Published by the Open Science Centre, University of Jyväskylä, Finland

ISSN: 1795-6889

HUMAN TECHNOLOGY

Volume 16, Number 3, November 2020

SPECIAL ISSUE

Mind, Music and Technology

Marc R. Thompson and Jonna K. Vuoskoski
Guest Editors

Jukka Jouhki
Editor in Chief

www.humantechnology.jyu.fi

HUMAN TECHNOLOGY

Volume #, Number #, Month Year

DOI: https://doi.org/10.17011/ht/urn.202011256762

Contents

Original Articles

HUMAN TECHNOLOGY

ISSN: 1795-6889
humantechnology.jyu.fi

Submission: humantechnologypublishing.jyu.fi

Human Technology is an interdisciplinary, scholarly journal publishing innovative, peer-reviewed articles exploring the issues and challenges within human–technology interaction and the human role in all areas of ICT-infused societies.

Human Technology, published by the Open Science Centre, University of Jyväskylä, is distributed without a charge online.

HUMAN TECHNOLOGY

ISSN: 1795-6889

www.humantechnology.jyu.fi Volume 16(3), November 2020, 227–232

Guest Editors' Introduction

MUSIC AS EMBODIED EXPERIENCE

Marc R. Thompson
Finnish Centre for Interdisciplinary Music Research
University of Jyväskylä
Finland

Jonna K. Vuoskoski
RITMO Centre for Interdisciplinary Studies in Rhythm, Time and Motion
University of Oslo
Norway

Perhaps more than ever, technology influences how people experience music. We live in an era of online music streaming services such as Spotify, app-based musical instrument lessons such as Yousician, and user-friendly digital audio workstations like Garageband. Respectively, these tools increase the consumption and awareness of new music, gamify instrumental tuition, and democratize the once "experts only" area of musical production. The technology in these applications has evolved to an extraordinary level whereby the software could be considered a coparticipant within the experience. Indeed, music technology can act as song recommender, teacher, and collaborator.

In addition to technological advancements, we also are witnessing in the current era a shift in perspectives toward music's societal function: Music is no longer simply a purely artistic pursuit but also a reliable therapeutic and salutogenic tool (see MacDonald, Kreutz & Mitchell, 2012). Music's healing power has been observed since antiquity (the story of David using his skillful harp-playing to cast out evil spirits residing inside King Saul comes to mind), yet the focus on music's role as a tool for emotional mood regulation (Saarikallio & Erkkilä, 2007) and how music-based interventions in clinical settings can aid in the process of stroke rehabilitation (Särkämö et al., 2008), just two diverse examples, is indeed unprecedented in history.

Taken together, technological advancement and changing attitudes toward music's societal function have led to new questions and topics for researchers interested in music perception and cognition. How is technology rewiring the musical brain? What role can music technology play in music therapy and music education?

In the mid- to late-20th century, these questions would have been answered using a cognitivist approach. Within that framework, mental processes would have been understood to be analogous to a digital computer with the mind operating as a central information processor of inputs and outputs. In a Cartesian sense, the mind would have been seen as separate from the physical

DOI: https://doi.org/10.17011/ht/urn.202011256763

world. Experimental designs from this era would have comprised stripped down, targeted listening tests in highly controlled settings; theories regarding how, for example, humans process musical structure would have been developed through proposed rule-based mental representations simulating real-world phenomena. In many ways, this disembodied approach to music cognition mirrored the then-recently discovered way of engaging with music. After all, it was just earlier in that century that humans first experienced music through recordings, far removed from the original sound source.

Yet throughout most of human history, musical experiences of all types required embodied actions. To participate in a musical activity, either as musician or audience member, required being within earshot of its source of production; most often, music was presented as a multimodal activity involving all the senses. The notion of music being a multimodal activity, one that incorporates all senses, is key to understanding why many music researchers started to adopt an embodied approach to music cognition. Since the late 1990s, the concept of embodiment has emerged as a viable and popular paradigm within music perception and cognition research, both as a concept to be explored philosophically and a phenomenon to be studied empirically. A recent collection edited by Lesaffre, Maes, and Leman (2017) demonstrates the multifaceted directions that embodied music cognition and music interaction have taken.

