encode a mapping from the experimental data output to a set of sound control parameters. (d) Requirement: Initial sounds are selected based on knowledge of sound design and previous experience with the dynamic qualities of the experimental data output. **(E)** Attunement Step 5: Across three subsystems, from Guided Exploration to Sound Generation. Process: (a) For Guided Exploration interface: Select each GP then select interface positions between generating points. (b) For Experimental Data Source: At each GP verify the state of the simulation phase space. Between GPs observe phase space transitions. (c) For Sound Generation: At each GP verify audible output. For interface positions between generating points verify audible transformations. (d) For TF_1 : Modify sound control space mappings to optimize for audible transformations that have a range of discernable differences corresponding to the dynamic range of the experimental data output. Requirement: Relative degree of audible change in sonification should correspond to relative degree of change in the interface—implying a relative degree of change in the simulation phase space. Establishing normalized degrees of interface action and audible response, differences in audible transformations will indicate non-linear properties of the Experimental Data Source rather than artifacts of interface or sound control.

dynamical systems (Choi, 2014a). The concept of space as a working metaphor is common in scientific practices especially in applications of simulation and modeling¹. The space metaphor is adopted to identify the formation of explorable regions of system states as having definable structure and function. For sonification applications the term *explorable space* is introduced to describe the collective high dimensional parameter space and circular causality of the extended sonification model illustrated in **Figures 3, 4A–C**. Appendix 3 in Supplementary Material summarizes functional explorable space comprising the extensible sonification model. The system is calibrated to ensure the sonification outputs and listeners' actions are meaningful with respect to the experimental system. Calibration involves adjustment of many control parameters, within each subsystem and between subsystems at each transfer function TF_{1-3} . Attunement is a calibration method that systematically reduces the dimensionality of the parameter space in two stages, first stabilizing the subsystems' internal parameter values then adjusting the transfer functions. TF_{1-3} are referred to as *tuning functions* when the design of the transfer functions formalizes the attunement of the sonification framework.

To perform attunement, the range of parameter values applied at one TF may require adjustments at other TFs. Each TF requires tuning to optimize for isochronous differences between subsystems, adjusting oversampling and undersampling between subsystems to minimize artifacts generated by aliasing.

Figures 5A–E describe the process of implementing attunement. The process begins with the user interface by defining a set of discrete interface states as *generating points* (GPs). In **Figure 5B** the interface GPs are assigned to a selected set of control states. To simplify the process of mapping M-dimensional GPs to N-dimensional data control parameters, a manifold interface technique is introduced (Choi, 2000a). In **Figure 5C** the explorable interface space is calibrated with control regions of the Experimental Data Source. In **Figure 5D** sound control parameter data sets that determine

audible features are aligned with selected states and features in the experimental data. In **Figure 5E** the listener associates audible features with generating points at the user interface, and audible transformations with explorable regions in the user interface. The attunement process is applied in cycles of iterative refinement while auditioning control input and sound output. Regions of interest in the experimental data are brought into correspondence with sound control data and audible transformations. Discovery of regions of interest in experimental data space may require iterative refinement. Adjustments to parameter values are applied at tuning functions TF_1 and TF_3 . Tuning function TF_2 is a representation of a listener's performance of heuristic listening.

Control Classification Using Sonification Design Patterns

A sonification design pattern (SDP) is a control structure for generating a data-driven audio stream. SDPs selectively control audible features to optimize audibility of features exhibited in experimental data. SDP structure is agnostic to audio content in the sense that one structure may control many types of sound generators. SDP are informed by Alexander's concept of design patterns (Alexander et al., 1977) developed for architecture, and SDP may be considered members of the superset of design patterns used in software engineering (Gamma et al., 1994), as SDP define procedural audio using instruction sets for sound generators. This author introduced the SDP method for interactive sound generation using non-linear simulation data (Choi and Bargar, 2014), and has applied SDP in multimodal performance with evolutionary systems (Choi, 2017a). SDP facilitate attunement by classifying control data features to align with audible features. In the sonification framework SDP are located in the Sound Generation subsystem and receive data from the output of TF_1 . SDP control parameters define the control space of the Sound Generation subsystem (**Figure 4B**), where SDP functions are modulated by the experimental data from TF_1 . SDP may be designed to control many different sound palettes. In the Swarm Chemistry case studies two SDP examples are applied.

Duration is the first organizing attribute for all SDP. Control parameters are defined for SDP depending upon the duration range and the type of sound generator to be controlled, presented in **Table 2**. Duration range of an SDP refers to signal processing time required to generate an audible signal combined with perceptual time required for a listener to register the audible attribute or pattern. We identify five sets of SDP attributes with characteristic duration ranges: SDP 1 controls pitch change

¹The concept of space is implicit in musical instruments; for example musicians' finger positions varying the lengths of strings and tubes are spatial manifestation of control structure for pitches. While varying in shapes and sizes, musical instruments commonly offer contact points and constrained space for performance navigation. The lesson to learn from this tradition is the optimization of instruments have been engineered with spatial abstraction, the distances and proportional relations for positioning physical elements; the optimization is highly targeted to achieve not only aesthetically sounding tones also physiologically coherent structure for the human body. System coherence is reflected in human capacity to explore the instrument and its audible space with facility.

TABLE 2 | Sonification design pattern classification by duration.

SDP timescale, micro to macro	Audible attributes that generate features in SDP	Duration range of audible SDP attributes
SDP1—micro	Pitch change, loudness change	50–200 ms
SDP2—micro	Timbre, Resonance, Filtering	200–500 ms
SDP3—meso	Sound Source Location and Distance cues; Spatial cues	450 ms–2.0 s
SDP4—meso	Sound Event (from onset to evolution to termination)	500 ms up to seconds
SDP5—macro	Patterns of Rhythm, Tempo, Spoken Words, Melody	1 s up to seconds

Multiple SDPs are combined to develop an audible palette. Audible features of SDP structure are dependent on the sound design and attributes of the sound generator. Features in SDP classes can be controlled independently at a sound generator while the audio output may elicit auditory co-dependencies due to psychoacoustics of perception.

and loudness change; SDP 2 controls timbre, resonance and filtering; SDP 3 controls sound source location and spatial cues; SDP 4 controls distinct sound events; and SDP 5 controls patterns made up of multiple sound events. A sound event must have sufficient duration to be audibly reproducible—meaning a pattern is recognizable because of its organization of elements (see Wiener in section Introduction: What Do We Listen to When We Listen to Data?). Up to a limit, a pattern is identifiable independent of tempo (rate of change), frequency range, and other attributes. **Table 2** provides duration ranges of five SDP classes. Appendix 4 in Supplementary Material reviews auditory perception at multiple time scales, relevant to SDP duration classification and the audibility of data.

RESULTS: APPLYING THE PROPOSED SONIFICATION FRAMEWORK TO MULTIPLE EXPERIMENTAL SYSTEMS

An extensible model of interactive sonification is introduced in section Data Acquisition Model for Sonification Signal Processing as a candidate sonification framework. The extensibility of the framework is compared across three sonification case studies using two experimental simulations that generate emergent behaviors. One case is implemented for Chua's Circuit and two for Swarm Chemistry. The three cases are each presented in terms of attunement method at TF_1 , listening scenario design, and observed outcomes with interpretation. Each implementation adopts a different level of abstraction between the experimental data source and the sound generation. The collective outcomes compare the framework performance in representing the experimental implementations. Performance measures include:

- Accuracy and relevance of the framework representation of an experimental sonification implementation
 - Accuracy of framework representation of an experimental system in terms of functional components, their sequence and their relationships
 - Whether the framework lacks components required in the experimental system, or the framework includes components extraneous to the experimental system
- Using the framework to compare multiple experimental sonification implementation

- Identify model components that have common roles in two+ implementations
- Identify model components that have different roles in two+ implementations
- Identify experimental components that have inconsistent relationships with model components in two+ implementations
- Using the framework to compare solutions for emergent behavior adopted in multiple sonification implementations
 - Similarity and difference of challenges presented by different types of emergent behavior
 - Accuracy of the solution space representation in the framework, compared to the application of the solution in the experimental system.

Sonification Framework Applied to Chua's Circuit Simulation

Chua's circuit is an experimental paradigm that satisfies the working definition of a chaotic system. It is well-suited to observe and analyze emerging behavior in physical systems, being comprised of the minimum number of elements required for a circuit to demonstrate chaotic behavior (Kennedy, 1993), illustrated in **Figure 6**. Chua's circuit has been implemented as a physical device and as a digital simulation using numerical models of the circuit elements². This author has applied interactive sonification to both physical and digital implementations (Choi, 1994). Many classes of biosignals exhibit oscillations with functional chaotic properties. Simulations are used to study and classify these behaviors, and Chua's circuit exhibits a diverse repertoire of intermittency, quasi-periodic patterns, and chaotic signals. (Kozma and Freeman, 2017) describe intermittent series of synchronized metastable brain states as essential to neocortex processing, with intermittency enabling serial phase transitions that advance cognition from one metastable pattern to the next, a model known as the cinematic

²To demonstrate the fidelity of the simulation to the experimental system, control voltages of the physical circuit are measured and applied as control parameter values in the simulation (Zhong et al., 1994). The digital signal output of the simulation at a 44.1 kHz sample rate produces an audible signal virtually identical to that of the physical experimental circuit tuned to an audible frequency range. Extended sequences of voltage-control changes applied in parallel to the physical and simulated circuits maintain the two systems in common regions of phase space and produce audibly identical signals.

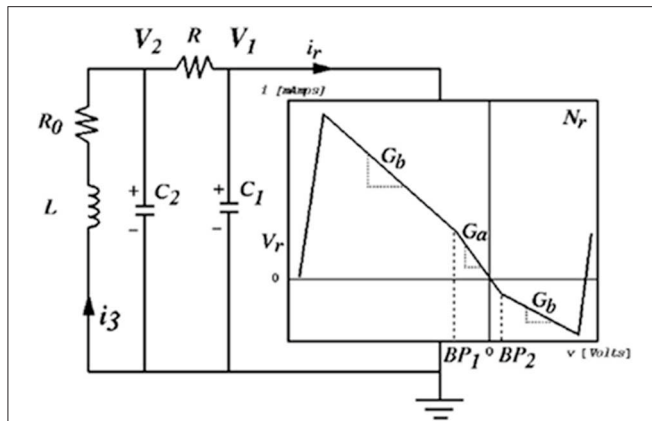


FIGURE 6 | Chua's circuit is comprised of seven elements: one locally active resistor where energy is introduced, three energy storage elements—two capacitors and a resistor, and one resistor programmed with a piecewise non-linear transfer function that induces chaotic behavior.

theory of cognition. Haken (1983) associates intermittent synchronization with information transfer between levels of neurons. (Tsuda et al., 2016) identifies biological dynamics under constraints such as embryonic development and differentiation of cortical functions, as having dependencies on properties of non-equilibrium systems such as bifurcation and attractor formations such as those well-observed in Chua's circuit, and outlines a critical connection from chaotic dynamics to the capacity for macroscopic self-organization in biological systems, providing mathematical models. Although the Chua's circuit is a deterministic system, emergent oscillatory features cannot be predicted by the states of the individual Chua oscillator elements or from preceding states of the output signal. **Figure 7** presents examples of emergent behaviors of the Chua's circuit oscillation. In addition Chua's circuit exhibits *hysteresis* (Zhong and Ayrom, 1985) where a given control state may generate more than one oscillatory state, resulting in more than one set of audible features for a set of control values. The sequence and duration of control state changes and the corresponding sequence of oscillation states influence hysteresis. This emergent behavior presents significant challenges to generating reliable sonification. Appendix 5 in Supplementary Material presents a technique adapting attunement for sonification to compensate for hysteresis.

Case Study 1: TF₁ Attunement for Chua's Circuit as a Sound Generator

The signal output by Chua's circuit oscillation exhibits properties that may be human audible. The attunement is illustrated in **Figure 8** and **Table 4** details functional components. The circuit oscillation is located in the phase space of the Experimental Data Source, with a seven-dimensional control space. Sonification is implemented by enabling the listener to access an audible signal from a direct data stream of the circuit's state.

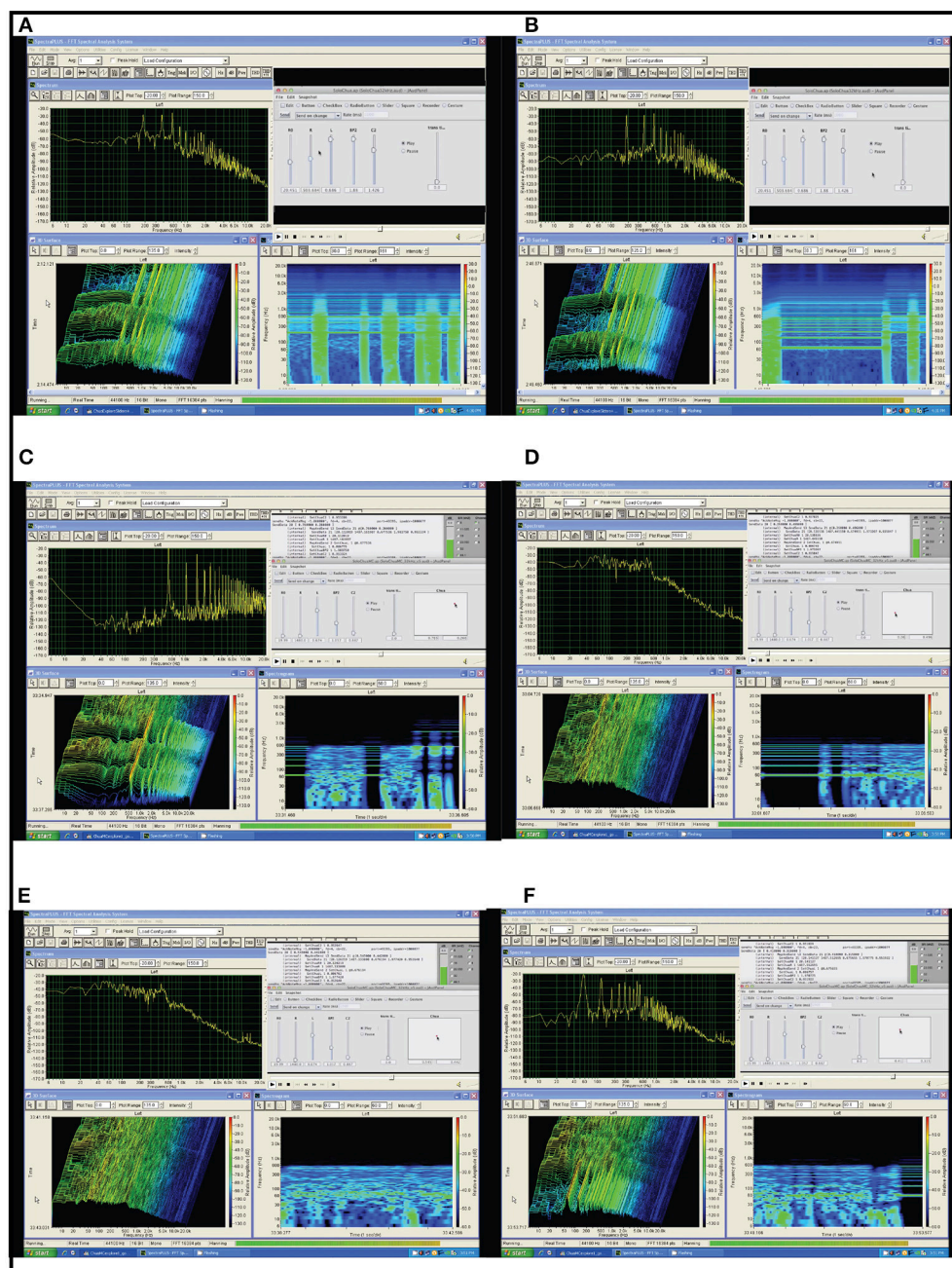
Attunement at TF₁: The tuning function TF₁ applies scaling to the control voltages of circuit elements to establish oscillation in a human-audible frequency range. TF₁ also identifies a single

circuit element, capacitor C2 (**Figure 6**), where the oscillation signal is extracted and routed to digital-to-analog conversion to generate an audible signal.

Listening Scenario: Listening directly to the experimental data in real time creates an affordance for highly interactive control of Chua's circuit, requiring covariance of seven circuit elements for agile navigation of non-linear phase space. A manifold interface technique was developed to facilitate interactive covariance for n -dimensional control parameters (Choi, 2000a). To regulate sound generation from emergent behaviors, *fiducial points* are a type of generating point used for attunement (**Figure 12**). Appendix 5 in Supplementary Material introduces the manifold interface technique for structured high dimensional control and discusses the use of fiducial points to empirically tune the exploration interface.

Observed Performance: The manifold interface tuned with fiducial points enables a listener to control the Chua's circuit in real time. Using the interface to covary seven control parameters, a listener can develop agility to manually guide the circuit states through regions of interest. Guidance from audible features improves the precision of navigating control regions that generate emergent behavior. Emergent behaviors are often exhibited in unstable regions that border unresponsive regions such as fixed points—where oscillation ceases, and limit cycles—where oscillation is fixed in a sine wave. Attunement aids exploration by identifying fiducial points with audible characteristics that indicate when the circuit is approaching undesirable regions. By attentive listening and applying micro-modifications to control parameters the listener can navigate circuit states away from a transition region or nudge the circuit state from one attractor region to another. This approach makes use of the sensitivity of unstable regions, and very small changes introduced at the manifold interface will influence the signal trajectory at boundaries of phase transitions.

Interpretation: Fiducial points can be used to mark behavioral trends in control space regions, and these points used to guide listeners through unstable regions in phase space. An interface providing stable control points bordering unstable regions supports the practice of Heuristic Listening by enhancing the reliability of free navigation in complex control space. These methods are compatible with the use of signal analysis and modeling to identify points in phase space that represent, predict and influence behavioral trends in chaotic systems. (Schiff et al., 1994) report a chaos control technique that applies signal analysis to identify unstable fixed points for learning the directions of signal approach and divergence. A small perturbation in the signal is introduced in these regions to prompt the signal to adopt preferred pathways. Faure and Korn (2001) report the use of recurrences plots in regions of phase space to characterize signal tendencies in the region and to predict the evolution of dynamics in the region, to a proximate future. These regions are sites for influencing phase transitions and signal behavior by applying weak perturbations to the control space. Attentive listeners using a manifold interface can explore highly unstable boundary conditions and influence the circuit to maintain quasi-stable states.



(Continued)

FIGURE 7 | time series a change of control parameter shifts the oscillation to an upper limit cycle attractor, which is nearly periodic at 500 Hz producing a fundamental harmonic tone and an ascending series of tones at integer multiples of 500 Hz. Note in the lower two windows during the bursts of chaos the amplitude heat maps show the frequency spectrum energy remains concentrated around the periods of the neighboring attractor. Voltage control change is applied in parallel to five circuit components using the cursor in the plane on the right side of the GUI. This is the 2D control surface for the manifold interface, mapping each 2D position to a 5D control signal. **(D)** Onset of Chaos. At the start of the time series a stable periodic attractor produces a harmonic tone with fundamental frequency at 60 Hz with additional periods producing harmonic tones at integer multiples above the fundamental. By introducing voltage control changes the system moves gradually out of the basin of attraction and falls into a chaotic region. Note in the lower two windows during the onset of chaos the amplitude heat maps show the frequency spectrum energy remains concentrated around the periods of the nearby attractor. Voltage control change is applied in parallel to five circuit components using the cursor in the plane on the right side of the GUI. This is the 2D control surface for the manifold interface, mapping each 2D position to a 5D control signal. **(E)** Chaotic Oscillation. A chaotic attractor produces oscillations constituting many rapid transitions between multiple periodic regions. The result is a noise-like signal with a faintly audible tone center shifting across the frequency range between 50 and 100 Hz. This persistent weak attractor region is visible in the spectral envelope of the upper left image and in the amplitude heat maps of the two lower images. Voltage control change is applied in parallel to five circuit components using the cursor in the plane on the right side of the GUI. This is the 2D control surface for the manifold interface, mapping each 2D position to a 5D control signal. **(F)** From Chaotic to Periodic Oscillation. At the start of the time series a chaotic attractor produces oscillations constituting many rapid transitions between multiple periodic regions. The result is a noise-like signal with a faintly audible tone center shifting across the frequency range between 50 and 100 Hz. This persistent weak attractor region is visible in the amplitude heat maps of the two lower images. During the time series example control voltage change is applied steadily and the oscillation exhibits intermittent stability then becomes stable on a periodic attractor at 50 Hz, with higher periods at integer ratios of the fundamental. Voltage control change is applied in parallel to five circuit components using the cursor in the plane on the right side of the GUI. This is the 2D control surface for the manifold interface, mapping each 2D position to a 5D control signal.

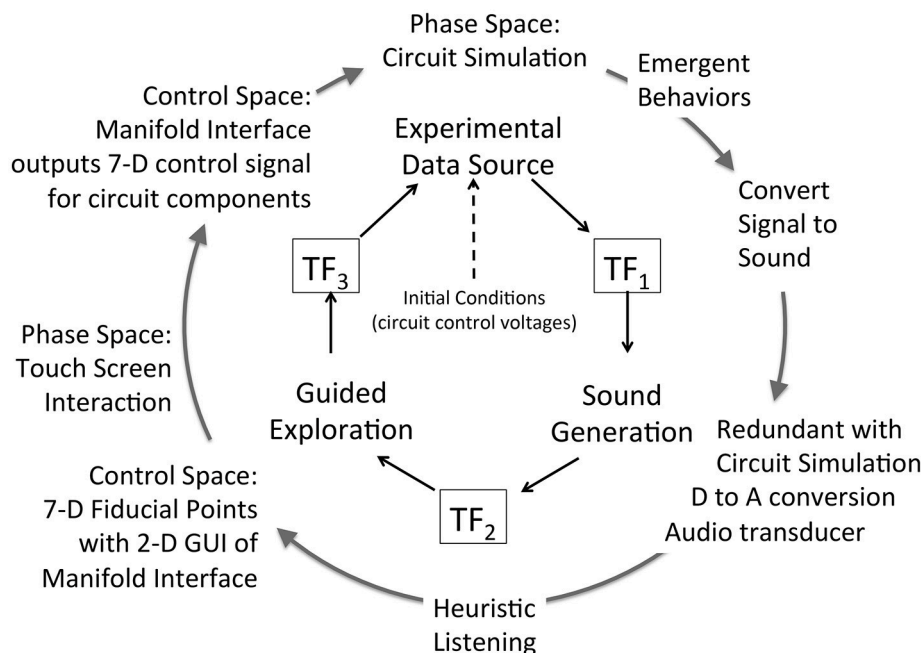


FIGURE 8 | Attunement applied to Chua's circuit using the sonification framework. The Experimental Data Source is Chua's circuit controlled by voltages of seven electronic components. The circuit oscillation range is tuned to a human-audible frequency range. The Sound Generation system is redundant with the circuit simulation; the circuit oscillation is sampled at a capacitor and converted into an audible signal. The listener acquires the audio signal while engaging a 2D interface, which visualizes fiducial points of stable control regions and enables exploration of other control regions. The Manifold Interface enables the 2D GUI to represent a 7D control space, continuously and differentially in a manifold subset of control space. Listener actions at the interface are converted to a 7D control signal and applied to the control voltages of circuit components. This enables heuristic listening in response to changes induced in the phase space of the circuit oscillation. **Table 4** details the functional components in this figure.

Sonification Framework Applied to Swarm Chemistry Simulation

Swarm Chemistry (Sayama, 2012) is based on Reynold's "boids" system (Reynolds, 1987). Sayama implements heterogeneous agents, each agent having autonomous social tendencies expressed as movement, and awareness of other agents sharing the movement space with social tendencies. The simulation specifies 100 to 300 agents, each agent initialized with movement

tendencies, perceptual radius, social responsiveness and a random initial position and velocity. Each agent is visualized as a low-polygon 2D graphical object animated in a bounded plane. An agent moves autonomously with a defined probability of random velocity until another agent enters its perceptual radius. At each time step in the simulation every agent responds to all other agents that are within its perceptual field. Agents' movement responses express these parameterized

TABLE 3 | Swarm Chemistry behavior control parameters and social conditions.

Agent control parameter	Agent's behavioral tendency	Social conditions	
		Alone	With others
Straying	Move randomly if no other agents are in perceptual range	Required	
Pace Keeping	Approximate current speed to normal speed	Optional	Optional
Cohesion	Move toward the average position of other agents in perceptual range		Required
Alignment	Accelerate toward the average velocity of other agents in perceptual range		Required
Separation	Avoid collisions with other agents		Required
Whim	Move randomly with a given probability		Required

TABLE 4 | Sonification Framework implementation for Chua's circuit.

Function		Component
Experimental simulation	Control space	7-dimensions: Control voltages for seven circuit components
	Phase space	Chua's circuit oscillation across seven circuit components
	Sample space	Oscillation signal at capacitor C2—a single channel of continuous frequency-amplitude oscillation
TF ₁	<ul style="list-style-type: none"> Chua's circuit fundamental frequency set at 50 Hz Signal extracted from component capacitor C2 routed to DAC 	
Sound generator (redundant with experimental simulation subsystem)	Control space	Co-variance of control voltages of seven circuit parameters in select control regions defined in TF ₃ and inherited from the simulation control space
	Phase space	DAC of circuit signal at 44.1 kHz to an audible signal
	Sample space	Continuous audio signal of Chua's circuit oscillation is routed to loudspeaker
TF ₂	Listener acquires sound output and engages Manifold interface to explore system properties in control regions, guided by audible features	
Guided exploration	Control space	Fiducial Points for Manifold Interface based upon stable audible properties related to control space of simulation
	Phase space	Listener navigating the continuous control plane of Manifold Interface
	Sample space	Listener's cursor path in Manifold interface is output as a 2D time series data stream of (x,y) positions
TF ₃	Bijective Mapping: continuous, differentiable and bi-directional mapping transforms 2D cursor positions into multidimensional time series data of control values for seven Chua circuit components	

Manifold Interface indicated in Guided Exploration is discussed in section Case Study 1: TF₁ Attunement for Chua's Circuit as a Sound Generator, and in Appendix 5 in Supplementary Material.

tendencies: straying, cohesion, alignment, separation, whim, and pace keeping, presented in **Table 3**. An agent's tendencies are quantified as strength of attraction to the average position and average velocity of all perceived neighbors, an imperative to avoid collision, a probability to move randomly, and strength of tendency to approximate its own average normal speed. Parameters for these attributes for each agent define the control space of the simulation. The state of an agent includes its attribute values, its current position in the movement plane, and its movement history required to determine acceleration and probability. At each simulation time step the movement response of each agent is solved for its attributes with respect to all other perceived agents, and the resulting positions of all agents are collectively updated. Other than state required for these calculations an agent has no memory of prior movement or location, and agents have no top-down spatial view of the

movement area or of other agents' formations or positions. A system was implemented for human socialization with agents enabling touch screen interventions in the simulation (Choi and Bargar, 2013). With or without external intervention collective patterns emerge across groups of agents. **Figure 9** illustrates examples of swarm agents' emergent behavior, induced by an observer using an interface to interact with agents' social tendencies.

Procedural Sound Generation for Swarm Chemistry Sonification

Swarm Chemistry simulation data cannot be converted directly into an audible signal. The Sound Generation subsystem for Swarm Chemistry adopts procedural sound synthesis, enabling a broad range of auditory representations including sonification design patterns (section Control Classification using Sonification

Design Patterns). For two cases studies, two methods of sonification are implemented as alternative attunements of transfer function TF_1 in the sonification framework, illustrated in **Figures 10, 11**. Case Study 2 (section Case 2: TF_1 Attunement for Parallel Data Streams of Many Agents) applies individual agents' data to directly control an equivalent number of individual sound sources. This method relies upon multiple sound source aggregate interaction to generate emergent features analogous to visual pattern emergence. Case Study 3 (section Case 3: TF_1 Attunement Using Feature Recognition Data) applies statistical measures to detect the emergence of salient features. This method applies swarm data indirectly via pattern recognition data to control sound synthesis. To maintain control for comparative assessment, Case studies 2 and 3 adopt a common attunement of transfer functions TF_2 and TF_3 , including a common Exploration Interface for the listener to interact with the Swarm Chemistry simulation. Common attunement of TF_2 and TF_3 enables the comparison of two sonification methods at TF_1 applied to a common source of emergent behaviors. At TF_2 the Exploration Interface is implemented using a multi-touch surface to display a graphic visualization of swarm agents. The visualization becomes an interface by enabling the user to interact directly with swarm agents by touching the screen. Appendix 6 in Supplementary Material introduces the *super agent* mechanism that enables listeners' social interaction with agents in simulation phase space. Adopting swarm data visualization as an interactive interface requires sounds generated in real-time presented synchronously with the visualization. As an audible channel parallel to visual patterns, sonification may provide either redundant or complementary information. Dynamic patterns emerging from agents in aggregate pose a challenge for fidelity of data representation in sound. Emergent patterns are transient features and introduce uncertainty in rendering these features in sound. Patterns that emerge visually in the swarm data may not emerge in sound, depending upon the relationship between the swarm data and the sound generator. This difference is demonstrated in Case Studies 2 and 3.

Case 2: TF_1 Attunement for Parallel Data Streams of Many Agents

To sonify the data of individual agents the movement of each agent is measured and transmitted as control data for a corresponding sound source. Assigning a unique sound source for each agent may be referred to as a "literal" method: n agents generate n concurrent data streams applied to control n individual sound sources. For swarms where $n = 300$ to 500 this data is high density and the sound computation is intensive for real-time interaction with the swarms. Auditory perception tends to limit recognition of concurrent sound sources; the attunement of many parallel sound streams anticipates their collective mixture. To generate sound for each agent a harmonic tone with uniform frequency spectrum, frequency range, and duration were applied to each agent's data to normalize the audibility of all agents. A high-density sound mix was anticipated to mask sounds of individual agents so that collective attributes will emerge. This attunement is illustrated in **Figure 10** and **Table 5** details functional components.

Attunement at TF_1 : To preserve a literal association to the visualization of agents, each (x,y) position was mapped to an auditory range that varied over fundamental frequency (pitch) and stereo position. To be highly literal in correspondence to the visualization, each agent's y-axis position was mapped a pitch (low to high with the position of the agent) and each x-axis position was mapped to stereo position (left to right with the position of the agent). These associations were selected as the most elementary with respect to simultaneous visual display of agent positions. Phase space dimensions are scaled to 1024×768 pixels; agent size is 4×4 pixels and agents move stepwise by units of 1 pixel.

Listening Scenario: The sonification design is hypothesized to render only general correspondences perceivable. For example, using a stereo field to represent lateral position of sound sources provides a range of perhaps a dozen virtual source positions that can be distinguished (Begault, 1994; Pedersen and Jorgensen, 2005), and then only when agents are in a relatively tight cluster. For 300 agents the tuning anticipated that when at least 70% (210 agents) are in a single cluster occupying no more than 15% of the x-axis range (153 pixels), a dominant stereo position will be audible. For pitch perception the frequency range of the y-axis was scaled between 800 and 1,200 Hz so that agents in close proximity will generate a focused pitch center (the harmonic ratio of 2:3 is a Perfect Fifth, a little more than half the perceptual difference of an octave). This tuning anticipates, when at least 60% (180 agents) are in a single cluster occupying no more than 15% of the y-axis range (114 pixels), that on the y-axis a dominant pitch center will be audible. This sonification design anticipates widely dispersed agents to render a broad tone cluster with no directional imaging of a sound source location. Evaluation aimed to test whether a separation of agents into two discrete clusters would be audible as two sound clusters separated by relative differences in pitch center and perceived source position.

Observed Performance: The literal attunement did not convey audible patterns that were easily recognized, compared to the clarity of visual patterns from the same data. Audible features were much less distinct than visible features of agent distributions and clusters. Source position and pitch center overall were weak. The dominant sound was a quiet broad pitch spectrum across the frequency range and stereo field. Even the most highly centralized clusters were weak in conveying pitch center and source position compared to the background sound. Stray agents in small percentage were sufficient to interfere with imaging. This lack of distinction is likely a result in part of the limited sound palette of the SDP. Many audible attributes were not included in data-driven transformation, and a higher-order sound pattern was not applied. The audible profile of the simple tones does not provide strong coherence for rendering spatial relationships among the agents.

Interpretation: The distributed sound source sonification is ineffective, in particular when compared to visualization features. Applying agent data to modulate multiple audio parameters would likely improve audible imaging of source location and pitch center. Interaction with the swarms improved recognition of audible features but the sound alone could not be used to perform accurate "blind" interpretations of swarm patterns.

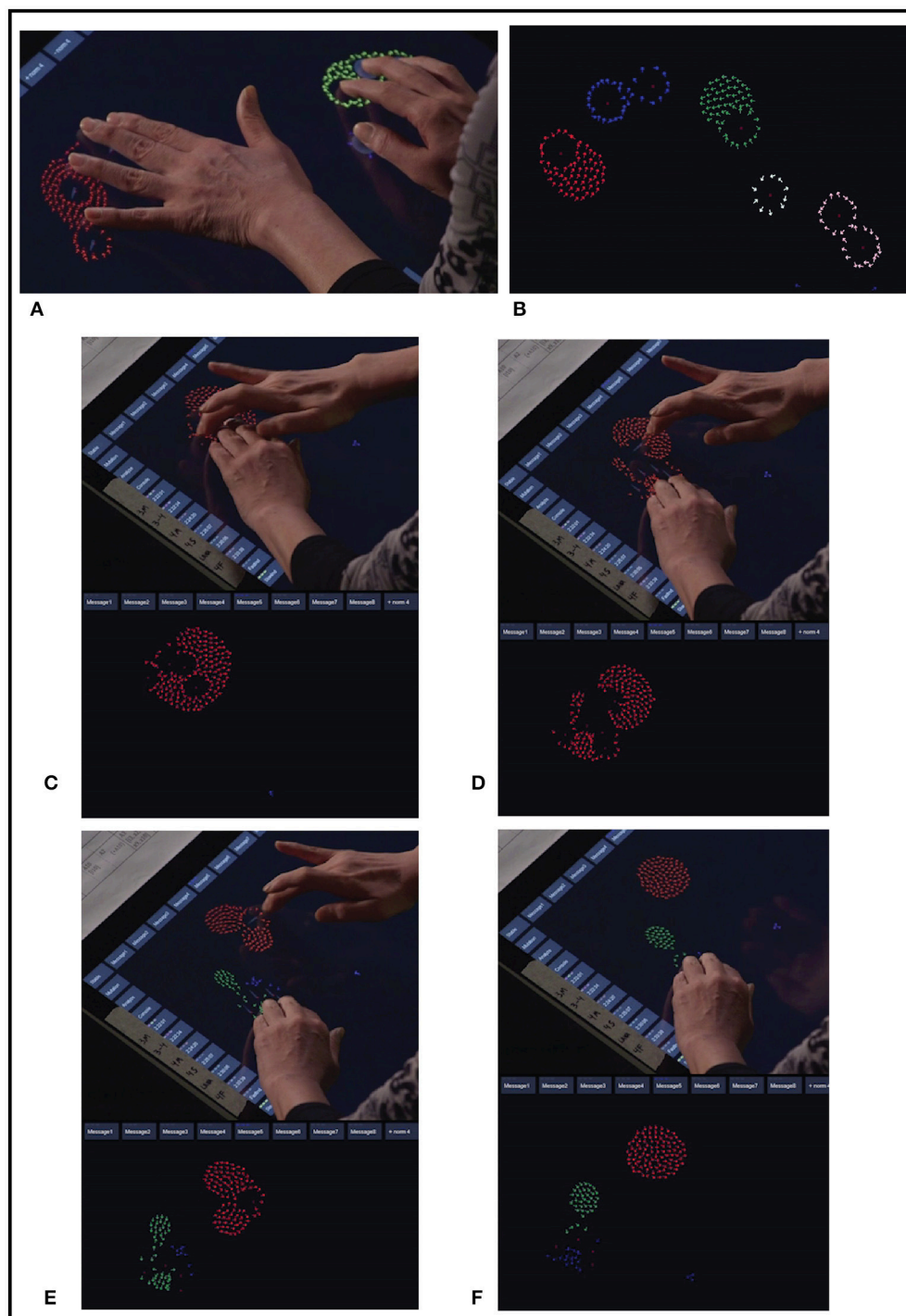
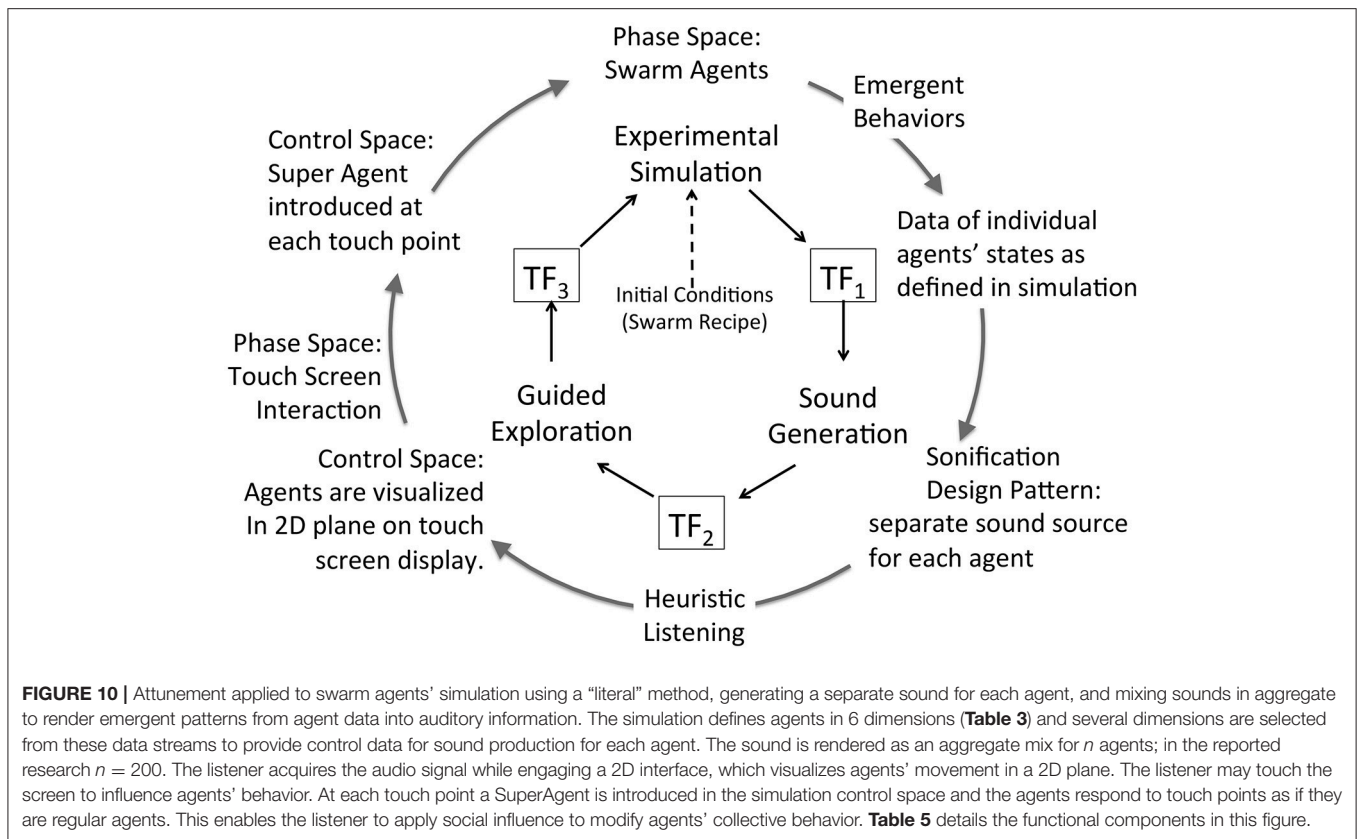


FIGURE 9 | Two examples of Swarm Chemistry emergent behavior induced by an observer's interaction. Swarm agents are visualized on a touch screen interface. Each touch point generates a SuperAgent in the swarm simulation control space. Agents' autonomous responses are determined by their behavioral rules for social interaction. Feature recognition is applied external to the simulation and is used to visualize agent clusters using color. **(A,B)** Two views of swarm agents gathered around 10 touch points from observer's interaction. In **(B)** five clusters are recognized by feature detection and visualized using color. Swarm clusters are an emergent property formed by agents' collective behavior. Individual agents are not aware of their membership in clusters. **(C–F)** A sequence illustrating induced bifurcation of one agent cluster into two clusters. In **(C,D)** an observer applies touch to agents in one cluster and leads the agents in two separate directions. In **(E)** some agents' collective movements break away from the original cluster and form a second cluster. Feature detection is applied external to the simulation and the two clusters are visualized using two colors. In **(F)** the observer has removed the touch points and the two clusters exhibit symmetry through the agents' social interaction.