Proponents of the embodied cognition paradigm argue that the interactions between the body and the environment are central to shaping mental processes (Shapiro, 2014). In this view, the manner in which an organism in any realm learns and develops lies primarily on its body's physical shape and sensing properties. When applied to music, embodied music cognition views the body as the center of musical experience, acting as mediator between musical thought and action (Cox, 2016; Fincher-Kiefer, 2019; Leman, 2007). As a result, this research approach has assisted researchers in understanding how, for example, musicians' gestures enable observers to perceive the emotional expression of performers (Dahl & Friberg, 2007; Vuoskoski, Thompson, Spence & Clarke, 2014) and how dancers parse musical structure (i.e., metrical hierarchies) by physically mimicking temporal and acoustic features of music (e.g., Toiviainen, Luck & Thompson, 2010).

The embodied approach often relies heavily on high resolution capturing of movements or the tracking of day-to-day musical activities that would not be possible without recent technological advancements. Alongside epistemological developments, new technologies have influenced research approaches significantly, allowing new questions and methodologies to flourish. Such technologies also open space for reconsidering what musical experiences represent in this new millennium. First, motion capture of varying types has become a common method to measure and model music-related movements. Ranging from small portable inertial measurement units to expansive multicamera infrared optical systems, these technologies capture information with high temporal and spatial accuracy, allowing researchers to test hypotheses regarding how musicality is embedded within actions associated with music production, performance, and perception. Indeed, motion capture allows researchers to study a wide spectrum of interests, from modeling performers' gestures to tracking the rehabilitation of motor impairments through music therapy. Second, mobile apps can be used for experience sampling, which enables researchers to track personal listening habits (Randall & Rickard, 2017). This opens up intriguing new possibilities for investigating the experience of music in ecological ways. From these data, researchers model how listening choices affect behavior and vice versa, which in turn sheds light on, for instance, music's role in mood regulation and well-being.

The aim of the current Special Issue of *Human Technology* is to highlight how the wide diversity of technologies have matured to the point where they affect the way music is created, performed, enjoyed, and researched. For the issue, we encouraged submissions that would demonstrate how technology enables researchers to computationally model the structure of embodied musical expression and experience throughout various music-related activities. We received 11 stimulating and highly relevant submissions, of which six have been accepted for publication in this issue; an additional two remain under consideration for publication at a later date. These papers cover a range of topics, investigating music as a creative endeavor through performance and dance, as well as a tool toward healing, well-being, and development in therapeutic and educational settings.

The papers accepted for this issue can be divided into two broad groups: those focused specifically on embodiment—that is, embodied expression and communication in the context of musical performance and dance and those adopting a more applied approach to musical embodiment, investigating educational and therapeutic perspectives. In the former group, technology is used primarily as a tool enabling the systematic measurement, analysis, and modeling of music-related body movement, as well as an interface for embodied musical expression. In the latter group, technology could be characterized as serving a mediating role, facilitating educational, research, and therapeutic goals and processes.

Embodiment, Expression and Communication

Employing optical, marker-based motion capture as a method of data collection and stimulus generation, **Birgitta Burger and Petri Toiviainen** investigated how dancers embody musical emotions in their movements. Burger and Toiviainen collected ratings of perceived emotion from observers who watched silent stick-figure animations of people dancing. They later related the observer ratings to objectively measured movement characteristics, revealing consistent patterns in how emotion is perceived through visual channels of musically embodied movements. Using a similar combination of optical motion capture and perceptual experiments, **Anna Siminoski, Erica Huynh, and Michael Schutz** explored the role of auditory and visual feedback in ensemble performance. By manipulating the type and amount of coperformer feedback available to clarinet−piano duos, the authors were able to generate stimuli (audio recordings and point-light animations) for subsequent experiments investigating the perceptual implications of limited coperformer feedback under different presentation modalities (audio-only, visual-only, and audiovisual). **Lindsay Warrenburg, Lindsey Reymore, and Daniel Shanahan** also employed variations in presentation modalities (visual-only and audiovisual) in exploring the perception of emotion and sociality in video-recorded dance performances expressing melancholy, grief, and fear. In addition to collecting observer evaluations of perceived social connection and emotion, the video-recorded performances also were coded independently and systematically regarding the amount of physical contact between and among the dancers, suggesting a link between physical contact and perceived social connection. Finally, in a paper oriented toward the generation of new musical interfaces, **Çağrı Erdem, Qichao Lan, and Alexander Refsum Jensenius** explored the shapes of muscle energy used when performing on an electric guitar and then related them to the shapes of the resultant sounds. Subsequently, machine learning was used to generate muscle−sound design mappings for considering a new "air" instrument based on electromyography (EMG) sensors. One of the papers still under consideration for future publication is that of **Luis Aly, Hugo Silva, Gilberto Bernardes, and Rui Penha**. These authors focus on embodiment and a broad range of biosensors in their review of interactive musical interfaces, covering 70 interactive musical artworks spanning more than 50 years.