Case 3: TF₁ Attunement Using Feature Recognition Data

As an alternative to direct sonification of each agent, sonification may use data generated by pattern recognition techniques applied to statistical analysis of agents' collective behavior. Cluster formation is a common emergent pattern, occurring when agents separate from a large swarm or gather from dispersed positions forming one or more cohesive subgroups. A swarm may self-organize into a variable number of clusters and undergo autonomous phase transitions where one cluster spontaneously separates into two, or two clusters come into proximity and merge into one. Symmetry and asymmetry of cluster shape is another common emergent pattern, with clusters achieving circular shapes at some times and other times dynamically deforming along the x- or y-axis. Change of density is another common emergent pattern, varying the number of agents in a cluster with respect to the cluster's geometric area (its visible "size"). This attunement is illustrated in Figure 11 and Table 5 details functional components.

As swarm agents collectively generate patterns the individual agents' actions exhibit emerging aggregate social dynamics across subsets of agents. Sonification can reflect different levels of information, from individual agents' details to emergent collective patterns. Sonification design patterns used in Case Study 3 expose multiple levels of detail in sound production and enables design of a scalable relationship between level of detail in data and level of detail in sound transformation. Adjusting the

relationship between data and SDP can modify level of detail that is transforming sound.

Attunement at TF₁: A pattern recognition process is located in TF₁ to detect and quantify swarm aggregate features (Figure 11). Target patterns are identified in advance from a selection of known emergent features. A recognition method is used to dynamically detect pattern formation in agents' positions. Data of all agents' positions is routed to the recognition process where the number and membership of clusters is determined at each time step. Each cluster is measured in area (visible size), density (number of agents/area), and symmetry or asymmetry (circular or deformed shape). The resulting pattern data is applied to control the SDP, which are designed with affordance for multiple clusters. Additional sonification is applied to phase transition events, for example signaling the separation, merging, and deformation of clusters.

Listening Scenario: The sonification design hypothesized that sound sources could be used to represent clusters and that listeners could understand audible features observed up to a limit of four concurrent sound sources, representing the largest four clusters. We further hypothesized that sound transformations can mirror visual transformations to enhance multimodal attention focus. SDP were designed in relation to the target feature set. For each target pattern, measurements used a single-value normalized scale [0,1] to quantify recognition, and this data controls the sound

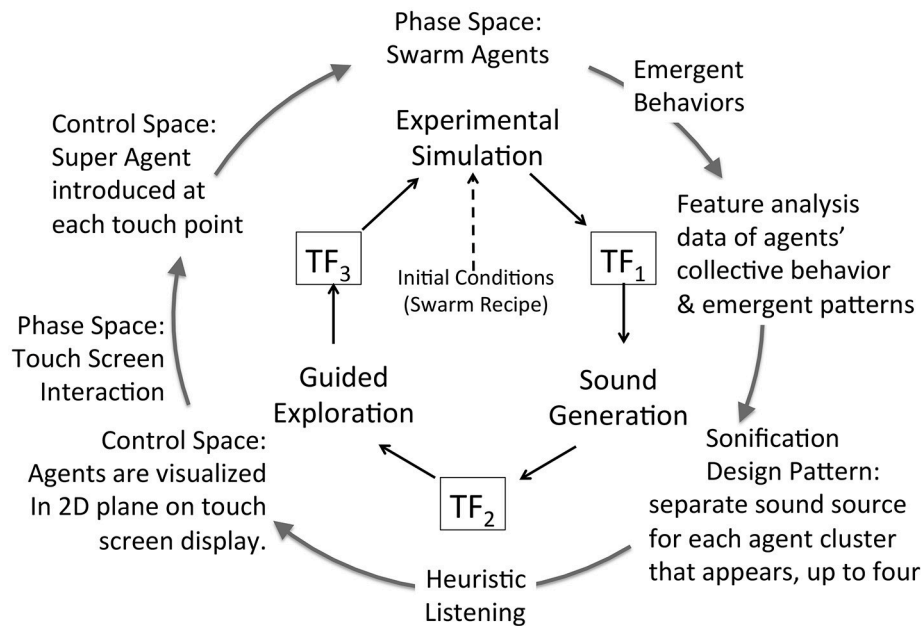


FIGURE 11 | Attunement applied to swarm agents' simulation using a pattern recognition method. Cluster formation among subgroups of agents is the primary feature reported. Secondary statistical data about each cluster is also reported, such as average velocity. Several dimensions are selected from these data streams to provide control data for sound production. The four largest clusters are selected as the sources of control data, and an independent sound source is generated to sonify the data of each cluster, up to four clusters. The sounds are transformed in direct correspondence to feature data measured in each cluster. The listener acquires the audio signal while engaging a 2D interface, which visualizes agents' movement in a 2D plane. The listener may touch the screen to influence agents' behavior. At each touch point a SuperAgent is introduced in the simulation control space and the agents respond to touch points as if they are regular agents. This enables the listener to apply social influence to modify agents' collective behavior. **Table 5** details the functional components in this figure.

palette. Intermediate values between several target patterns generate corresponding interpolations of SDP related to each pattern.

Observed Performance: Applying feature detection data to control sonification design patterns, the audible transformations clearly corresponded to visible features that were measured by pattern recognition. In addition the sounds' qualitative differences enhanced the visualization by enabling fine-grained audible comparisons of relative size and dynamic properties of clusters. Listeners did not report confusion from four sound sources, in part because of coupling to visual cues. The range and variety of sounds enhanced quantitative understanding of the swarms without requiring a one-to-one relationship between the number of agents and the number of sounds. Sounds and their transformations were designed to represent the range and variety of features that comprise the target patterns.

Interpretation: Sonification design patterns modulated by cluster feature recognition is effective, in particular in providing sounds that correspond to visualization features. The number of sound sources required may be independent of the number of agents, and a modest number of sound sources can sonify the features of large population swarms. A weakness of this method is that swarm behaviors that are not part of predetermined target patterns are not emphasized in the sonification.

DISCUSSION

Perceptualization is concerned with an observer and her disposition with respect to objects of study. Framing interactive sonification includes models of listening as well as models that exhibit emergent behavior. The aim is to demonstrate feasibility of extensible modeling for interactive sonification and feasibility of applying a sonification framework to biological information. The proposed sonification framework provides circular causality as a signal pathway for modeling a listener interacting with an experimental system that generates emergent behavior. A framework for interactive sonification is a step toward community development of a theory of sonification, which is underdeveloped in a growing field. For example, the research area of real time EEG sonification, has seen publication growth more than 5 times (from 25 to 140) between 2002 and 2012 (Väljamäe et al., 2013).

Biological information is very likely to exhibit unpredictability such as non-linear and chaotic behavior as well as experimental system fluctuations and noise. To help disambiguate these data conditions scientific observation methods often provide numerical reference models and simulations. Adopting this approach, the sonification framework provides methods developed with models and simulations that exhibit well-known behaviors of biological systems. Working with biologically inspired simulations offers opportunity to exercise attunement

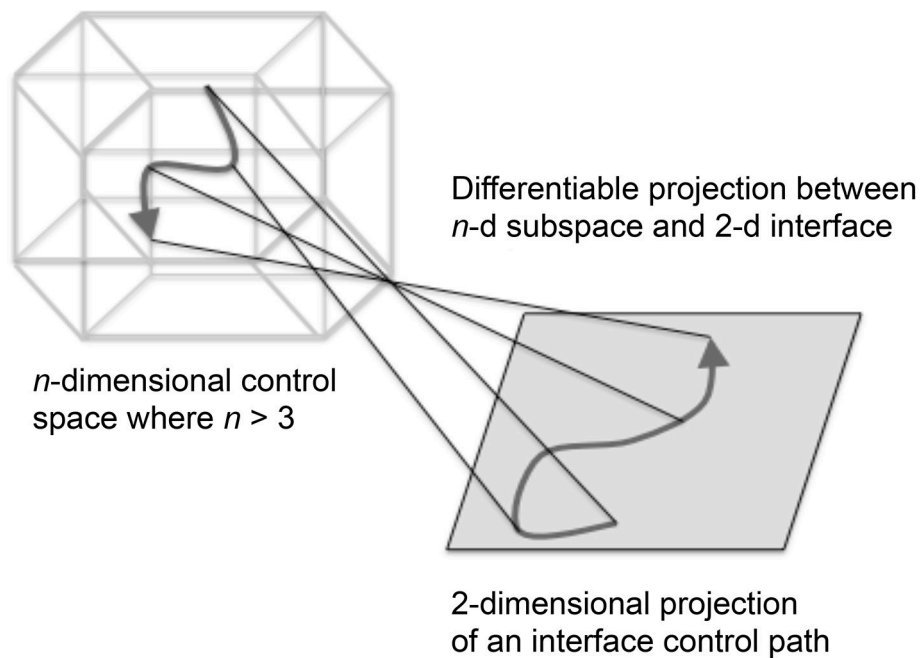


FIGURE 12 | Schematic illustration of a manifold interface with a control path and four fiducial points. A manifold is a locally Euclidean topological space representing linear ranges of n control parameters in n -dimensional Euclidean space, where each point in the space is an n -tuple of real numbers corresponding to a unique set of control values of a parameterized system. An n -dimensional control path is illustrated on the left and the path on the right is generated by differentiable 2D projection from a bounded sub-region of nD space. The 2D actuation space is differentiable and bi-directional with the nD sub-region. Fiducial points are indicated by the endpoints of the four lines projected between spaces.

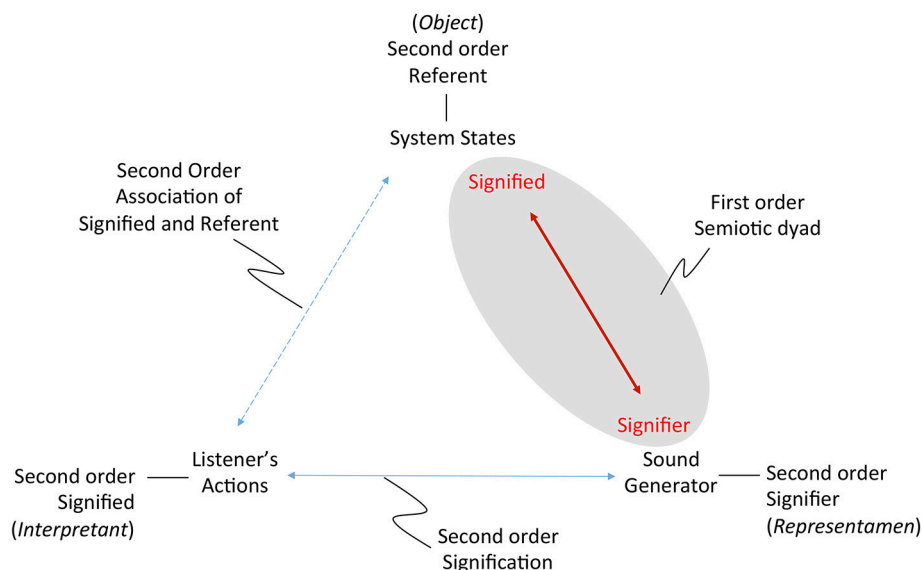


FIGURE 13 | Semiotic relationships of the sonification framework. In the first-order semiotic dyad, data-driven sounds signify experimental system states. In the second-order semiotic triad, data-driven sounds signify the listener's actions that bring about the system states. The listener's actions are identified as a second-order signified. In attunement the two semiotic layers are concurrent. Terms in italics originate from Peirce's triadic semiotic model (Peirce, 1955). The vertices of the semiotic triad align with the component subsystems of the sonification framework in **Figure 3**.

with an interactive architecture addressing the entire application context including the listener's mode of interaction in the loop. This larger context accounts for observation as a secondary

information system coupled to the primary data source being studied. Attunement articulates the coupling between models of observation and models exhibiting unpredictable behavior,

TABLE 5 | Sonification Framework implementation for Swarm Chemistry.

Function	Component	
Experimental simulation	Control space	Total number of agents is defined as a constant; 6 control dimensions initialize agents' social behavior; heterogeneous agent types may be initialized with different control values to increase control space dimensions by $6 \times n$ agent types
	Phase space	Agents' autonomous movements in a bounded 2D plane
	Sample space	Data of individual agents' movements; Statistical analysis of agents' behaviors; feature detection of phase transitions such as emergent cluster formation
TF ₁	Agent data and swarm aggregate feature data are translated by multidimensional mapping to create sound generator control data	
Sound generator	Control space	Sonification Design Patterns receive control data from swarm agent data and feature analysis data
	Phase space	Real-time sound synthesis with integration of recorded sound sources and digital signal processing
	Sample space	Sonification Design Patterns with transformations controlled by agent data
TF ₂	Listener acquires sound output and engages swarm visualization touch screen interface. Listener is attentive to explore agents' behaviors and agents' responses to perturbations guided by auditory features	
Guided exploration	Control space	Agents are visualized and displayed on a multi-touch screen. Listener's touch points in the swarm introduce "super agents" into the simulation
	Phase space	Listener observes agents' autonomous movements and determines when, where, and how to perturb agents
	Sample space	Listeners' actions applied to the multi-touch screen are transmitted to TF ₃ as 2D (x,y) positional data streams
TF ₃	Each touch position is added as a super agent to the simulation control space, with the super agent's behavior determined by the listener, not by simulation control parameters. Autonomous agents respond to a super agent as they respond to all other agents. TF ₃ enables a listener to introduce perturbations into the simulation via engagement directly with simulation Phase space	

TABLE 6 | Comparison of Attunement parameters for Chua's circuit and Swarm Chemistry, based on the affordance Compatible Temporality of Events.

Affordance Type: Compatible Temporality of Events	Data Type: Frequency	
	Chua	Swarm
Measurement of Attunement configuration		
Control rate: Experimental Simulation Control Space	15–20 Hz	10 Hz
Frequency rate: Experimental Simulation Phase space	20 kHz	20 Hz
Sample rate: Experimental Simulation Sample space	44.1 kHz	12–15 Hz
Control rate: Sound Generation Control Space	15–20 Hz	20 Hz
Frequency rate of Interface update in Guided Exploration	24–30 Hz	15–30 Hz
Temporal Ratio in TF ₁ : Experimental Simulation sample rate/Sound Generation control rate	1/1	1/2
Temporal Ratio in TF ₃ : Guided Exploration sample rate/Experimental Simulation control rate	1.5/1–2/1	1.2/1–1.5/1

(See **Table 1**, *Affordances of sonification listening scenarios*).

to increase reliability in sonification applied to non-linear and chaotic systems.

In a discussion of bioinformatics Biro introduces a distinction between biological signal, data, and information (Biro, 2011). A signal emitted by a biological system is initially "data translocation" and becomes "information transmission" only when biological receptors exhibit local state changes in response to the signal. Data is transmitted throughout the system; information is transmitted only at points of responsive reception. The recipient mechanism determines what data is information. Biro's use of "information" suggests the etymology of "inform" from the Latin *informare* meaning "to give form to."³ In sonification, data is formed into information when a listener is attentive to the sound and can associate it to a data source. Biro describes reception of information as a semiotic

function of state change and system response. Semiotic principles provide a perspective for understanding how listeners may disambiguate sounds having representative meanings. Appendix 7 in Supplementary Material discusses related semiotic functions in sonification.

Performance Assessment of the Proposed Sonification Framework for Representing Multiple Experimental Cases

To assess accuracy and relevance of the sonification framework, **Tables 4, 5** compare three case study implementations across each of the subsystems' functional components, showing correspondence to the models illustrated in **Figures 4A–C**. The Case study implementations also align with the sonification framework subsystems illustrated in **Figure 8**, corresponding to **Table 4**, and **Figures 10, 11** corresponding to **Table 5**. The framework provides a reference model for analyzing sonification

³<http://www.etymonline.com/>

TABLE 7 | Listening scenario comparison of two sonification implementations using classification by affordances.

Listening Scenario Comparison using Classification by Affordance		
Indicative class of users experience	Chua's circuit	Swarm chemistry
Affordance 1: Sound for a domain investigation Means: Sound generation and sound design		
What is heard	A continuous electronic signal	Sound samples and Synthesized sounds
Range of audible change	Frequency range averages from tenor to soprano; Tone quality range from pure pitch to pure noise without pitch center; Rhythmic patterns both regular and irregular are imposed upon the frequency and tone	Sonification Design Patterns express data through multidimensional transformations including audible primitives such as frequency, duration, and timbre as well as sound events and sequences
Relation to experimental data	Sound represents oscillation dynamics and emergent properties that are signatures of non-linearity and chaos	SDP provide an aggregate audible summary of high level features; SDP are transformed by data from individual agents and groups of agents
Affordance 2: Interface for domain investigation Means: Control paradigm		
Interface presentation	Graphical user interface with cursor. Capture of cursor positions as control points. Linear potentiometers to control individual parameter values.	Graphical visualization of 2D agents autonomous movements in real-time
Degrees of freedom	2DoF as cursor moves in bounded area	4DoF afforded by two hands each having 2DoF for independent (x,y) position. Each hand position is extended by multi-finger touch constrained by limits of wrist rotation
Multidimensional control	Highly efficient Manifold interface—position of 2D cursor in a control area determines values of seven control parameters simultaneously	Touch interaction in agents' phase space provides indirect access to simulation control space. No direct modification of swarm agent control parameters is enabled for exploratory interaction
Affordance 3: Temporal relationship of events Means: Scheduling of data		
Latency from simulation control space to phase space	Very low latency—Instantaneous response	Agents' movements are updated in real time instantaneously. Emergent features evolve over extended durations of multiple seconds.
Latency across TF_1 :	Very low latency—Instantaneous response	400 ms to 2 s. Audio response to agents requires statistical analysis.
Ratio of temporal resolution of data to sound	Nearly 1 to 1	M to N. Agents' individual temporal dynamics are applied to transform an independent number of sound sources
Affordance 4: Multisensory information Means: Engaging two or more modalities of an observer		
Visible interface	GUI's discrete spatial locations in control area represent simulation control states; linear distances represent relative change of control states; marked points in control plane aid user memory of surrounding regions	Visualization of agents movements
Data visualization	None	Visualization of agents movements
Responsive interface	Audible change synchronizes with interface cursor movement	Direct touch screen interaction with visualization of agents, while listening to audible responses
Affordance 5: Complexification Means: Increasing or reducing level of detail under user control		
Ratio of simulation data to sound control data	1 to 1	1 to 1 in the first Swarm Chemistry case study. 300 to 4 in the second Swarm Chemistry case study
Reduction/compression of experimental data	Sample rate conversion of data to sound at 44.1 kHz; direct representation of data as sound	1 to 1 in the first Swarm Chemistry case study. Extensive statistical analysis applied in the second case study, reporting 5 global measures for overall agents' behavior and 4 measures for each agent cluster up to four clusters

(Continued)

TABLE 7 | Continued

Listening Scenario Comparison using Classification by Affordance		
Indicative class of users experience	Chua's circuit	Swarm chemistry
Affordance 6: Learning Provision Means: Rehearsal for developing competence		
Access to example scenarios	Yes, an initialized Manifold interface with generating points is provided for exploration	Yes, an initialized Swarm Chemistry simulation with prepared sounds was provided for exploration
Rehearsal time given	Yes	Yes

implementation. For example the framework may be applied to disambiguate the model of Sound Generation for the Chua's circuit, which unlike the Swarm Chemistry sonification does not apply a separate procedural sound generator nor sonification design patterns. Instead audible sonification is generated by D-to-A conversion of the Chua's circuit oscillation. Experimental data that is converted directly to sound have been regarded as a special class of sonification, indicating the need for a more inclusive model in order to perform comparative assessment. By adopting the circuit signal as a direct sound source, the relationship of data to sound may collapse into a trivial—and ambiguous—representation. Presented in the framework component model of a sonification design, the circuit functions as a signal generator component in two subsystems, Experimental Data Source and Sound Generator (Figure 8). The circuit's dual position resolves ambiguity by referring to an underlying data acquisition model (Figure 2) that is shared in all subsystems. The shared structure of signal generator allows for two framework components to share an implementation of the circuit simulation.

The framework also aids comparison of case studies to identify common sonification structures having different experimental implementations. Case Studies 2 and 3 use different models of sound generation; one adopts a simple mapping of data to sound, the other applies sonification design patterns. The sonification framework provides a structure at transfer function TF_1 for representing attunement between experimental data and sound generation. Multiple alternative sound generator implementations can be compared at TF_1 (see Figures 10, 11), in terms of measuring how the sound generator is tuned to the control data and how the tuning satisfies requirements of a listening scenario.

The sonification framework provides a structure to interpret and compare attunement for Chua's Circuit and Swarm Chemistry. Table 6 identifies sonification components to define affordances based on temporality of events. At each of these components the two simulations are compared in terms of frequency measurement and frequency ratio. The degree of similarity in frequency provides quantitative comparison of how two listening scenarios are defined by temporal affordance. The selection of attunement configuration parameters in Table 6 was determined by their relevance for measuring impact of processing frequency across the two simulations.

Table 7 provides a comprehensive comparison of the affordances that define listening scenarios for Chua's Circuit

and Swarm Chemistry sonifications. Metrics and quantitative measurements can be devised for many of the fields in Table 7, including data that measures user interactions. The framework enables the impact comparison of each affordance across multiple listening scenario implementations.

The framework also identifies how a common function may be implemented in different subsystems. For example in both Chua's Circuit and Swarm Chemistry, attunement methods are applied to increase the reliability of sonifying emergent behavior. In Chua's Circuit emergent behavior is tuned at transfer function TF_3 , using fiducial points located in the user interface. In Swarm Chemistry emergent behavior is tuned using automated feature recognition to control sonification design patterns, located at transfer function TF_1 .

In summary, a sonification framework identifies structural requirements and functional components to support comparisons of multiple implementations using measurement of attributes and related affordances. An extensible framework that can be used to establish symmetry for measuring and comparing system implementations, can further be applied as a reference model for measurement and comparison of user interactions and user experiences with diverse sonification systems. Table 8 uses the sonification framework to summarize structures that align and variations in component implementation across three case studies. Structural symmetries identified in the framework can be used in the future to define human performance metrics, measuring and comparing user experience and productivity outcomes across multiple sonification system.

Sonifying Emergence With Attunement and Interactive Exploration

Emergent behaviors in non-linear dynamical systems are only partially predictable in best-case scenarios. A sonification framework facing probabilistic behavior intends to provide an affordance to observe all possible states of a system. The tension between indeterminate behavior and reproducibility poses a dilemma in terms of sonification objectives:

1. Reproduce experimental system behaviors as audible signatures;
2. Render the salient features of emergent behaviors.

Restated as a methodology problem: Emergent features may not provide a data model for linear coupling to predetermined audible features. Addressing this problem, two approaches for

TABLE 8 | Summary of case studies of emergent behavior sources, attunement techniques, tuning functions, and outcomes.

System elements and state data	Attunement techniques for sonification design	TF ₁ Tone production for sonification	TF ₂ Listener's actuation of state changes	TF ₃ Experimental data source state and transformation	Attunement technique and outcomes
Chua's Circuit	Seven circuit elements that determine states & control oscillation	System states generate oscillation data. TF ₁ applies rescaling to convert data stream to audible frequency and amplitude range.	Manifold Interface projects 7D Control Space as a 2D plane. Listener manipulates cursor in 2D to actuate 7D Control Space	User's actions modify voltage control data for 7 circuit elements, inducing oscillation state changes	Fiducial points provide stable reference points with stable auditory signatures. A set of fiducial points identify stable states to aid exploration of boundaries of unstable regions
Swarm Simulation	Autonomous 300 Social Agents moving in a bounded plane	"Literal" method: For each agent one sound is generated and the aggregate sound mixture reflects emergence	Actuation applied to system Phase Space: Listener observes a swarm simulation visualized on a 2D touch screen and interacts with the swarm by touching the swarm agents in the interface while hearing sounds.	Swarm simulation converts Listener's touch into a "Superagent" and all simulated swarm agents respond to the Listener's actions using their regular rules for agent social behavior. Swarm behaviors and patterns change as a result	Aggregating 300 sound sources as a literal sonification of agents' positions does not yield audible emergent patterns that correspond to the visible emergent patterns of agents
	Pattern recognition method: Target emergent patterns are preselected and a sound pallet is assigned to targets	Procedural Sound generated based on swarm cluster attributes and events. Preselected sound pallet is applied to salient features and transformations			Predetermined sound palette is transformed by data from pattern recognition engine; emergence of target patterns and events is translated into sound palette mix and synchronizes well with visual events

Column one indicates the system elements that collectively exhibit emergent behavior. Column two indicates attunement techniques for sonifying emergent patterns.

mapping data to sound have been presented. One approach is to preselect sound to represent known features and use these audible signatures to define boundaries of unstable regions, aiding reliable exploration by close association. This is demonstrated in Case Study 1, the Chua's circuit sonification using fiducial points to anchor the listener's interface with stable boundaries of unstable regions. Appendix 5 in Supplementary Material presents a related technique using attunement to compensate for hysteresis. The other approach is to render sound to represent all data points and listen for patterns in the aggregate, applying a design that aims to ensure emergent data features will generate parallel emergent audible features. This is demonstrated in Case Study 2, the Swarm Chemistry sonification using a separate sound source to represent each of 200 swarm agents. Neither approach offers a complete solution to the study of emergence. Case Study 3 demonstrates the application of sonification design patterns to provide strategies that aim to close the gap between these two approaches.

Modes of interactive exploration expand the conditions for interpreting sonification by enhancing the listener's measure of temporal dynamics. As a listener explores a dynamical system, behaviors and sounds co-evolve. Dynamic navigation across control space provides a context to anchor sounds by association to movement and transformation. Heuristic listening involves attentive movement dynamics that complement the dynamics of control states and system responses. Modeling a listener provides criteria for measuring audible feature identification, both to assess target acquisition in known states of experimental systems, and to identify the salience of new features in emergent behaviors. Interactive exploration optimizes for emergence, with exploration supported by attunement.

Feasibility of Applying Sonification Attunement to Data of Biological Signals

Future work will study feasibility of sonification attunement applied to biological signals, anticipating a two-phase approach: (1) apply the framework to sonification of a biological reference model; (2) in the sonification framework replace the reference model with a biological data source. The relative instability of biological systems presents challenges. Measurements of biological information can experience signal fluctuations introduced from an experimental apparatus. Noise may be introduced from data recording instrumentation and the surrounding environment. A biological system's states during a data acquisition trial period are unlikely to remain in a narrow mean that represents a constant value. Experimentally recorded biological data often requires disambiguation of information from noise. In line with research practices that use models and simulations as references for comparison of noisy data, sonification attunement adopts models to generate audible reference features for comparison to data that exhibits unstable system behaviors.

Physical constraints of biological systems may require specialized adaptation of the sonification framework. Feasibility study of an experimental biological system is required prior

to direct application of interactive sonification. Working from a simulation to an experimental system enables comparison of workflows—of the model and the physical system—to determine symmetry between the experimental design and the simulated control configuration. Instrumentation determines where and how attunement may be applied to extend an experimental workflow. For biological experimental systems, real-time attunement feedback from tuning functions TF_2 to TF_3 may be challenged by physical limitations of interaction. Response characteristics of experimental biological systems may limit the capacity for real-time exploration. Time latency required to actuate state changes in a biological system may reduce the listener's sense of interaction. Establishing a parameterized exploration space for experimental acquisition of biological information requires system precision for inducing and measuring state changes. Attunement utilizes initial system exploration to identify salient features and boundaries of unstable regions. With biological information, the initial exploration process is qualified by the control parameters of the observing apparatus. Constraints in implementation of experimental control space will qualify the initial exploration of the system, which is required to identify salient features of emergent behaviors. Detecting emergence will depend upon the instrumentation and the ability to identify fiducial points in experimental control space. Finally, significant latency in experimental systems may impede interactive exploration required for heuristic listening.

Summary of Sonification Framework Requirements

Sonification as data driven tone production requires interpretive and representative techniques for perceptual relevance, therefore generates design requirements. Equally, sonification requires rigor to accurately interpret and represent data or systems with reproducibility, therefore generates scientific requirements.

The sonification framework provides a reference model to generate requirements for implementation of interactive data processing coupled to sound generation. The framework is designed as a canonical model of interactive sonification, providing a small number of variables to represent the model, a simple tripartite structure based on symmetry of data flow, and a reusable template that can be applied to many systems. The framework supports an attunement process to provide solutions for sonification of unpredictable data of non-linear and chaotic systems. The framework adopts a tripartite semiotic structure, which constructs the position of a sonification listener analogous to the position of a bioinformatician (**Figure 13**). The flexibility of the canonical model is demonstrated with two models that exhibit characteristics of biological systems, Chua's circuit and Swarm Chemistry.

To conclude, the sonification framework may be summarized as a set of requirements for design, implementation, and application of attunement.

Architecture Requirements:

1. A controllable experimental data source that exhibits emergent behavior output as a digitized signal;

2. A serial relationship of three subsystems: an active observer controlling a dynamical system, the observed dynamical system generating experimental data, and a sound generator responding to the data to generate information acquired by listening;
3. An interactive signal path that provides circular causality across the three subsystems, defining an explorable space for active listening;

Functional Requirements:

4. An interface to vary control parameters in real time for inducing changes in the experimental data source;
5. A sound generating engine with data driven control parameters;
6. Engineered distinction between data (elicited from a signal) and information (interpretation of a received signal);
7. Three transfer functions (tuning functions) where conversion between data and information is performed for routing control signals from one subsystem to another;

Procedural Requirements:

8. For the dynamical system that is the experimental data source, enable observers to induce changes in the system states;
9. Measure the output signals of the experimental data source to identify salient features and emergent behaviors, and transmit this information across the framework;
10. Given salient features in an output signal, enable observers to annotate the related system states by creating fiducial points in the user interface;
11. Given unstable behavior in the experimental system, identify control space boundaries of unstable regions and apply fiducial points in the user interface to mark the boundaries;
12. Apply data output by the experimental system to provide control of the sound generator, such that fiducial points have recognizable associated sounds;

User Experience Requirements:

13. For exploring an experimental system, provide gestalt orientation for listeners to learn system behaviors through multimodal experiences;

14. For multimodal engagement, provide coupling through the experimental data source, between affordances of sound production and affordances of the user interface;
15. To generate a requisite variety of sounds, adopt sound synthesis design to generate a range of audible transformations of sounds with respect to a domain of measured changes in the experimental data;
16. To connect outputs of data pattern recognition to a semiotic function in sound, design audible signatures for known salient features of emergent behaviors.

Attunement implemented with pattern recognition provides a hybrid methodology to support reproducible observation, identification and feature discernment across multiple types of dynamic data sources using multiple types of sonification. The framework provides an efficient and extensible reference that integrates models of emergent behavior and models of a listener's attentive interaction with data. It may be applied to compare diverse sonification systems and applications, to identify common functions implemented in different subsystems, and to compare the impact of affordances across multiple implementations of listening scenarios.

AUTHOR CONTRIBUTIONS

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

ACKNOWLEDGMENTS

The author would like to thank Leon Chua and Hiroki Sayama for their scientific and engineering guidance; Robin Bargar, G.Q. Zhong, Arthur Peters, Jeff Meyers, and Kevin Bolander for technology support.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fnins.2018.00197/full#supplementary-material>

REFERENCES

- Alain, C., Snyder, J. S., He, Y., and Reinke, K. S. (2007). Changes in auditory cortex parallel rapid perceptual learning. *Cereb. Cortex* 17, 1074–1084. doi: 10.1093/cercor/bhl018
- Albrecht, A. R., and Scholl, B. J. (2010). Perceptually averaging in a continuous visual world: extracting statistical summary representations over time. *Psychol. Sci.* 21, 560–567. doi: 10.1177/0956797610363543
- Alexander, C., Ishikawa, S., and Silverstein, M. (1977). *A Pattern Language: Towns, Buildings, Construction*. New York, NY: Oxford University Press.
- Andersen, R. A., Musallam, S., and Pesaran, B. (2004). Selecting the signals for a brain-machine interface. *Curr. Opin. Neurobiol.* 14, 720–726. doi: 10.1016/j.conb.2004.10.005
- Bargar, R., Choi, I., Das, S., and Goudeseune, C. (1994). “Model-based interactive sound for an immersive virtual environment,” in *Proc. 1994 International Computer Music Conference, ICMC 1994* (Aarhus).
- Begault, D. (1994). *3D Sound for Virtual Reality and Multimedia*. Academic Press; NASA TM-2000-209606. Available online at: <http://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/20010044352.pdf>
- Best, V., Ozmeral, E. J., and Shinn-Cunningham, B. G. (2007). Visually-guided attention enhances target identification in a complex auditory scene. *J. Assoc. Res. Otolaryngol.* 8, 294–304. doi: 10.1007/s10162-007-0073-z
- Biro, J. (2011). Biological information—definitions from a biological perspective. *Information* 2, 117–139. doi: 10.3390/info2010117
- Bohadana, A., Izbicki, G., and Kraman, S. S. (2014). Fundamentals of lung auscultation. *N. Engl. J. Med.* 370, 744–751. doi: 10.1056/NEJMra1302901
- Brechmann, A., and Scheich, H. (2005). Hemispheric shifts of sound representation in auditory cortex with conceptual listening. *Cereb. Cortex* 15, 578–587. doi: 10.1093/cercor/bhh159
- Bregman, A. (1990). *Auditory Scene Analysis: The Perceptual Organization of Sound*. Boston, MA: MIT Press.
- Bregman, A. S. (1978). Auditory streaming is cumulative. *J. Exp. Psychol. Hum. Percept. Perform.* 4, 380–387. doi: 10.1037/0096-1523.4.3.380

- Castiglioni, A. (1947). *A History of Medicine, Trans and Ed. E. B. Krumbhaar*. New York, NY: Alfred A. Knopf.
- Charbonneau, D., Blonder, B., and Dornhaus (2013). "Social insects: a model system for network dynamics," in *Temporal Networks*, eds P. Holme and J. Saramäki (Heidelberg: Springer), 217–241.
- Choi, I. (1994). Interactive exploration of a chaotic oscillator for generating musical signals in real-time concert performance. *J. Franklin Inst.* 331, 785–818. doi: 10.1016/0016-0032(94)90089-2
- Choi, I. (2000a). "A manifold interface for kinesthetic notation in high-dimensional systems," in *Trends in Gestural Control of Music*, eds M. Battier and M. Wanderly (Paris: IRCAM).
- Choi, I. (2000b). "Gestural Primitives and the context for computational processing in an interactive performance system," in *Trends in Gestural Control of Music*, eds M. Battier and M. Wanderly (Paris: IRCAM), 139–172.
- Choi, I. (2014a). "A Priori Attunement for Two Cases of Dynamical Systems," in *Proceedings of the 20th International Conference on Auditory Display (ICAD-2014)*, eds A. Roginska et al. Georgia Institute of Technology/International Community for Auditory Display, Available online at: <https://smartech.gatech.edu/handle/1853/52079>.
- Choi, I. (2014b). Advancing Performability in Playable Media: a simulation-based interface as a dynamic score. *EAI Endorsed Trans. Creat. Technol.* 14:e5. doi: 10.4108/ct.1.1.e5
- Choi, I. (2017a). "Structured reciprocity for musical performance with swarm agents as a generative mechanism," in *Proc. 14th International Conference on Advances in Computer Entertainment Technology, Lecture Notes in Computer Science* (Heidelberg: Springer), 689–712.
- Choi, I. (2017b). "Interactive composition and performance framework with evolutionary computing," in *Proc. 2017 International Computer Music Conference* (Shanghai).
- Choi, I., and Bargar, R. (2011). "A playable evolutionary interface for performance and social engagement," in *Proceedings 4th International ICST Conference on Intelligent Technologies for Interactive Entertainment (INTETAIN 2011)*, Genova, eds A. Camurri and C. Costa LNICST 78, ISSN 1867-8211, ISBN 978-3-642-30213-8, Springer, Heidelberg, 2012, 170–182. Available online at: http://link.springer.com/chapter/10.1007/978-3-642-30214-5_19
- Choi, I., and Bargar, R. (2013). "Between music and games: interactive sonic engagement with emergent behaviors," in *Advances in Computer Entertainment Lecture Notes in Computer Science, Vol 8253*, eds D. Reidsma, K. Haruhiro, and A. Nijholt (Heidelberg: Springer), 519–523.
- Choi, I., and Bargar, R. (2014). "Sounds shadowing agents generating audible features from emergent behaviors," *Proc. Fourteenth International Conference on Synthesis and Simulation of Living Systems* (New York, NY: MIT Press), 726–733.
- Chua, L. (2005). Local activity is the origin of complexity. *Int. J. Bifurcation Chaos* 15, 3435–3456. doi: 10.1142/S0218127405014337
- Cusack, R., Deeks, J., Aikman, G., and Carlyon, R. P. (2004). Effects of location, frequency region, and time course of selective attention on auditory scene analysis. *J. Exp. Psychol. Hum. Percept. Perform.* 30, 643–656. doi: 10.1037/0096-1523.30.4.643
- Downar, J., Crawley, A. P., Mikulis, D. J., and Davis, K. D. (2000). A multimodal cortical network for the detection of changes in the sensory environment. *Nat. Neurosci.* 3, 277–283. doi: 10.1038/72991
- Faure, P., and Korn, H. (2001). Is there chaos in the brain? I. Concepts of nonlinear dynamics and methods of investigation. *C.R. Acad. Sci. III* 324, 773–793.
- Faure, P., and Korn, H. (2003). Is there chaos in the brain? II. Experimental evidence and related models. *C. R. Biol.* 326, 787–840. doi: 10.1016/j.crv.2003.09.011
- Fritz, J. B., Elhilali, M., and Shamma, S. A. (2005). Differential dynamic plasticity of A1 receptive fields during multiple spectral tasks. *J. Neurosci.* 25, 7623–7635. doi: 10.1523/JNEUROSCI.1318-05.2005
- Fritz, J. B., Elhilali, M., David, S. V., and Shamma, S. A. (2007). Auditory attention—focusing the searchlight on sound. *Curr. Opin. Neurobiol.* 17, 1–19. doi: 10.1016/j.conb.2007.07.011
- Gamma, E., Helm, R., Johnson, R., and Vlissides, J. (1994). *Design Patterns*. Boston, MA: Addison-Wesley.
- Haken, H. (1983). *Synergetics: An Introduction*. Berlin: Springer.
- Hansen, C., and Johnson, C. (2005). *The Visualization Handbook*. Oxford: Elsevier.
- Hermann, T. (2011). "Model based sonification," in *The Sonification Handbook*, eds T. Hermann and A. Hunt (Berlin: Logos Verlag), 399–427.
- Hillyard, S. A., Hink, R. F., Schwent, V. L., and Picton, T. W. (1973). Electrical signs of selective attention in the human brain. *Science* 182, 177–180. doi: 10.1126/science.182.4108.177
- Holcombe, A. O. (2009). Seeing slow and seeing fast: two limits on perception. *Trends Cogn. Sci.* 13, 216–221. doi: 10.1016/j.tics.2009.02.005
- Hotting, K., Rosler, F., and Roder, B. (2003). Crossmodal and intermodal attention modulate event-related brain potentials to tactile and auditory stimuli. *Exp. Brain Res.* 148, 26–37. doi: 10.1007/s00221-002-1261-z
- Huang, M. X., Lee, R. R., Miller, G. A., Thoma, R. J., and Hanlon, F. M., Paulson, et al. (2005). A parietal-frontal network studied by somatosensory oddball MEG responses, and its cross-modal consistency. *Neuroimage* 28, 99–114. doi: 10.1016/j.neuroimage.2005.05.036
- Ivry, R. B., and Spencer, R. M. (2004). The neural representation of time. *Curr. Opin. Neurobiol.* 14, 225–232. doi: 10.1016/j.conb.2004.03.013
- Izhikevich, E. (2003). Simple model of spiking neurons. *IEEE Trans. Neural Netw.* 14, 1569–1572. doi: 10.1109/TNN.2003.820440
- James, C. J., and Hesse, C. W. (2005). Independent component analysis for biomedical signals. *Physiol. Meas.* 26, R15–R39. doi: 10.1088/0967-3334/26/1/R02
- Kaneko, K. (2015). From globally coupled maps to complex-systems biology. *Chaos* 25, 097608. doi: 10.1063/1.4916925
- Kaniasus, E. (2012). *Biomedical Signals and Sensors I, Biological and Medical Physics, in Biomedical Engineering*. Berlin: Springer-Verlag.
- Kayser, C., Petkov, C. I., Augath, M., and Logothetis, N. K. (2007). Functional imaging reveals visual modulation of specific fields in auditory cortex. *J. Neurosci.* 27, 1824–1835. doi: 10.1523/JNEUROSCI.4737-06.2007
- Kennedy, P. M. (1993). Three steps to chaos, part II: a Chua's circuit primer. *IEEE Trans. Circ. Syst.* 40, 657–674. doi: 10.1109/81.246141
- Kozma, R., and Freeman, W. J. (2017). Cinematic operation of the cerebral cortex interpreted via critical transitions in self-organized dynamic systems. *Front. Syst. Neurosci.* 11:10. doi: 10.3389/fnsys.2017.00010
- Laennec, R. T. H., and Forbes, J. John, Sir (1835). *A Treatise on the Diseases of the Chest and on Mediate Auscultation*. New York, NY: Samuel Wood & Sons.
- Lange, K., and Roder, B. (2006). Orienting attention to points in time improves stimulus processing both within and across modalities. *J. Cogn. Neurosci.* 18, 715–729. doi: 10.1162/jocn.2006.18.5.715
- Liljenström, H. (2010). "Inducing transitions in mesoscopic brain dynamics," in *Modeling Phase Transitions in the Brain*, eds D. A. Steyn-Ross and M. Steyn-Ross (New York, NY: Springer), 147–177.
- Liljenström, H. (2012). *Mesoscopic Brain Dynamics. Scholarpedia*. Available online at: http://www.scholarpedia.org/article/Mesoscopic_brain_dynamics#Mesoscopic_brain_dynamics_and_consciousness
- Lin, Y.-P., Yang, Y.-H., and Jung, T.-P. (2014). Fusion of electroencephalographic dynamics and musical contents for estimating emotional responses in music listening. *Front. Neurosci.* 8:94. doi: 10.3389/fnins.2014.00094
- Madl, T., Baars, B. J., and Franklin, S. (2011). The Timing of the cognitive cycle. *PLoS ONE* 6: e14803. doi: 10.1371/journal.pone.0014803
- Martikainen, M. H., Kaneko, K., and Hari, R. (2005). Suppressed responses to self-triggered sounds in the human auditory cortex. *Cereb. Cortex* 15, 299–302. doi: 10.1093/cercor/bhh131
- Mattei, T. (2014). Unveiling complexity: non-linear and fractal analysis in neuroscience and cognitive psychology. *Front. Comput. Neurosci.* 8:17. doi: 10.3389/fncom.2014.00017
- McDermott, J. H., Schemitsch, M., and Simoncelli, E., P (2013). Summary statistics in auditory perception. *Nat. Neurosci.* 16, 493–498. doi: 10.1038/nn.3347
- Mealla, S., Oliveira, A., Marimon, X., Steffert, T., Jordà S., and Våljamäe, A. (2014). "The role of personalization and multiple EEG and sound features selection in real time sonification for neurofeedback," in *Proceedings of the International Conference on Physiological Computing Systems, Vol. 1, PhyCS* (Lisbon), 323–330.
- Molholm, S., Martinez, A., Ritter, W., Javitt, D. C., and Foxe, J. J. (2005). The neural circuitry of pre-attentive auditory change-detection: an fMRI study of pitch and duration mismatch negativity generators. *Cereb. Cortex* 15, 545–551. doi: 10.1093/cercor/bhh155