Educational and Therapeutic Applications

Embodied musical experience encompasses far more than just the production and consumption of music. Indeed, we editors felt it important to include in this thematic issue research that demonstrated novel applications of embodied technology in, for example, education and psychological therapy. The two accepted studies are examples of how music processing tools (e.g., Max, pure data) and general use interaction tools (e.g., Bela boards) have a near limitless applicability to projects highlighting gestural interaction with new technology and how these lend themselves seamlessly to creative educational and therapeutic endeavors. In the first paper in this area, **Kjetil Falkenberg, Hans Lindetorp, Adrian Benigno Latupeirissa, and Emma Frid** investigated the method of vocal sketching as a means to inspire young children collaborating with master of arts students in designing novel digital musical instruments. During a workshop between the graduate students and children, the children described their ideal musical interface and what it would sound like. The graduate students then developed novel instruments based on the children's ideas. While vocal sketching ended up playing a limited role in the completed products, the authors highlighted how the technique could be useful in a more inclusive design process of new musical interfaces by opening creative channels to young children or other nonexpert groups. Then, **Gabriela Patiño-Lakatos, Hugues Genevois, Benoît Navarret, Irema Barbosa-Magalhaes, Cristina Lindenmeyer, Maurice Corcos, and Aurélie Letranchant** presented the outcome of a clinical platform trial that combined sounds, music, and vibrotactile mediation. In their pilot study, eight adolescents diagnosed with anorexia nervosa were exposed to a variety of common sounds and musical excerpts mediated through five audiovibrotactile objects (e.g., ball, table, headrest pillow, blanket, and microphone). By allowing the adolescents full control over the sounds, music, and audiovibrotactile objects, the mediating objects offered the youths a multimodal sensory interaction with different parts of their bodies. The authors noted how the mediating objects triggered various associations, memories, and feelings in each adolescent that acted as springboards for communicating inner experiences during clinical therapy sessions. Finally, the other paper still under consideration for later publication is that of **Andrew Danso and Rebekah Rousi**, which focused on music learning by primary school students. They studied use of the iPad and the KAiKU Music Glove as part of the pupils' regular music class pedagogy and analyzed the data via the Technology Acceptance Model to investigate the role of each device in students' academic performance and music experience.

A thematic issue dedicated to technology's current and future impact on the embodied musical experience is both timely and necessary. Collectively, the contributions paint a picture of a multifaceted area that, albeit young, comes from a rich lineage. This Special Issue on Music, Mind & Technology (MMT) is named after an international graduate degree program that was offered by the University of Jyväskylä (Finland) from 2005 to 2018. MMT was an interdisciplinary program that attracted students from around the world interested in bridging the fields of musicology, cognitive science, and music technology. The program was also associated with the Academy of Finland funded Centre of Excellence (CoE) in Interdisciplinary Music Research (2008–2013). In the intervening years, the work of the MMT and CoE has continued through the recently formed Finnish Centre in Interdisciplinary Music Research. Meanwhile the field at large is growing with new established centers. For example, in 2018, the RITMO Centre for Interdisciplinary Studies in Rhythm, Time and Motion, was established in 2018 at the University of Oslo, Norway. In the same year, the University of Oslo, in collaboration with the Norwegian University of Science and Technology, opened a new master's program in Music, Communication and Technology (MCT). This new international program has taken up the mantle in training tomorrow's researchers enthusiastic about discovering the relationships among music, cognition, and technology. The above-mentioned groups are just examples of some of the exciting active research groups at the forefront of contemporary and innovative research on the musical experience. Though technology may have changed the way people engage with music, it remains a human endeavor. Research groups such as these ensure that the interdisciplinary field of music, cognition and technology will be pushed for years to come.

REFERENCES

Cox, A. (2016). *Music and embodied cognition: Listening, moving, feeling, and thinking*. Bloomington, IN: Indiana University Press.

Dahl, S., & Friberg, A. (2007). Visual perception of expressiveness in musicians' body movements. *Music Perception, 24*(5), 433–454. https://doi.org/10.1525/mp.2007.24.5.433

Fincher-Kiefer, R. (2019). *How the body shapes knowledge: Empirical support for embodied cognition*. Washington, DC, USA: American Psychological Association. https://doi.org/10.1037/0000136-000

Leman, M. (2007). *Embodied music cognition and mediation technolog*y. Cambridge, MA, USA: MIT Press.

Lesaffre, M., Maes, P.-J., & Leman, M. (Eds.). (2017). *The Routledge companion to embodied music interaction*. New York, NY, USA: Routledge.

MacDonald, R., Kreutz, G., & Mitchell, L. (Eds.). (2012). *Music, health, and well-being*. Oxford, UK: Oxford University Press.

Randall, W. M., & Rickard, N. S. (2017). Personal music listening: A model of emotional outcomes developed through mobile experience sampling. *Music Perception, 34*(5), 501–514. https://doi.org/10.1525/mp.2017.34.5.501

Saarikallio, S., & Erkkilä, J. (2007). The role of music in adolescents' mood regulation. *Psychology of Music, 35*(1), 88–109. https://doi.org/10.1177/0305735607068889

Särkämö, T., Tervaniemi, M., Laitinen, S., Forsblom, A., Soinila, S., Mikkonen, M., Autti, T., Silvennoinen, H. M., Erkkilä, J., Laine, M., Peretz, I., & Hietanen, M. (2008). Music listening enhances cognitive recovery and mood after middle cerebral artery stroke. *Brain, 131*(3), 866–876. https://doi.org/10.1093/brain/awn013

Shapiro, L. (Ed.). (2014). *The Routledge handbook of embodied cognitio*n. New York, NY, USA: Routledge. https://doi.org/10.1525/mp.2017.34.5.501

Toiviainen, P., Luck, G., & Thompson, M. R. (2010). Embodied meter: Hierarchical eigenmodes in music-induced movement. *Music Perceptio*n, *28*(1), 59–70. https://doi.org/10.1525/mp.2010.28.1.59

Vuoskoski, J., Thompson, M., Spence, C., & Clarke, E. (2014). Crossmodal interactions in the perception of expressivity in musical performance. *Attention, Perception, & Psychophysics, 76*(2), 591–604. https://doi.org/10.3758/ s13414-013-0582-2

Authors' Note

All correspondence should be addressed to
Marc R. Thompson
Department of Music, Art & Culture Studies (MACS)
PO Box 35(M)
FI-40014 University of Jyväskylä
Finland
marc.thompson[at]jyu.fi

Human Technology
ISSN 1795-6889
www.humantechnology.jyu.fi

HUMAN TECHNOLOGY

ISSN: 1795-6889

www.humantechnology.jyu.fi

Volume 16(3), November 2020, 233–256

SEE HOW IT FEELS TO MOVE: RELATIONSHIPS BETWEEN MOVEMENT CHARACTERISTICS AND PERCEPTION OF EMOTIONS IN DANCE

Birgitta Burger
Institute for Systematic Musicology
University of Hamburg
Germany
and
Finnish Centre for Interdisciplinary Music Research
Department of Music, Art and Culture Studies
University of Jyväskylä
Finland

Petri Toiviainen
Finnish Centre for Interdisciplinary Music Research
Department of Music, Art and Culture Studies
University of Jyväskylä
Finland