- Morris, C., and Lecar, H. (1981). Voltage oscillations in the Barnacle giant muscle fiber. *Biophys. J.* 193, 193–213. doi: 10.1016/S0006-3495(81)84782-0
- Nagarajan, S., Blake, D. T., Wright, B., Byl, N., and Merzenich, M. (1998). Practice-related improvements in somatosensory interval discrimination are temporally specific but generalize across skin location, hemisphere, and modality. *J. Neurosci.* 18, 1559–1570.
- Pastor, M. A., Macaluso, E., Day, B. L., and Frackowiak, R. S. (2006). The neural basis of temporal auditory discrimination. *Neuroimage* 30, 512–520. doi: 10.1016/j.neuroimage.2005.09.053
- Pearson Education (2006). *Prentice Hall Heart and Lung Sounds Workbook and CD*. Pearson Education.
- Pedersen, J., and Jorgensen, T. (2005). “Localization performance of real and virtual sound sources,” in *The RTO Human Factors and Medicine Panel (HFM) Symposium* (Amersfoort: NATO OTAN Science and Technology Collaboration Support Office; RTO-MP-HFM-123).
- Peirce, C. (1955). “Logic and semiotic: the theory of signs,” in *Philosophical Writings of Peirce*, ed J. Buchler (New York, NY: Dover), 98–119.
- Piazza, E. A., Sweeny, T. D., Wessel, D., Silver, M. A., and Whitney, D. (2013). Humans use summary statistics to perceive auditory sequences. *Psychol. Sci.* 24, 1389–1397. doi: 10.1177/0956797612473759
- Reynolds, C. (1987). “Flocks, herds and schools: a distributed behavioral model,” in *SIGGRAPH '87: Proceedings of the 14th Annual Conference on Computer Graphics and Interactive Techniques*. Anaheim, CA: Association of Computing Machinery, 25–34.
- Rosenboom, D., and Number, L. M. (1997). *Extended Musical Interface with the Human Nervous System. LEONARDO MONOGRAPH SERIES International Society for the Arts, Sciences and Technology (ISAST), no. 1.*
- Sayama, H. (2007). “Decentralized control and interactive design methods for large-scale heterogeneous self-organizing swarms,” in *ECAL 2007*, eds F. Almeida e Costa, L. M. Rocha, E. Costa, I. Harvey, and A. Coutinho (Berlin: Springer-Verlag), 675–684.
- Sayama, H. (2012). “Swarm-based morphogenetic artificial life,” in *Morphogenetic Engineering*, ed R. Doursat, H. Sayama, and O. Michel (Berlin: Heidelberg: Springer-Verlag), 191–208.
- Schiff, S., Jerger, K., Duong, D., Chang, T., Spano, M., and Ditto, W. (1994). Controlling chaos in the brain, *Nature* 370, 615. doi: 10.1038/370615a0
- Silva, C. P. (1993). Shil'nikov's theorem - a tutorial. *IEEE Trans. Circuits Syst. I* 40, 675–682. doi: 10.1109/81.246142
- Spitzer, H., Desimone, R., and Moran, J. (1988). Increased attention enhances both behavioral and neuronal performance. *Science* 240, 338–340. doi: 10.1126/science.3353728
- Sussman, E. S., Horvath, J., Winkler, I., and Orr, M. (2007). The role of attention in the formation of auditory streams. *Percept. Psychophys.* 69, 136–152. doi: 10.3758/BF03194460
- Teng, X., Tian, X., and Poeppel, D. (2016). Testing multi-scale processing in the auditory system. *Sci. Rep.* 6:34390. doi: 10.1038/srep34390
- Tsuda, I., Yamaguti, Y., and Watanabe, H. (2016). Self-organization with constraints—a mathematical model for functional differentiation. *Entropy* 18:74. doi: 10.3390/e18030074
- Ulanovsky, N., Las, L., Farkas, D., and Nelken, I. (2004). Multiple time scales of adaptation in auditory cortex neurons. *J. Neurosci.* 24, 10440–10453. doi: 10.1523/JNEUROSCI.1905-04.2004
- Väljamäe, A., Steffert, T., Holland, S., Marimon, X., Benitez, R., Mealla, S., et al. (2013). “A review of real-time EEG sonification research,” in *Proceedings of the 19th International Conference on Auditory Display (ICAD 2013)* (Lodz), 85–93.
- Voisin, J., Bidet-Caulet, A., Bertrand, O., and Fonlupt, P. (2006). Listening in silence activates auditory areas: a functional magnetic resonance imaging study. *J. Neurosci.* 26, 273–278. doi: 10.1523/JNEUROSCI.2967-05.2006
- Wang, X. J., and Rinzel, J. (1998). “Oscillatory and bursting properties of neurons,” in *Handbook of Brain Theory and Neural Networks*, ed M. Arbib (Cambridge: MIT Press), 686–691.
- Ward, L. (1994). Supramodal and modality-specific mechanisms for stimulus-driven shifts of auditory and visual attention. *Can. J. Exp. Psychol.* 48, 242–259. doi: 10.1037/1196-1961.48.2.242
- Webber, C. L., and Zbilut, J. P. (1994). Dynamical assessment of physiological systems and states using recurrence plot strategies. *J. Appl. Physiol.* 76, 965–973. doi: 10.1152/jappl.1994.76.2.965
- Welch, R. B., and Warren, D. H. (1980). Immediate perceptual response to intersensory discrepancy. *Psychol. Bull.* 3, 638–667. doi: 10.1037/0033-2909.88.3.638
- Wiener, N. (1954). *The Human Use of Human Beings*. Boston, MA: Da Capo Press.
- Woldorff, M. G., and Hillyard, S. A. (1991). Modulation of early auditory processing during selective listening to rapidly presented tones. *Electroencephalogr. Clin. Neurophysiol.* 79, 170–191. doi: 10.1016/0013-4694(91)90136-R
- Zhong, G. Q., and Ayrom, F. (1985). Experimental confirmation of chaos from Chua's circuit. *Int. J. Circuit Theor. Appl.* 13, 93–98. doi: 10.1002/cta.4490130109
- Zhong, G. Q., Bargar, R., and Halle, K. (1994). Circuits for voltage tuning the parameters of chua's circuit: experimental application for musical signal generation. *J. Franklin Inst.* 331B, 743–784. doi: 10.1016/0016-0032(94)90088-4

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.

Copyright © 2018 Choi. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



The Use of Footstep Sounds as Rhythmic Auditory Stimulation for Gait Rehabilitation in Parkinson's Disease: A Randomized Controlled Trial

Mauro Murgia^{1,2*}, Roberta Pili³, Federica Corona⁴, Fabrizio Sors¹, Tiziano A. Agostini¹, Paolo Bernardis¹, Carlo Casula³, Giovanni Cossu³, Marco Guicciardi² and Massimiliano Pau⁴

¹Department of Life Sciences, University of Trieste, Trieste, Italy, ²Department of Pedagogy, Psychology, Philosophy, University of Cagliari, Cagliari, Italy, ³AOB "G. Brotzu" General Hospital, Cagliari, Italy, ⁴Department of Mechanical, Chemical and Materials Engineering, University of Cagliari, Cagliari, Italy

OPEN ACCESS

Edited by:

Diego Minicacchi,
Università degli Studi di Firenze,
Italy

Reviewed by:

George C. McConnell,
Stevens Institute of Technology,
United States
Erwin Van Wegen,
VU University Amsterdam,
Netherlands

*Correspondence:

Mauro Murgia
mmurgia@units.it

Specialty section:

This article was submitted
to Neuroprosthetics,
a section of the journal
Frontiers in Neurology

Received: 01 July 2017

Accepted: 30 April 2018

Published: 24 May 2018

Citation:

Murgia M, Pili R, Corona F, Sors F, Agostini TA, Bernardis P, Casula C, Cossu G, Guicciardi M and Pau M (2018) The Use of Footstep Sounds as Rhythmic Auditory Stimulation for Gait Rehabilitation in Parkinson's Disease: A Randomized Controlled Trial. *Front. Neurol.* 9:348. doi: 10.3389/fneur.2018.00348

Background: The use of rhythmic auditory stimulation (RAS) has been proven useful in the management of gait disturbances associated with Parkinson's disease (PD). Typically, the RAS consists of metronome or music-based sounds (artificial RAS), while ecological footstep sounds (ecological RAS) have never been used for rehabilitation programs.

Objective: The aim of this study was to compare the effects of a rehabilitation program integrated either with ecological or with artificial RAS.

Methods: An observer-blind, randomized controlled trial was conducted to investigate the effects of 5 weeks of supervised rehabilitation integrated with RAS. Thirty-eight individuals affected by PD were randomly assigned to one of the two conditions (ecological vs. artificial RAS); thirty-two of them (age 68.2 ± 10.5 , Hoehn and Yahr 1.5–3) concluded all phases of the study. Spatio-temporal parameters of gait and clinical variables were assessed before the rehabilitation period, at its end, and after a 3-month follow-up.

Results: Thirty-two participants were analyzed. The results revealed that both groups improved in the majority of biomechanical and clinical measures, independently of the type of sound. Moreover, exploratory analyses for separate groups were conducted, revealing improvements on spatio-temporal parameters only in the ecological RAS group.

Conclusion: Overall, our results suggest that ecological RAS is equally effective compared to artificial RAS. Future studies should further investigate the role of ecological RAS, on the basis of information revealed by our exploratory analyses. Theoretical, methodological, and practical issues concerning the implementation of ecological sounds in the rehabilitation of PD patients are discussed.

Clinical Trial Registration: www.ClinicalTrials.gov, identifier NCT03228888.

Keywords: rhythm, ecological sounds, auditory stimuli, rhythmic auditory stimulation, Parkinson disease, gait, spatio-temporal parameters, gait analysis

INTRODUCTION

Individuals affected by Parkinson's disease (PD) typically exhibit motor (e.g., tremor, rigidity, postural instability, gait disturbance) and non-motor symptoms, which progressively affect their quality of life. In order to cope with motor symptoms, patients are generally treated with pharmacological therapies (e.g., L-DOPA, dopamine agonists). However, the symptoms tend to become more severe with the progression of the disease and, at the same time, they become more resistant to medication (1), determining the gradual increase of doses and, consequently, the onset of serious side effects. To optimize the use of medication and cope with patients' impairments, pharmacological therapies are usually accompanied by physical therapy, which is essential for effectively contrasting the motor symptoms and (at least partially) restoring the motor functions. Given that the loss of motor functions increases the risk of falling and gradually affects patients' independence, researchers have directed their attention to the methods enhancing the efficacy of physical therapy. One of them is the rhythmic auditory stimulation (RAS) developed by Thaut and colleagues (2) and widely studied in the past 20 years (3–6).

The RAS method consists in a gait training, in which patients' gait is guided by an auditory rhythm. Typically, patients are provided with auditory rhythms (metronome or music), whose beats per minute (BPM) depends on patients' cadence (steps per minute) at baseline. Usually the BPM is equal to one's own cadence or slightly increased/decreased (e.g., ± 5 –10%), depending on the characteristics of the patients and on the methodological choices of the researchers (7–13). The logic of RAS can be understood by analyzing the source of gait disturbance in PD. The damages of basal ganglia, typical of PD, would compromise the functionality of patients' internal clock, consequently affecting the coordination and the execution of movements (14, 15). Thus, the dysfunctions of the internal clock would be one of the causes of the scarce movement fluidity and of gait impairment. In order to reduce these symptoms, it is necessary to "guide" the internal clock and this can be done by using an external rhythm, that is, RAS. Hence, RAS would facilitate the activity of the internal clock and would help in regulating the fluidity of muscular activation, improving coordination, and facilitating the execution of automatic movements, such as walking. The neural mechanisms underpinning the effectiveness of RAS are not totally clear; it has been proposed that RAS would either rely on residual activity in cortico-striatal circuitry or facilitate compensation by bypassing the damaged areas and relying on alternative pathways (e.g., cerebello-thalamo-cortical circuitry) (16).

The first empirical evidence supporting the efficacy of RAS in the rehabilitation of PD patients was provided by Thaut and colleagues (2). In their study participants were randomly assigned to one of three conditions: RAS, self-paced training, and no training. In both RAS and self-paced training, participants performed daily walking and other exercises for 30 min, for 3 weeks. The only difference was that the RAS group did the exercises with the auditory stimulation, while the self-paced group had no external triggers. The results revealed that both training groups improved in terms of spatio-temporal parameters; however, the RAS group exhibited significantly better results than the other two groups in

step cadence, gait speed, and stride length. Significant results were obtained also as concerns the EMG activity of the leg muscles.

In the subsequent years, the efficacy of RAS has been widely confirmed by many studies [for reviews see Ref. (3–6)]. For instance, it has been shown that during a RAS session there is a close synchronization between auditory rhythm and cadence in both PD and healthy participants, suggesting that rhythmic entrainment occurs even with damaged basal ganglia (17). Other studies investigated the immediate effects of RAS (18–29) and its role in training protocols (8, 9, 11–13, 30–32), manipulating important variables (e.g., number of weeks and sessions, duration of each session, tempo of the stimuli), and the majority of them consistently reported positive effects of RAS. The improvements of training protocols have been observed in different kinds of variables: (1) clinical measures, such as unified Parkinson's Disease Rating Scale (UPDRS), freezing of gait questionnaire (FOGQ), Tinetti test, and timed up and go test (TUG) (10, 11); (2) spatio-temporal parameters of gait, in particular cadence and gait speed (9, 17); and (3) amplitude and timing of the muscular contraction, in particular with regards gastrocnemius, tibialis anterior, and vastus lateralis (2, 12). Recently, improvements have been observed also in terms of kinematics. Indeed, it has been shown that the typical gait of PD patients (33) would be modified by a rehabilitation program integrated with RAS, with the hip flexion–extension movement closer to the normality after rehabilitation (13).

Overall, the positive effects of RAS training based on metronome or music (artificial RAS) are quite well-established in the PD literature. Given that the tempo of RAS is usually determined on the basis of patient's own cadence, in a certain degree RAS represents the perceptualization of biological information associated with gait. However, from the perceptual point of view, the experience of artificial sounds like metronome or music is quite far from the auditory experience of walking. Indeed, artificial RAS only provide rhythmic information, while more ecological stimuli such as footstep sounds would provide both rhythmic information and other gait-related information (e.g., posture, force, gait cycle). Surprisingly, to the best of our knowledge, nobody has explored whether RAS based on ecological footstep sounds can be advantageous, compared to RAS based on artificial sounds (e.g., metronome), within a PD rehabilitation program.

The effects of ecological versus artificial sounds on motor processes have been explored in various domains, such as breathing (34) and motor learning (35), with apparently contradictory results probably because of the different methods employed. In the domain of PD, the role of ecological sounds has been explored in laboratory experiments on walking, revealing interesting results. In one of these experiments, Rodger, Young, and Craig (36) found that the use of synthesized footstep sounds reduced the coefficient of variation of stride length and stride duration, compared to normal walk without sounds. In another study on PD patients (37), the same authors found that the variability of step length and step duration was lower when administering footstep sounds compared to metronome sounds, in a real-time imitation task. This evidence suggests that the complexity of ecological sounds can provide more information than simple beats and this pattern of information can be used as guidance for

walking. Recently, it has been argued that the information conveyed by footstep sounds may have important implications for the enhancement of gait in PD (4, 38), and it has been questioned whether a rehabilitation program based on ecological sounds may be more advantageous compared to a metronome-based program.

The rationale of the hypothesized greater effects of the ecological sounds originates from both neurophysiological evidence and perceptual-motor theories. As regards the neurophysiological evidence, it is well-known that the neurons with mirror properties (39) are associated with imitation (40). Indeed, the main areas of the mirror system (i.e., inferior parietal lobule, precentral gyrus, inferior frontal gyrus) activate when humans see or perform an action, but even when they hear the same action (41–44). Based on this evidence, it seems reasonable to hypothesize that listening to footstep sounds would activate some of the brain regions involved in the control of walking, possibly triggering imitative behaviors.