Abstract: *Music makes humans move in ways found to relate to, for instance, musical characteristics, personality, or emotional content of the music. In this study, we investigated associations between embodiments of musical emotions and the perception thereof. After collecting motion capture data of dancers moving to emotionally distinct musical stimuli, silent stick-figure animations were rated by a set of observers regarding perceived discrete emotions, while 10 movement features were computationally extracted from the motion capture data. Results indicate kinematic profiles—emotion-specific sets of movement characteristics—that furthermore conform with dimensional models of valence and arousal, suggesting that observers rated the emotions consistently according to distinct movement features prevalent in the animations. Outcomes show commonalities and differences to a previous study that linked these movement features to auditory perception of musical emotion, providing insights into how emotional expression of music-induced movement could be conveyed and understood through auditory and visual channels, respectively.*

Keywords: *music-induced movement, emotion, perception, computational movement feature extraction.*

DOI: https://doi.org/10.17011/ht/urn.202011256764

VERTAISARVIOITU KOLLEGIALT GRANSKAD PEER-REVIEWED www.tsv.fi/tunnus

INTRODUCTION

Oftentimes, listening to music makes humans move, and it can even be difficult to avoid spontaneous bodily responses, ranging from foot tapping or head nodding to full-body dance movement (Keller & Rieger, 2009; Lesaffre et al., 2008). Such movements usually are spontaneous and not choreographed; however, most often these responses are regular and organized, for instance, synchronized with the pulse of the music or mimicking instrumentalists' gestures (Leman & Godøy, 2010). Several factors have been found to affect music-induced movement characteristics: music-intrinsic features such as beat strength and pulse clarity (Burger, Thompson, Luck, Saarikallio, & Toiviainen, 2013; van Dyck, Moelants, et al., 2013), individual features such as personality and music preference (Luck, Saarikallio, Burger, Thompson, & Toiviainen, 2010) or mood (Saarikallio, Luck, Burger, Thompson, & Toiviainen, 2013), and the emotional content of the music (Burger, Saarikallio, Luck, Thompson, & Toiviainen, 2013; van Dyck, Maes, Hargreaves, Lesaffre, & Leman, 2013).

Such involvement and utilization of the body have contributed to the notion of embodied (music) cognition, which claims that human cognition and intelligent behavior require goal-directed interaction between the mind/brain, sensorimotor capabilities, body, and environment—a coupling of action and perception, rather than merely passive perception (e.g., Varela, Thompson, & Rosch, 1991). Musical involvement in this framework can be seen then as linking the perception of music to body movements, in that body movements reflect, imitate, and help listeners parse and understand musical elements (Leman, 2007). Leman suggested several (coexisting) components of corporeal articulations, differing in the degree of musical involvement and the kind of action–perception couplings employed. Synchronization constitutes the fundamental component and relates to entraining and predicting the beat structure. Embodied Attuning refers to linking body movement to musical features more complex than basic beat structure (e.g., melody, harmony, timbre, or rhythm). Finally, Empathy is seen as the component that links musical features to expressivity and social functions of music, including emotional expression.

As research on emotions and nonverbal behavior has shown, it is possible to successfully express and communicate emotional states to observers through body movements. Already in 1872, Darwin assigned certain body movements and postures quite specifically to emotional states; joyful movements, for example, were described as jumping, stamping, body thrown backwards and shaking, and upright torso and head, whereas anger was characterized by trembling body, shaking fist, erected head, and expanded chest. Sad movements were described as passive, motionless, and a downward directed head. Wallbott (1998) conducted a study in which he used a scenario-based approach with professional actors finding distinctive movement features for each emotion category; for instance, sadness was expressed with a collapsed body, whereas joy and anger were associated with an upright torso. Anger was the most active emotion, followed by joy and sadness. Lateral arm movement and dynamics/power related mostly to anger, less for joy, and even less for sadness. Spatially expansive movements typically represented anger and joy, whereas sad movements covered less space. De Meijer (1989) used actors performing several movement characteristics instead of emotions. These movements differed in general features, such as trunk or arm movement, velocity, and spatial direction. In a subsequent step, observers attributed emotional characteristics to the movements. The results suggested that fast, active, open, light, and upward directed movements with raised arms and a stretched trunk were perceived as happy, whereas strong and fast movements with a bowed torso and a high force were perceived as angry. In contrast, sad movements were described as slow, light, downward directed, with arms close to the body. Coulson (2004) used pictures of static body postures of computer-generated figures that he manipulated systematically in several parameters. Observers associated anger with a forward bent body having hands in front of the chest and bent elbows, whereas a forward-leaning and downward crouched body with downward-directed hands related to sadness. A backwards bent body with upraised arms and hands characterized as happy. Halovic and Kroos (2018) investigated emotion recognition in gait and found that a bouncing gait and increased arm movement portrayed happiness, whereas a slow and slouching gait suggested sadness, and a fast and stomping gait characterized anger.