As regards the perceptual-motor theories, it has been suggested that the perceptual and the motor systems share a common representational organization, and continuously influence each other (45, 46). Thus, the footstep sounds would evoke a representation of walking in the common representational system, which would reinforce a representation of the same gesture already stored in memory, due to previous perceptual-motor experience. The same representation would not be triggered by the metronome sounds, because they are not intrinsically related to walk and would not be able to adequately resemble a human walking representation. As a consequence, footstep sounds would activate a powerful representation of human walking in the common representational system (45, 46), which with a higher probability would influence the corresponding motor outcomes (4, 34), namely, walking. In sum, while the artificial RAS (i.e., metronome) would provide patients with rhythmic cues only, the ecological RAS (i.e., footsteps) would provide patients with rhythmic cues, which are also meaningful and able to evoke a mental representation of walking.

In this study, we aim to investigate whether the hypothesized superiority of the ecological over artificial RAS has an actual impact in a PD rehabilitation program. In particular, we intend to better understand whether the type of sound (i.e., footsteps or metronome) is relevant in a typical rehabilitation protocol including a gait training with RAS. We hypothesize that patients treated with footstep sounds would improve more than those treated with metronome sounds. Indeed, we expect that both groups would benefit from the rhythmic information of the stimuli, but the former would also take advantage of the priming effect elicited by the ecological information of footstep sounds.

MATERIALS AND METHODS

Participants

Thirty-eight individuals affected by PD participated in this study and 32 of them ($M_{\text{age}} = 68.2$ years; $SD = 10.5$) completed it; see **Figure 1** for participant flow diagram and **Table 1** for baseline demographic and clinical characteristics of participants. Patients

were enrolled by RP, CC, and GC between December 2014 and February 2015 at “G. Brotzu” General Hospital (Cagliari, Italy), where they were informed about the present study. The sample size was calculated by means of the G*Power software (parameters: $\alpha = 0.05$; power = 0.80; effect size = 0.25), and the result was 28 participants. All patients included in the study met the following criteria: diagnosis of PD according to the UK Brain Bank criteria (47) ability to walk independently; absence of relevant hearing impairments which could prevent the correct perception of the auditory cues (i.e., ability to have a regular conversation with medical doctors during interview without the use of hearing aids and without the doctor shouting); absence of significant cognitive impairment [i.e., mini-mental status examination (MMSE) >24 ; frontal assessment battery (FAB) >13]; absence of psychiatric or severe systemic illnesses; mild-to-moderate disability assessed by means of the modified Hoehn and Yahr (H&Y) staging scale ($1.5 \leq \text{H\&Y} \leq 3$); and no engagement in any rehabilitative program in the 3 months before the beginning of the study. When participants were recruited, all of them were treated with L-DOPA and five of them were also taking dopamine agonists. The experimental protocol was approved by local ethics committee (Prot. PG/2014/19654). Written informed consent was obtained by all participants.

Stimuli

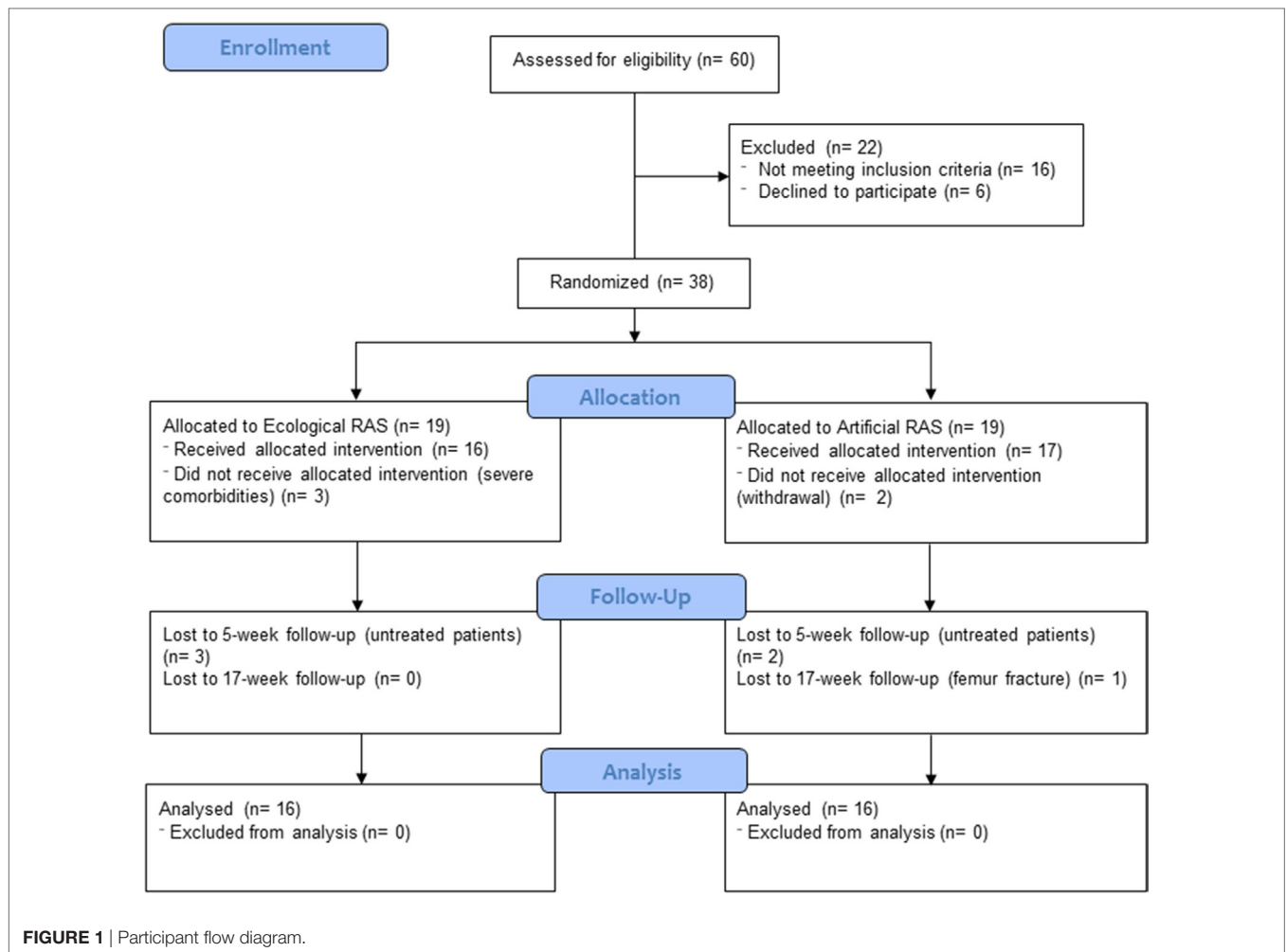
Footsteps Recording

The recordings were carried out in a soundproof room with a parquet floor. Footstep sounds were recorded by means of a fixed-cardioid, large-diaphragm condenser sE2200A microphone. The microphone was fixed on an elastic shock mount in order to isolate it from mechanically transmitted noise; the elastic shock, in turn, was fixed on a stick. The microphone was connected to a M-AUDIO Fast Track Ultra 8R external sound card, which was connected to a laptop running the Logic Pro X software.

A database of footstep sounds was created using the following procedure. Fourteen healthy young adults (7F, 7M) participated in the recording phase. These volunteers were recruited on the basis of their weight. Specifically, females' weight ranged from 45 to 75 kg with intervals of 5 kg, while males' weight ranged from 60 to 90 kg, always with 5 kg intervals. The volunteers were required to wear garments free of synthetic fabrics (to avoid potential side noises) and a pair of their own sneakers with a rubber sole. Each volunteer was required to take six steps of 70 cm at the pace of 100 BPM. A set of strips were marked on the floor to provide the correct distance, while the pace was provided by means of earplugs conveying a metronome sound from a portable MP3 player. The recordings were carried out by an experimenter following the walk of the volunteer from a side, without walking, by using the microphone stick.

Stimuli Editing

Two kinds of RAS stimuli (i.e., ecological and artificial) were created. Ecological stimuli consisted of footstep recordings taken from the above-described database; artificial stimuli consisted of metronome sounds. Each patient was provided with one stimulus, either ecological or artificial (depending

**TABLE 1 |** Baseline demographic and clinical characteristics for each group.

	Ecological rhythmic auditory stimulation (RAS) group	Artificial RAS group
Age (years)	66.5 ± 10.9	69.9 ± 10.1
Parkinson's disease (PD) duration (years)	6.9 ± 5.4	5.8 ± 6.2
Hoehn & Yahr (H&Y)	1.5 ≤ H&Y ≤ 2.5	1.5 ≤ H&Y ≤ 3
Unified PD rating scale (UPDRS III)	18.0 ± 9.1	20.2 ± 9.6
Mini-mental status examination	27.1 ± 1.6	27.9 ± 1.5
Frontal assessment battery	17.1 ± 1.5	16.5 ± 1.2

on the assigned condition); the stimuli were personalized for each patient. In this regard, ecological stimuli were assigned to patients on the basis of their own gender and weight, thus providing patients with sounds similar to those produced by themselves. Moreover, for both ecological and artificial stimuli, the BPM of the soundtracks provided to patients were calculated considering one's own cadence measured before the beginning of the rehabilitation program (at T0, see “experimental design” paragraph), and the cadence of healthy individuals of

the same age (48, 49). In particular, the BPM were calculated following the procedure of Pau and colleagues (13), namely: (a) if a patient's cadence was below the normality, the BPM of the stimulus was set at a value of 10% higher than one's own cadence; (b) if a patient's cadence was below, but close to normality (less than 10% difference), the BPM of the stimulus was set at normality values; (c) if a patient's cadence was above the normality, the BPM of the stimulus was set at a value equal to one's own cadence. The interval between one beat/step and the subsequent one was constant, in both conditions. Prototypical examples of artificial and ecological sounds are illustrated in **Figure 2** and attached in Supplementary Material.

Assessment Protocol

Assessment was carried out when patients were in “ON” state 60–90 min after intake of the usual morning l-DOPA dose. The assessment was carried out at the Laboratory of Biomechanics and Industrial Ergonomics of the Department of Mechanical, Chemical and Materials Engineering, University of Cagliari (Italy), lasted about 90–100 min and included both biomechanical and clinical evaluations.

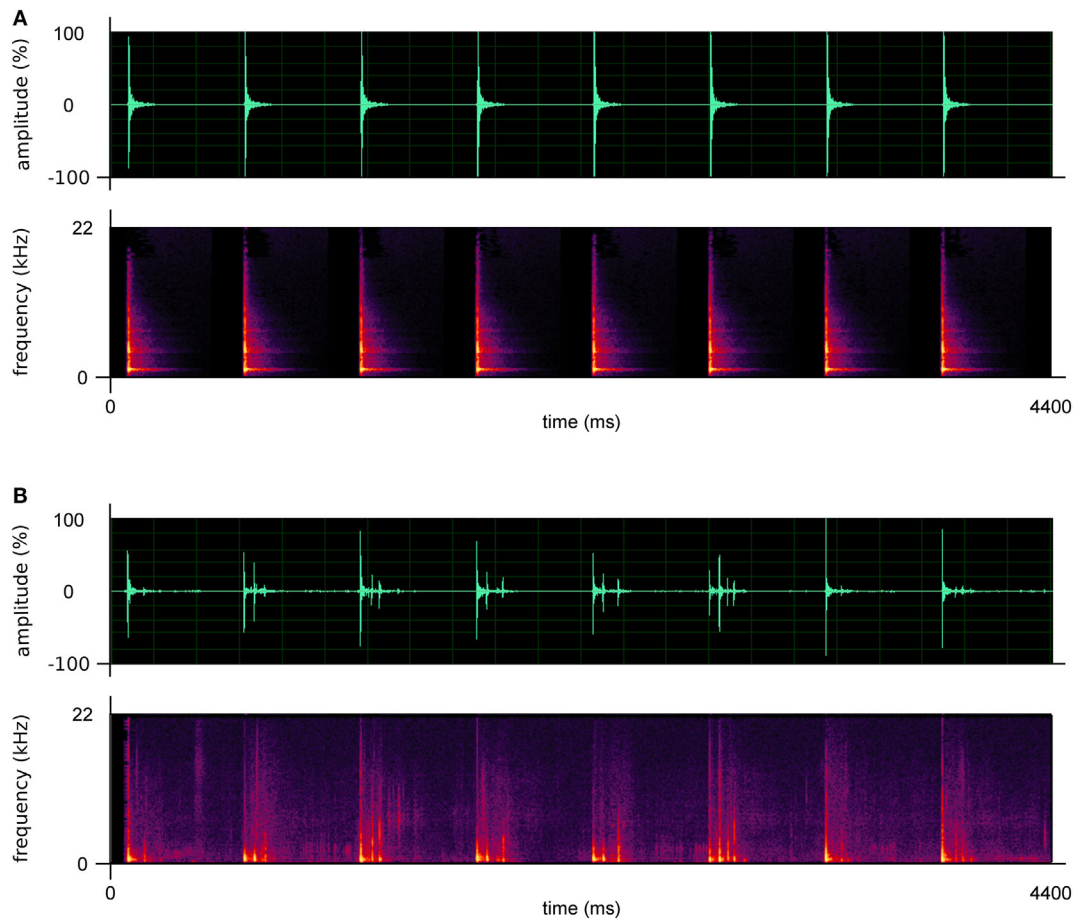


FIGURE 2 | Prototypical examples of the artificial (A) and ecological sounds (B) used in this study. The upper graphs describe the amplitude of the sounds; the lower graphs describe their frequencies.

Biomechanical Evaluation

We measured the spatio-temporal parameters of gait. The acquisition of these parameters was performed using a motion capture system composed of 8 infrared cameras (Smart-D system, BTS Bioengineering, Italy) set at a frequency of 120 Hz. Before the tests, a number of anthropometric features (i.e., height, weight, anterior superior iliac spines distance, pelvis thickness, knee and ankle width, leg length) were collected. Then, 22 reflective passive markers (14 mm diameter) were placed on specific landmarks of individual's lower limbs and trunk according to the protocol described by Davis et al. (50). Participants were asked to walk at a self-selected speed in the most natural manner possible on a 10 m walkway for at least six times, allowing suitable rest were needed in order to avoid fatiguing effects. The raw data were then processed using dedicated software (Smart Analyzer, BTS Bioengineering, Italy) to calculate the following spatio-temporal parameters, which were the primary outcome measures of the present study: gait speed, step length and width, stride length, cadence, stance, swing, and double support phase duration (the latter three parameters expressed as percentage of the gait cycle duration).

Clinical Evaluation

A clinical evaluation of the patients was performed by a team of clinicians' expert in PD. Clinical measures were the secondary outcome measures of the present study. The patients were evaluated by using the following tests:

- Modified H&Y to assess the severity of symptoms of each patient on one of the seven levels of the scale (51, 52).
- Unified Parkinson's disease rating scale (UPDRS) part 3 to specifically evaluate the motor symptoms (53).
- Functional independence measure (FIM) to measure the patients' disability, revealed by their independence, by the amount of care needed and by their cognitive skills (54).
- Tinetti test to evaluate the static and dynamic balance (55).
- Short physical performance battery (SPPB) to evaluate the functionality of the lower limbs. This test is composed of three sections: balance (assessed in three different trials), walking speed (4 m test), and sit to stand (STS) transfer (56).
- Geriatric depression scale (GDS)—short version (15 items), a self-report scale to assess depression in elderly (57).

- Parkinson's disease quality of life questionnaire (PDQ-8), a specific scale to evaluate the quality of life of patients affected by PD (58).
- Falls efficacy scale (FES) to evaluate the patients' fear of falling (59).
- Activities specific balance confidence (ABC), a self-efficacy scale regarding activities which require specific balance confidence (60).
- Freezing of gait questionnaire (FOGQ) to evaluate freezing episodes reported by patients during walking (61).

Rehabilitation Protocol

Participants were engaged in a supervised rehabilitative treatment which lasted 5 weeks; during this period patients were engaged in two sessions per week, whose duration was 45 min each. Patients were individually assisted in the training by a physical therapist, under the supervision of a physical medicine specialist; the treatment sessions were held at "G. Brotzu" General Hospital.

The treatment sessions consisted of standard and personalized exercises aimed at enhancing mobility, balance, and posture. Twenty minutes of each session were dedicated to specific gait training with RAS, with participants engaged in walking while listening to their own personalized soundtrack (either ecological or artificial). Moreover, during the 5 weeks of treatment, participants were invited to train at least three times a week at their homes, performing a subset of the same exercises typically performed at the hospital and 30 min of gait training with RAS (they were provided with an MP3 player). Participants were asked to set the volume at a comfortable level and were allowed to modify it anytime.¹ The rehabilitation protocol was similar to that used by Pau and colleagues (13); a detailed description of the exercises is reported in Appendix.

After the 5 weeks in which participants were engaged in the supervised treatment, they were invited to daily perform their home-exercises for the subsequent 12 weeks. The activities performed by patients during these 12 weeks were unsupervised, however, participants were asked to keep a log of their home-exercises and such data were discussed with clinicians during regularly scheduled meetings.

Experimental Design

Participants were randomly assigned to the groups (ecological or artificial RAS) in a 1:1 fashion, using blocked randomization (62). Randomization was generated by MM, by means of an online sequence generator (www.random.org) inserting 1 and 38 as smallest and largest numbers and calculating random sequences in 2 columns of 19 numbers each. MM also assigned participants to interventions. Participants were evaluated in three moments: before the rehabilitative treatment (T0), at the end of the 5-week rehabilitative treatment (T5), and 3 months after the end of the treatment, namely 17 weeks after the first assessment (T17). Thus, in this study there were two independent variables: (1) RAS, between subjects, two levels (ecological RAS, artificial RAS); (2)

Time, within subjects, three levels (T0, T5, T17). The dependent variables were all the biomechanical and clinical measurements described above. The researcher who assigned participants to the conditions was not involved in the enrollment and evaluation of patients, while the researchers involved in the evaluation of patients (MP, FC, RP, CC, GC) were not aware of the conditions under which the patients were treated (observer-blind trial). Data collection was concluded in April 2016.

Statistical Analyses

As regards the biomechanical variables, a preliminary *t*-test was run to test for possible differences between the left and right limbs (when separate data were available). As no significant difference was revealed by the analyses, we used the mean of the two limbs for each parameter, for each participant, in the subsequent analyses. We conducted a 3×2 mixed MANOVA, using all biomechanical variables as dependent measures. *Post hoc* comparisons were adjusted with LSD test; the alpha level was set at 0.05.

As regards the clinical variables, for each dependent variable, a 3×2 mixed ANOVA (Time \times RAS) was applied. *Post hoc* analyses were calculated by using repeated measures ANOVAs and *t*-tests. The alpha level was set at 0.05 for the omnibus tests and was adjusted with the Bonferroni formula for *post hoc* analyses (p value = $0.05/n$ comparisons).

Moreover, independently of the outcomes of the previous analyses, we planned to conduct a set of additional exploratory analyses, to better examine the potential of the footstep sounds. To this purpose, we separately tested the two groups of participants, by conducting two repeated measures MANOVAs and a set of contrasts adjusted with LSD test on the primary outcome measures of this study (i.e., the biomechanical measures).

All the analyses were performed using the SPSS Statistics software.

RESULTS

A preliminary set of analyses was run to compare the two groups at baseline for each variable, and no significant difference was found. The effects of the two treatments (rehabilitation with ecological vs. artificial RAS) across time (T0, T5, T17) were observed using both biomechanical and clinical measures.