Moreover, emotions are an essential component of musical engagement and experiences (e.g., Juslin & Sloboda, 2010; in particular, Gabrielsson, 2010; Juslin & Västfjäll, 2008) with strong influences, for instance, on the listener's mood (e.g., Saarikallio, 2011), as well as being one of the most important reasons for many to listen to music (Krumhansl, 2002). Various perceptual experiments have shown that listeners are able to recognize emotion in music (e.g., Eerola & Vuoskoski, 2011). Various approaches to categorizing emotions have been proposed: discrete emotions (e.g., Balkwill & Thompson, 1999), domain-specific emotion models (such as the GEMS model, see Zentner, Grandjean, & Scherer, 2008), or dimensional models (Ilie & Thompson, 2006). The discrete model described emotions as unidimensional, independent from each other, and derived from a limited number of universal basic emotions, such as happiness, anger, or sadness. An advantage of using this approach is the possibility to name the emotions in clear terms. However, a disadvantage is that they can be mixed and confused with other emotions, and the terms might be too generic for all possible emotions elicited by music. Domain-specific models, such as the GEMS (Geneva Emotion Music Scale), have been developed based on this approach to create a repertoire of emotions being highly music-related; however, these models might still bear the above-mentioned disadvantage. A dimensional approach could reduce this challenge. In such models, emotions are represented as a combination of two (usually valence and arousal) or three (valence, arousal, and tension) mutually independent dimensions, often illustrated as a two-(or three-) dimensional space with valence on the horizontal axis and arousal on the vertical axis (tension would form the third axis, yielding a cubical design). In such a model, happiness would be represented as an emotion high on arousal and high on valence, for instance, whereas sadness would be judged low on arousal and low on valence, and thus located opposite to happiness in a two-dimensional space. Anger, high on arousal and low on valence, thus, would be next to happiness and on top of sadness. Tenderness, commonly judged as low on arousal and high on valence, thus would be located below happiness and next to sadness (see Figure 1 for an illustration). Examples of such approaches are the circumplex models of affect by Russell (1980) and Schlossberg (1954). Eerola and Vuoskoski (2011) compared the different models and found that discrete and dimensional models performed similarly well in unambiguous, easily identifiable emotions, while the dimensional model outperformed the discrete model in case of ambiguous emotions.

Despite emotions and emotional impact being an essential component of music experiences, and emotions showing relationships to nonverbal behavior and movement, studies that link emotions with music-related movement are relatively rare. A few studies have been conducted in the field of music performance. Dahl and Friberg (2007), for instance, recorded

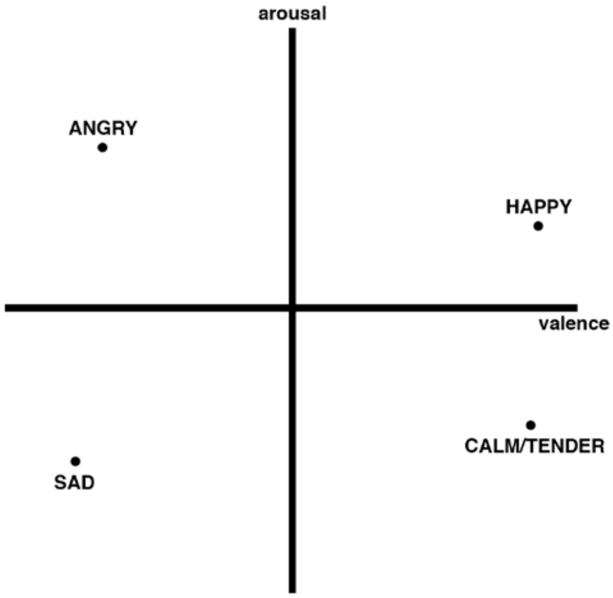


Figure 1. Locations of discrete emotions in a two-dimensional model as a combination of valence and arousal (based on the circumplex model of affect by Russell, 1980).