Biomechanical Measures

Overall, the analyses run on biomechanical measures (Table 2) did not reveal any significant result for the interaction Time \times RAS and for the main effect of RAS. Although the results did not reveal a statistical significance for interaction—which was the primary interest of our investigation—the majority of the considered parameters indicated a significant effect of the variable Time [Wilk's $\lambda = 0.098$, $F(16, 15) = 8.652$, $p < 0.001$, $\eta_p^2 = 0.902$]. In particular, this was found for cadence [$F(2, 60) = 4.595$; $p = 0.01$; $\eta_p^2 = 0.133$], gait speed [$F(2, 60) = 8.538$; $p < 0.001$; $\eta_p^2 = 0.222$], step width [$F(2, 60) = 12.647$; $p < 0.001$; $\eta_p^2 = 0.297$], step length [$F(2, 60) = 17.752$; $p < 0.001$; $\eta_p^2 = 0.372$], stride length [$F(2, 60) = 3.681$; $p < 0.05$; $\eta_p^2 = 0.109$], percentage of double support phase [$F(2, 60) = 4.911$; $p = 0.01$; $\eta_p^2 = 0.141$], and percentage of swing phase [$F(2, 60) = 6.843$; $p < 0.005$; $\eta_p^2 = 0.186$].

¹ It is noteworthy that participants were asked to use the auditory stimulation for 30 min almost every day, thus it was important that they feel comfortable with it. Moreover, they needed to adapt the volume during home-training depending on the environmental noise.

TABLE 2 | Comparison between spatio-temporal parameters assessed before and after rehabilitation for each group.

		Spatio-temporal parameters of gait		
		T0	T5	T17
Ecological RAS group	Gait speed (m/s)	1.08 ± 0.24	1.21 ± 0.25	1.24 ± 0.22
	Cadence (steps/min)	115.81 ± 11.71	123.40 ± 9.07	123.49 ± 11.98
	Stride length (m)	1.16 ± 0.20	1.20 ± 0.19	1.21 ± 0.17
	Step length (m)	0.52 ± 0.12	0.58 ± 0.10	0.60 ± 0.09
	Step width (m)	0.18 ± 0.03	0.18 ± 0.04	0.20 ± 0.02
	Stance phase (% of gait cycle)	60.48 ± 3.31	59.59 ± 1.74	59.98 ± 1.87
	Swing phase (% of gait cycle)	38.68 ± 2.59	40.42 ± 1.74	39.98 ± 1.90
	Double support (% of gait cycle)	11.34 ± 2.44	10.07 ± 2.44	9.96 ± 2.02
Artificial RAS group	Gait speed (m/s)	1.05 ± 0.29	1.11 ± 0.24	1.15 ± 0.29
	Cadence (steps/min)	113.43 ± 14.54	115.53 ± 11.34	116.50 ± 11.62
	Stride length (m)	1.13 ± 0.20	1.16 ± 0.18	1.18 ± 0.21
	Step length (m)	0.52 ± 0.12	0.56 ± 0.10	0.59 ± 0.11
	Step width (m)	0.17 ± 0.03	0.17 ± 0.03	0.20 ± 0.02
	Stance phase (% of gait cycle)	60.79 ± 2.92	60.32 ± 2.36	60.12 ± 2.20
	Swing phase (% of gait cycle)	38.89 ± 2.57	39.63 ± 2.33	39.43 ± 2.62
	Double support (% of gait cycle)	11.73 ± 2.78	10.69 ± 2.42	10.70 ± 2.55

Values are expressed as mean ± SD.

Conversely, the effect of time on the percentage of stance phase was not significant.

Then, we investigated more deeply how these parameters changed across time. We found that the cadence significantly increased between T0 and T5 ($p = 0.021$) and between T0 and T17 ($p = 0.029$), while it remained constant between T5 and T17. Analogous results were revealed by the analysis on gait speed, indeed it significantly increased between T0 and T5 ($p = 0.006$) and between T0 and T17 ($p = 0.001$) and remained constant between T5 and T17. Similarly, the percentage of swing phase significantly increased between T0 and T5 ($p = 0.001$) and between T0 and T17 ($p = 0.033$), while it remained constant between T5 and T17. Consistently, the percentage of double support phase significantly decreased between T0 and T5 ($p = 0.009$) and between T0 and T17 ($p = 0.013$) and remained constant between T5 and T17. As concerns step length, we found an increase between T0 and T5 ($p = 0.004$), between T0 and T17 ($p < 0.001$), and also between T5 and T17 ($p = 0.01$). Surprisingly, we did not find a difference between T0 and T5 regarding step width ($p = 0.568$), but the values observed in T17 were significantly higher than those observed in T0 ($p < 0.001$) and T5 ($p < 0.001$). Moreover, as concerns stride length, there was no difference between T0 and T5 ($p = 0.108$), while it significantly increased between T0 and T17 ($p = 0.011$); no difference was found between T5 and T17.

Clinical Measures

The analyses run on clinical measures (Table 3) did not reveal any significant result for the time × RAS interaction and for the main effect of RAS. As regards time, similarly to the analyses on biomechanical measures, we found that the main effect was significant for almost all clinical measures. In particular, a significant main effect of time was observed for the UPDRS—part 3 [$F(2, 52) = 32.749$; $p < 0.001$; $\eta_p^2 = 0.557$], ABC [$F(2, 56) = 5.418$; $p < 0.01$; $\eta_p^2 = 0.162$], FES [$F(2, 58) = 4.819$; $p < 0.05$; $\eta_p^2 = 0.143$], FOGQ [$F(2, 60) = 3.926$; $p < 0.05$; $\eta_p^2 = 0.116$], GDS [$F(2, 58) = 3.663$; $p < 0.05$; $\eta_p^2 = 0.112$], PDQ8 [$F(2, 58) = 7.343$;

$p < 0.001$; $\eta_p^2 = 0.202$], Tinetti test [$F(2, 58) = 3.945$; $p < 0.05$; $\eta_p^2 = 0.12$], SPPB [$F(2, 58) = 5.330$; $p < 0.01$; $\eta_p^2 = 0.155$] and its subcomponents STS [$F(2, 42) = 15.390$; $p < 0.001$; $\eta_p^2 = 0.423$], and 4-m test [$F(2, 40) = 7.382$; $p < 0.01$; $\eta_p^2 = 0.270$]. Conversely, no significant effect was found for the FIM scale.

On the basis of the ANOVAs results, we decided to further explore how the clinical measures changed across time. We found that the FES scores decreased between T0 and T5 [$t(30) = 2.375$; $p < 0.05$; $d = 0.377$] and between T0 and T17 [$t(30) = 2.367$; $p < 0.05$; $d = 0.406$], while they remained constant between T5 and T17. Analogous results were observed for SPPB, with an increase between T0 and T5 [$t(30) = 2.456$; $p = 0.01$; $d = 0.304$] and between T0 and T17 [$t(30) = 2.367$; $p < 0.05$; $d = 0.253$], and no difference between T5 and T17. This trend was also observed in two of the subcomponents of SPPB, namely STS and 4-m test. There was a decrease of time necessary to complete the test between T0 and T5 for both the STS [$t(22) = 4.082$; $p < 0.001$; $d = 1.016$] and the 4-m test [$t(22) = 2.331$; $p < 0.05$; $d = 0.647$], and between T0 and T17 for both the STS [$t(22) = 4.651$; $p < 0.001$; $d = 1.061$] and the 4-m test [$t(22) = 2.666$; $p < 0.01$; $d = 0.569$]; in no case a difference between T5 and T17 was observed.

The FOGQ scores also decreased between T0 and T5 [$t(31) = 2.459$; $p = 0.01$; $d = 0.217$], while the difference between T0 and T17 was no longer significant after the Bonferroni correction [$t(31) = 1.797$; $p < 0.05$; $d = 0.173$]; no difference was observed between T5 and T17. A similar trend was observed for the Tinetti test, with marginally significant increases between T0 and T5 [$t(30) = 2.129$; $p < 0.05$; $d = 0.376$] and between T0 and T17 [$t(30) = 2.037$; $p < 0.05$; $d = 0.438$], which were no longer significant after the Bonferroni adjustment; again no difference was observed between T5 and T17.

The UPDRS—part 3 scores decreased between T0 and T5 [$t(27) = 7.701$; $p < 0.001$; $d = 0.548$], between T0 and T17 [$t(27) = 6.261$; $p < 0.001$; $d = 0.781$], and also between T5 and T17 [$t(31) = 2.598$; $p < 0.01$; $d = 0.269$]. Analogously, the ABC scores

TABLE 3 | Comparison between clinical scores assessed before and after rehabilitation for each group.

		Clinical scores		
		T0	T5	T17
Ecological rhythmic auditory stimulation (RAS) group	UPDRS—part 3	16.93 ± 9.14	11.64 ± 7.00	10.54 ± 6.96
	Functional independence measure (FIM)	123.47 ± 3.91	123.87 ± 3.27	123.87 ± 3.25
	Tinetti test	27.19 ± 1.76	27.75 ± 0.45	27.62 ± 0.62
	Short physical performance battery (SPPB)	11.33 ± 1.54	11.87 ± 0.35	11.87 ± 0.35
	SPPB—4 m test (s)	2.56 ± 0.28	2.25 ± 0.22	2.42 ± 0.22
	SPPB—sit to stand (STS) (s)	9.83 ± 1.80	8.47 ± 1.20	8.17 ± 1.34
	Geriatric depression scale (GDS)	3.33 ± 2.26	3.33 ± 2.53	2.47 ± 2.20
	Parkinson's disease quality of life questionnaire (PDQ)-8	20.21 ± 13.46	20.00 ± 11.92	14.17 ± 11.56
	Falls efficacy scale (FES)	4.73 ± 6.95	2.40 ± 3.22	2.27 ± 2.34
	Activities specific balance confidence (ABC)	81.96 ± 13.21	85.23 ± 7.49	88.38 ± 8.75
Artificial RAS group	Freezing of gait questionnaire (FOGQ)	4.00 ± 5.14	2.44 ± 3.76	2.62 ± 4.08
	UPDRS—part 3	20.21 ± 9.59	15.36 ± 9.20	12.79 ± 6.78
	FIM	117.94 ± 13.12	118.50 ± 10.91	119.00 ± 10.48
	Tinetti test	26.93 ± 1.58	27.47 ± 0.74	27.67 ± 0.49
	SPPB	10.94 ± 2.29	11.75 ± 0.77	11.69 ± 0.70
	SPPB—4 m test (s)	2.72 ± 0.70	2.30 ± 0.36	2.34 ± 0.44
	SPPB—STS (s)	10.00 ± 1.52	8.43 ± 1.21	8.47 ± 1.42
	GDS	3.56 ± 2.85	2.75 ± 2.27	2.19 ± 2.23
	PDQ-8	17.97 ± 10.86	16.21 ± 11.87	12.69 ± 10.67
	FES	5.50 ± 5.49	3.62 ± 3.01	3.56 ± 3.60
	ABC	79.37 ± 14.98	83.81 ± 8.73	87.12 ± 8.37
	FOGQ	3.00 ± 4.27	2.56 ± 3.72	2.81 ± 4.00

Values are expressed as mean ± SD.

increased between T0 and T5 [$t(29) = 1.997$; $p < 0.05$; $d = 0.302$], between T0 and T17 [$t(29) = 2.556$; $p < 0.01$; $d = 0.609$] and between T5 and T17 [$t(30) = 2.240$; $p < 0.05$; $d = 0.385$], however, the first and the last comparisons were no longer significant after Bonferroni correction.

The PDQ8 revealed the best results at T17, with higher scores compared to both T0 [$t(30) = 2.950$; $p < 0.01$; $d = 0.490$] and T5 [$t(31) = 2.966$; $p < 0.01$; $d = 0.285$], while no difference was observed between T0 and T5. As concerns the GDS, a marginal decrease was observed between T0 and T17 [$t(30) = 2.171$; $p < 0.05$; $d = 0.476$], which disappeared after the Bonferroni correction, while the other comparisons did not reach any significant value.

Additional Exploratory Analyses

We conducted a set of exploratory analyses by separately examining the two groups of participants on the primary outcome measures (i.e., biomechanical measures). The repeated measures MANOVAs revealed significant results for the ecological RAS group [Wilk's $\lambda = 0.23$, $F(16, 46) = 3.149$, $p = 0.001$, $\eta_p^2 = 0.523$], but not for the artificial RAS group. The MANOVA conducted on the ecological RAS group data revealed significant values for the following measures: cadence [$F(2, 30) = 4.367$; $p < 0.05$; $\eta_p^2 = 0.225$], gait speed [$F(2, 30) = 5.914$; $p < 0.01$; $\eta_p^2 = 0.283$], percentage of swing phase [$F(2, 30) = 5.533$; $p < 0.01$; $\eta_p^2 = 0.269$], step length [$F(2, 30) = 10.23$; $p < 0.001$; $\eta_p^2 = 0.405$], and step width [$F(2, 30) = 5.322$; $p < 0.05$; $\eta_p^2 = 0.262$]. In our opinion, of particular interest are the results concerning cadence and gait speed (see **Figure 3**), since these are the two measures more directly related to the auditory stimuli. The contrasts for these

variables indicate that cadence significantly increased from T0 to T5 ($p = 0.011$), but this advantage was not maintained at T17 ($p = 0.072$), while gait speed increased from T0 to T5 ($p = 0.017$) and this advantage was maintained at T17 ($p = 0.13$).

DISCUSSION

The aim of this study was to investigate whether a PD rehabilitation program integrated with ecological RAS (i.e., footstep sounds) can be more effective than the same program integrated with artificial RAS (i.e., metronome sounds). We hypothesized that both groups would benefit from the rhythmic information of the stimuli, but the group exposed to ecological RAS would also take advantage of the priming effect elicited by the ecological information of footstep sounds. The results observed on the primary outcome measures (biomechanical measures) suggest that the treatments are equally effective as no significant interaction was revealed by analyses. The same results were given by the secondary outcome measures (clinical measures).

As a whole, our results indicate that—independently of the type of sound—the rehabilitation programs integrated with RAS are effective. Indeed, comparing the biomechanical and clinical data collected before and after the treatment, we observed noticeable improvements in the majority of the variables. This evidence is in line with previous literature, which clearly proved the efficacy of rehabilitation with RAS, in terms of both spatio-temporal parameters and clinical variables (2, 9–11, 17). Moreover, we observed that these improvements were largely maintained at the follow-up, 3 months after the end of the supervised period, which represents a longer term

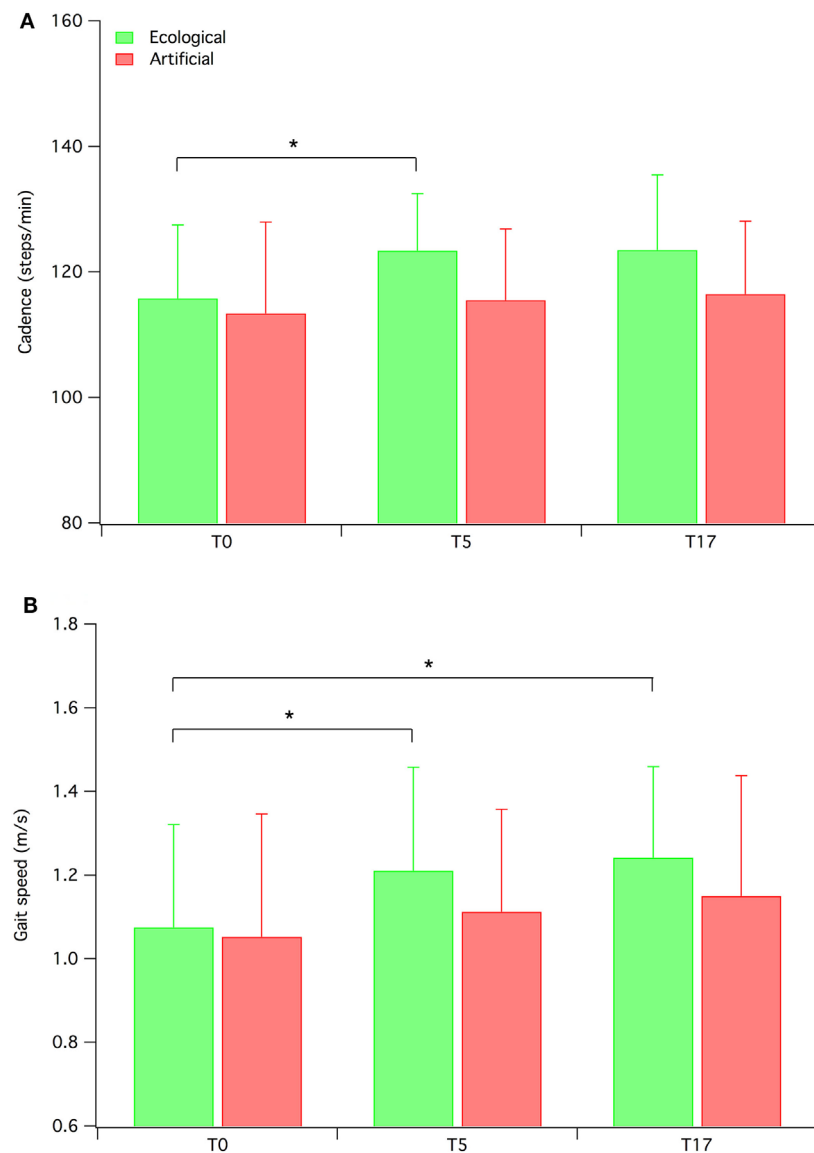


FIGURE 3 | Mean cadence (A) and gait speed (B) for both ecological rhythmic auditory stimulation (RAS) participants and artificial RAS participants. Error bars represent SD.

compared to the majority of previous RAS studies. However, given that the RAS efficacy is well-documented in literature, our main aim was not to further confirm it, but to examine whether the type of sound (i.e., ecological or artificial) can actually influence the efficacy of rehabilitation.

The novelty of this study is that, for the first time, we used biological motion sounds as RAS in a rehabilitation program with PD patients. The same rehabilitation program was integrated either with footstep sounds (ecological RAS condition) or with metronome sounds (artificial RAS condition) and the overall results indicated that the effects of the two sounds—in a rehabilitation context—are equivalent. However, to better examine the potential of the footstep sounds and obtain additional information that could be used as starting point for future investigations, we ran a set of exploratory analyses on biomechanical measures

for each group, separately. These analyses revealed that only the patients assigned to the ecological RAS condition significantly improved. Among the various measures showing significant improvements, in our opinion a particular attention should be dedicated to cadence and gait speed, since these are the two parameters more directly linked to the auditory stimuli. However, we acknowledge that our exploratory analyses are only informative but not conclusive; indeed these analyses do not compare the two groups and cannot prove a superior effect of ecological sounds over artificial sounds. Future studies should better clarify the potential of footstep sounds in the rehabilitation context.

The effects of ecological and artificial sounds on motor tasks in PD patients were examined in previous research (37). In their study, Young and colleagues found that in some spatio-temporal parameters patients performed better in the ecological sound

condition than in the artificial sound condition. However, the methods used in the previous and in this study are quite different: the patients tested by Young and colleagues were engaged in a real-time imitation task, while in our case the patients were engaged in a 5-week rehabilitation program. To the best of our knowledge this is the first time that a rehabilitation program is integrated with ecological RAS, since previous RAS experiments generally used music or metronome sounds (4–6). Therefore, our study can be considered as a first attempt to investigate in this direction and further research is needed to understand whether this line of investigation can be fruitful.

Our hypothesis, based on the audiovisual mirror neurons (41–44) and on the common coding perceptual-motor theories (45, 46), was that the use of ecological RAS would constitute an advantage. In particular, we hypothesized that footstep sounds would evoke a mental representation of walking, directly activating the motor systems and, consequently, facilitating patients' walking. The overall results we observed are not consistent with this hypothesis. Only the additional exploratory analyses are in line with this hypothesis, however, further research stressing the role of RAS is necessary to confirm our exploratory observations. Indeed, we examined the effects of ecological RAS integrated with a standard rehabilitation protocol performed at a hospital. On the one hand, this has a strong ecological validity, since it represents a typical context in which RAS can be employed; on the other hand, the effects of RAS can be somehow masked by the effects of the other exercises performed by the patients. Thus, in our opinion, future studies should isolate the effects of gait training with RAS from the effects of other exercises, making more salient the possible differences between ecological and artificial RAS.

Like every study, also our work does have some limitations. The severity of the disease of our participants was mild–moderate, so we cannot extend our results to patients with more severe impairments. Moreover, by experimentally testing a sample of individuals with a higher level of motor impairment it could be possible to observe stronger effects. This might be helpful to better understand the possible different effects between the two types of sound. Another limitation is that we did not control/manipulate important parameters of sounds (i.e., volume, frequency). In our opinion, specific research of how these features of the stimuli affect gait training should be performed in more controlled laboratory experiments, rather than in ecological settings as in our study. Moreover, the subjective pleasantness of the two types of sound (ecological vs. artificial) should be investigated.

From an applied perspective, our results suggest that ecological RAS is equally effective compared to artificial RAS, therefore, it would be possible to let patients decide what type of sound they prefer for gait training. This is particularly important for their

compliance with the treatment, because the administration of sounds which are perceived as annoying by patients might lead to a low adherence to the training. Moreover, it is noteworthy that several participants reported that the footstep sounds were meaningful and reminded them some physical activities they used to do in the past (e.g., military march), thus evoking motor images directly linked with walking. For future research, the implementation of ecological RAS training in the standard rehabilitation protocols at the hospitals could represent an important source of data, allowing researchers to examine the efficacy of ecological RAS on larger samples of participants.

ETHICS STATEMENT

This study was carried out in accordance with the recommendations of the Independent Ethics Committee of the A.O.U. Cagliari with written informed consent from all subjects. All subjects gave written informed consent in accordance with the Declaration of Helsinki. The protocol was approved by the Independent Ethics Committee of the A.O.U. Cagliari.

AUTHOR CONTRIBUTIONS

MM, RP, FC, FS, TA, PB, CC, GC, MG, and MP designed the study. CC and RP performed the physical capacity assessment and prepared the rehabilitation protocol. GC performed the neurological evaluations. MP and FC collected and processed the biomechanical data. MM, TA, FS, and PB prepared the stimuli. MM and MG performed the statistical analyses. MM, FS, and MP wrote the manuscript. RP, FC, TA, PB, CC, GC, and MG revised the manuscript.

ACKNOWLEDGMENTS

This study was partly supported by the Autonomous Region of Sardinia (grant CRP-78543 L.R. 7/2007). The author MM was supported by the Autonomous Region of Sardinia, Master and Back Programme 2013 (PRR-MAB-A2013-19330). The authors wish to express their gratitude to the physical therapists Marilena Fara, Giovanna Ghiani, Alessandra Pani, Elsa Sau, Gino Sedda, and Mauro Usala for their valuable support during the training program.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at <https://www.frontiersin.org/articles/10.3389/fneur.2018.00348/full#supplementary-material>.

REFERENCES

- Blin O, Ferrandez AM, Serratrice G. Quantitative analysis of gait in Parkinson patients: increased variability of stride length. *J Neurol Sci* (1990) 98:91–7. doi:10.1016/0022-510X(90)90184-O
- Thaut MH, McIntosh GC, Rice RR, Miller RA, Rathbun J, Brault JM. Rhythmic auditory stimulation in gait training for Parkinson's disease patients. *Mov Disord* (1996) 11:193–200. doi:10.1002/mds.870110213
- Lim I, van Wegen E, de Goede C, Deutekom M, Nieuwboer A, Willems A, et al. Effects of external rhythmical cueing on gait in patients with Parkinson's disease: a systematic review. *Clin Rehabil* (2005) 19:695–713. doi:10.1191/0269215505cr9060a
- Murgia M, Corona F, Pili R, Sors F, Agostini T, Casula C, et al. Rhythmic auditory stimulation (RAS) and motor rehabilitation in Parkinson's disease: new frontiers in assessment and intervention protocols. *Open Psychol J* (2015) 8:220–9. doi:10.2174/1874350101508010220

5. Nombela C, Hughes LE, Owen AM, Grahn JA. Into the groove: can rhythm influence Parkinson's disease? *Neurosci Biobehav Rev* (2013) 37:2564–70. doi:10.1016/j.neubiorev.2013.08.003
6. Thaut MH, Abiru M. Rhythmic auditory stimulation in rehabilitation of movement disorders: a review of current research. *Music Percept Interdiscip J* (2010) 27:263–9. doi:10.1525/mp.2010.27.4.263
7. Fernandez del Olmo M, Cudeiro J. Temporal variability of gait in Parkinson disease: effects of a rehabilitation programme based on rhythmic sound cues. *Parkinsonism Relat Disord* (2005) 11:25–33. doi:10.1016/j.parkreldis.2004.09.002
8. Fernandez del Olmo M, Arias P, Furio MC, Pozo MA, Cudeiro J. Evaluation of the effect of training using auditory stimulation on rhythmic movement in Parkinsonian patients – a combined motor and [18F]-FDG PET study. *Parkinsonism Relat Disord* (2006) 12:155–64. doi:10.1016/j.parkreldis.2005.11.002
9. Ford MP, Malone LA, Nyikos I, Yelissety R, Bickel CS. Gait training with progressive external auditory cueing in persons with Parkinson's disease. *Arch Phys Med Rehabil* (2010) 91:1255–61. doi:10.1016/j.apmr.2010.04.012
10. Frazzitta G, Maestri R, Uccellini D, Bertotti G, Abelli P. Rehabilitation treatment of gait in patients with Parkinson's disease with freezing: a comparison between two physical therapy protocols using visual and auditory cues with or without treadmill training. *Mov Disord* (2009) 24:1139–43. doi:10.1002/mds.22491
11. Kadivar Z, Corcos DM, Foto J, Hondzinski JM. Effect of step training and rhythmic auditory stimulation on functional performance in Parkinson patients. *Neurorehabil Neural Repair* (2011) 25:626–35. doi:10.1177/1545968311401627
12. Miller RA, Thaut MH, McIntosh GC, Rice RR. Components of EMG symmetry and variability in parkinsonian and healthy elderly gait. *Electroencephalogr Clin Neurophysiol* (1996) 101:1–7. doi:10.1016/0013-4694(95)00209-X
13. Pau M, Corona F, Pili R, Casula C, Sors F, Agostini T, et al. Effects of physical rehabilitation integrated with rhythmic auditory stimulation on spatio-temporal and kinematic parameters of gait in Parkinson's disease. *Front Neurol* (2016) 7:126. doi:10.3389/fneur.2016.00126
14. Avanzino L, Pelosin E, Vicario CM, Lagravinese G, Abbruzzese G, Martino D. Time processing and motor control in movement disorders. *Front Hum Neurosci* (2016) 10:631. doi:10.3389/fnhum.2016.00631
15. Bieńkiewicz MMN, Craig CM. Parkinson's is time on your side? Evidence for difficulties with sensorimotor synchronization. *Front Neurol* (2015) 6:249. doi:10.3389/fneur.2015.00249
16. Dalla Bella S, Benoit C-E, Farrugia N, Keller PE, Obrig H, Mainka S, et al. Gait improvement via rhythmic stimulation in Parkinson's disease is linked to rhythmic skills. *Sci Rep* (2017) 7:42005. doi:10.1038/srep42005
17. McIntosh GC, Brown SH, Rice RR, Thaut MH. Rhythmic auditory-motor facilitation of gait patterns in patients with Parkinson's disease. *J Neurol Neurosurg Psychiatry* (1997) 62:22–6. doi:10.1136/jnnp.62.1.22
18. Arias P, Cudeiro J. Effect of rhythmic auditory stimulation on gait in parkinsonian patients with and without freezing of gait. *PLoS One* (2010) 5:e9675. doi:10.1371/journal.pone.0009675
19. Baker K, Rochester L, Nieuwboer A. The immediate effect of attentional, auditory, and a combined cue strategy on gait during single and dual tasks in Parkinson's disease. *Arch Phys Med Rehabil* (2007) 88:1593–600. doi:10.1016/j.apmr.2007.07.026
20. Baker K, Rochester L, Nieuwboer A. The effect of cues on gait variability – reducing the attentional cost of walking in people with Parkinson's disease. *Parkinsonism Relat Disord* (2008) 14:314–20. doi:10.1016/j.parkreldis.2007.09.008
21. Fernandez del Olmo M, Cudeiro J. A simple procedure using auditory stimuli to improve movement in Parkinson's disease: a pilot study. *Neurol Clin Neurophysiol NCN* (2003) 2003:1–7.
22. Freedland RL, Festa C, Sealy M, McBean A, Elghazaly P, Capan A, et al. The effects of pulsed auditory stimulation on various gait measurements in persons with Parkinson's disease. *NeuroRehabilitation* (2002) 17:81–7.
23. Hove MJ, Suzuki K, Uchitomi H, Orimo S, Miyake Y. Interactive rhythmic auditory stimulation reinstates natural 1/f timing in gait of Parkinson's patients. *PLoS One* (2012) 7:e32600. doi:10.1371/journal.pone.0032600
24. Howe TE, Lövgren B, Cody FWJ, Ashton VJ, Oldham JA. Auditory cues can modify the gait of persons with early-stage Parkinson's disease: a method for enhancing parkinsonian walking performance? *Clin Rehabil* (2003) 17:363–7. doi:10.1191/0269215503cr6210a
25. Nieuwboer A, Baker K, Willems A-M, Jones D, Spildooren J, Lim I, et al. The short-term effects of different cueing modalities on turn speed in people with Parkinson's disease. *Neurorehabil Neural Repair* (2009) 23:831–6. doi:10.1177/1545968309337136
26. Rochester L, Burn DJ, Woods G, Godwin J, Nieuwboer A. Does auditory rhythmic cueing improve gait in people with Parkinson's disease and cognitive impairment? A feasibility study. *Mov Disord* (2009) 24:839–45. doi:10.1002/mds.22400
27. Suteerawattananon M, Morris GS, Etnyre BR, Jankovic J, Protas EJ. Effects of visual and auditory cues on gait in individuals with Parkinson's disease. *J Neurol Sci* (2004) 219:63–9. doi:10.1016/j.jns.2003.12.007
28. Willems AM, Nieuwboer A, Chavret F, Desloovere K, Dom R, Rochester L, et al. The use of rhythmic auditory cues to influence gait in patients with Parkinson's disease, the differential effect for freezers and non-freezers, an explorative study. *Disabil Rehabil* (2006) 28:721–8. doi:10.1080/09638280500386569
29. Willems A-M, Nieuwboer A, Chavret F, Desloovere K, Dom R, Rochester L, et al. Turning in Parkinson's disease patients and controls: the effect of auditory cues. *Mov Disord* (2007) 22:1871–8. doi:10.1002/mds.21445
30. Marchese R, Diverio M, Zucchi F, Lentino C, Abbruzzese G. The role of sensory cues in the rehabilitation of parkinsonian patients: a comparison of two physical therapy protocols. *Mov Disord* (2000) 15:879–83. doi:10.1002/1531-8257(200009)15:5<879::AID-MDS1018>3.0.CO;2-9
31. Rochester L, Nieuwboer A, Baker K, Hetherington V, Willems A-M, Chavret F, et al. The attentional cost of external rhythmic cues and their impact on gait in Parkinson's disease: effect of cue modality and task complexity. *J Neural Transm (Vienna)* (2007) 114:1243–8. doi:10.1007/s00702-007-0756-y
32. Ellis T, de Goede CJ, Feldman RG, Wolters EC, Kwakkel G, Wagenaar RC. Efficacy of a physical therapy program in patients with Parkinson's disease: a randomized controlled trial. *Arch Phys Med Rehabil* (2005) 86:626–32. doi:10.1016/j.apmr.2004.08.008
33. Corona F, Pau M, Guicciardi M, Murgia M, Pili R, Casula C. Quantitative assessment of gait in elderly people affected by Parkinson's disease. *IEEE International Symposium on Medical Measurements and Applications (MeMeA)*. Benevento (2016). p. 1–6.
34. Murgia M, Santoro I, Tamburini G, Prpic V, Sors F, Galmonte A, et al. Ecological sounds affect breath duration more than artificial sounds. *Psychol Res* (2016) 80:76–81. doi:10.1007/s00426-015-0647-z
35. Effenberg AO, Fehse U, Schmitz G, Krueger B, Mechling H. Movement sonification: effects on motor learning beyond rhythmic adjustments. *Front Neurosci* (2016) 10:219. doi:10.3389/fnins.2016.00219
36. Rodger MWM, Young WR, Craig CM. Synthesis of walking sounds for alleviating gait disturbances in Parkinson's disease. *IEEE Trans Neural Syst Rehabil Eng* (2014) 22:543–8. doi:10.1109/TNSRE.2013.2285410
37. Young WR, Rodger MWM, Craig CM. Auditory observation of stepping actions can cue both spatial and temporal components of gait in Parkinson's disease patients. *Neuropsychologia* (2014) 57:140–53. doi:10.1016/j.neuropsychologia.2014.03.009
38. Rodger MWM, Craig CM. Beyond the metronome: auditory events and music may afford more than just interval durations as gait cues in Parkinson's disease. *Front Neurosci* (2016) 10:272. doi:10.3389/fnins.2016.00272
39. Rizzolatti G, Fadiga L, Gallese V, Fogassi L. Premotor cortex and the recognition of motor actions. *Brain Res Cogn Brain Res* (1996) 3:131–41. doi:10.1016/0926-6410(95)00038-0
40. Cook R, Bird G, Catmur C, Press C, Heyes C. Mirror neurons: from origin to function. *Behav Brain Sci* (2014) 37:177–92. doi:10.1017/S0140525X13000903
41. Galati G, Committeri G, Spitoni G, Aprile T, Di Russo F, Pitzalis S, et al. A selective representation of the meaning of actions in the auditory mirror system. *Neuroimage* (2008) 40:1274–86. doi:10.1016/j.neuroimage.2007.12.044
42. Gazzola V, Aziz-Zadeh L, Keysers C. Empathy and the somatotopic auditory mirror system in humans. *Curr Biol* (2006) 16:1824–9. doi:10.1016/j.cub.2006.07.072
43. Lewis JW, Brefczynski JA, Phinney RE, Janik JJ, DeYoe EA. Distinct cortical pathways for processing tool versus animal sounds. *J Neurosci* (2005) 25:5148–58. doi:10.1523/JNEUROSCI.0419-05.2005
44. Tettamanti M, Buccino G, Saccuman MC, Gallese V, Danna M, Scifo P, et al. Listening to action-related sentences activates fronto-parietal motor circuits. *J Cogn Neurosci* (2005) 17:273–81. doi:10.1162/0898929053124965

45. Hommel B, Müsseler J, Aschersleben G, Prinz W. The theory of event coding (TEC): a framework for perception and action planning. *Behav Brain Sci* (2001) 24:849–878; discussion 878–937. doi:10.1017/S0140525X01000103
46. Prinz W. Perception and action planning. *Eur J Cogn Psychol* (1997) 9:129–54. doi:10.1080/713752551
47. Gibb WR, Lees AJ. The relevance of the Lewy body to the pathogenesis of idiopathic Parkinson's disease. *J Neurol Neurosurg Psychiatry* (1988) 51:745–52. doi:10.1136/jnnp.51.6.745
48. Hollman JH, McDade EM, Petersen RC. Normative spatiotemporal gait parameters in older adults. *Gait Posture* (2011) 34:111–8. doi:10.1016/j.gaitpost.2011.03.024
49. Oberg T, Karsznia A, Oberg K. Basic gait parameters: reference data for normal subjects, 10–79 years of age. *J Rehabil Res Dev* (1993) 30:210–23.
50. Davis RB, Öunpuu S, Tyburski D, Gage JR. A gait analysis data collection and reduction technique. *Hum Mov Sci* (1991) 10:575–87. doi:10.1016/0167-9457(91)90046-Z
51. Hoehn MM, Yahr MD. Parkinsonism: onset, progression and mortality. *Neurology* (1967) 17:427–42. doi:10.1212/WNL.17.5.427
52. Goetz CG, Poewe W, Rascol O, Sampaio C, Stebbins GT, Counsell C, et al. Movement disorder society task force report on the Hoehn and Yahr staging scale: status and recommendations. *Mov Disord* (2004) 19:1020–8. doi:10.1002/mds.20213
53. Goetz CG, Fahn S, Martinez-Martin P, Poewe W, Sampaio C, Stebbins GT, et al. Movement disorder society-sponsored revision of the unified Parkinson's disease rating scale (MDS-UPDRS): process, format, and clinimetric testing plan. *Mov Disord* (2007) 22:41–7. doi:10.1002/mds.21198
54. Granger C, Hamilton B, Keith RA, Zielezny MA, Sherwin F, Keith RA, et al. Advances in functional assessment for medical rehabilitation. *Top Geriatr Rehabil* (1986) 1(3):59–74. doi:10.1097/00013614-198604000-00007
55. Tinetti ME, Williams TF, Mayewski R. Fall risk index for elderly patients based on number of chronic disabilities. *Am J Med* (1986) 80:429–34. doi:10.1016/0002-9343(86)90717-5
56. Guralnik JM, Simonsick EM, Ferrucci L, Glynn RJ, Berkman LF, Blazer DG, et al. A short physical performance battery assessing lower extremity function: association with self-reported disability and prediction of mortality and nursing home admission. *J Gerontol* (1994) 49:M85–94. doi:10.1093/geronj/49.2.M85
57. Marc LG, Raue PJ, Bruce ML. Screening performance of the geriatric depression scale (GDS-15) in a diverse elderly home care population. *Am J Geriatr Psychiatry* (2008) 16:914–21. doi:10.1097/JGP.0b013e318186bd67
58. Franchignoni F, Giordano A, Ferriero G. Rasch analysis of the short form 8-item Parkinson's disease questionnaire (PDQ-8). *Qual Life Res* (2008) 17:541–8. doi:10.1007/s11136-008-9341-6
59. Tinetti ME, Richman D, Powell L. Falls efficacy as a measure of fear of falling. *J Gerontol* (1990) 45:239–43. doi:10.1093/geronj/45.6.P239
60. Powell LE, Myers AM. The activities-specific balance confidence (ABC) scale. *J Gerontol A Biol Sci Med Sci* (1995) 50A:M28–34. doi:10.1093/gerona/50A.1.M28
61. Giladi N, Shabtai H, Simon ES, Biran S, Tal J, Korczyn AD. Construction of freezing of gait questionnaire for patients with parkinsonism. *Parkinsonism Relat Disord* (2000) 6:165–70. doi:10.1016/S1353-8020(99)00062-0
62. Schulz KF, Grimes DA. Generation of allocation sequences in randomised trials: chance, not choice. *Lancet* (2002) 359:515–9. doi:10.1016/S0140-6736(02)07683-3

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.

Copyright © 2018 Murgia, Pili, Corona, Sors, Agostini, Bernardis, Casula, Cossu, Guicciardi and Pau. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

APPENDIX: REHABILITATION PROTOCOL

Targets	Exercises
<ul style="list-style-type: none">• Prevention of inactivity and fear of falling• Prevention of falls• Improving physical activity levels• Recognizing the onset of fluctuations and adopting suitable movement strategies• Learning simple motor exercises of increasing difficulty to be self-administered at home	<ul style="list-style-type: none">• Segmental exercises of active or assisted mobilization (flexion–extension, prono-supination) to increase strength, mobility, and coordination of four limbs• Stretching of anterior and posterior muscular kinetic chains• Improvement of static balance: standing (uni- and bipedal), sitting, and quadrupedal posture• Improvement of dynamic balance: ambulation on paths of increasing levels of difficulty (e.g., turns, obstacles, etc.)• Postural changes: from sitting/quadrupedal to standing, from supine/prone to lateral• Occupational therapy exercises• Gait training with rhythmic auditory stimulation (for about 50% of the duration of each session)

Advantages of publishing in Frontiers



OPEN ACCESS

Articles are free to read
for greatest visibility
and readership



FAST PUBLICATION

Around 90 days
from submission
to decision



HIGH QUALITY PEER-REVIEW

Rigorous, collaborative,
and constructive
peer-review



TRANSPARENT PEER-REVIEW

Editors and reviewers
acknowledged by name
on published articles

Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org

Contact us: info@frontiersin.org | +41 21 510 17 00



REPRODUCIBILITY OF RESEARCH

Support open data
and methods to enhance
research reproducibility



DIGITAL PUBLISHING

Articles designed
for optimal readership
across devices



FOLLOW US

@frontiersin



IMPACT METRICS

Advanced article metrics
track visibility across
digital media



EXTENSIVE PROMOTION

Marketing
and promotion
of impactful research



LOOP RESEARCH NETWORK

Our network
increases your
article's readership