players of the marimba, bassoon, and saxophone performing with different emotional intentions (happy, angry, sad, and fearful) and had observers evaluate the emotion perceived in the movements and indicate movement cues conveying these emotions. Observers could identify successfully happy, angry, and sad performances (although they failed for fear) and indicated distinct cues: happy performances were communicated using medium regularity and fluency, high speed, and high amount of movement, while anger was conveyed with medium regularity, low fluency, high speed, and medium-to-high amount of movement. Sad performances were communicated with very regular and fluid movement, low speed, and small amount of movement. Based on these findings, Burger and Bresin (2010) developed a small mobile robot displaying these three emotions using a limited set of movements: happiness, for instance, reflected large, fluid, and circular movements; anger presented large, irregular, and jerky movements; whereas sadness was conveyed through slow, regular, and reduced movements. Their subsequent perceptual experiment demonstrated that subjects successfully recognized the emotions.

Camurri, Lagerlöf, and Volpe (2003) adopted a similar approach as Dahl and Friberg (2007) when asking professional dancers to perform the same dance in four different emotional characters. In their qualitative approach, they identified happiness as performed with frequent tempo changes with longer stops, dynamic movements, and changes between high and low tension; associated anger with short movements with short stops between, frequent tempo changes, as well as dynamic and tense movements; and portrayed sadness with long and smooth movements, few tempo changes, and low tension. Using a video-based analysis tool, Castellano, Villalba, and Camurri (2007) found two main movement characteristics that differentiated happy, angry, sad, and pleasurable dance gestures: The “Quantity of Motion” (QoM), that is, the overall amount of detected movement, and “Contraction Index” (CI), the contraction and expansion of the body, were measured in a two-dimensional space. QoM referred to the difference between high and low arousal (happiness and anger would be high, sadness and pleasure low on QoM), whereas CI to the difference between positive and negative emotions (happiness and pleasure high, anger and sadness low on CI). Boone and Cunningham (1998) found in a dance study in which actors displayed basic emotions that happiness was characterized by upward arm movements, angry movement by high tempo changes of the torso, sad movements by low tension, and fearful movements as rigid and upright.

In case of emotions in spontaneous movement induced by music (for an overview, see van Dyck, Burger, & Orlandatou, 2017), van Dyck, Maes et al. (2013) conducted a motion-capture study in which participants underwent an emotion induction task to feel either happy or sad and were subsequently asked to dance to an emotionally neutral piece of music. When induced to feel happy, participants moved faster and more accelerated, as well as more expanded and impulsive, as compared to when feeling sad. In a subsequent perceptual experiment (van Dyck, Vansteenkiste, Lenoir, Lesaffre, & Leman, 2014), observers successfully recognized the emotional state of the dancers to a high degree and paid most attention to movement of the dancers' chests. Burger, Saarikallio et al. (2013) conducted a study on perceived emotions in which they collected motion-capture data of spontaneous dancing, computationally extracted movement features from these data, and linked these to ratings of emotions perceived in the stimuli presented to the dancers. They found distinct movement feature combinations for each emotion. Happiness was linked to small foot distance as well as rotating and complex movements; anger was related to jerky and nonrotational, straight movements; sadness was associated with low movement complexity; and tenderness was linked to low head and hand acceleration, fluid movements, and a forward-bent posture.

In summary, emotions can be recognized in posture (e.g., Coulson, 2004), movement (e.g., de Meijer, 1989), gait (e.g., Halovic & Kroos, 2018), music performance (e.g., Dahl & Friberg, 2007), as well as the gestures of professional dancers (e.g., Camurri et al., 2003). Emotion perception in spontaneous dance/music-induced movement has been investigated much less frequently—with the exception of van Dyck et al. (2014), who used only happy and sad stimuli—despite the fact that visual information has a significant effect on how music is perceived (Tsay, 2013; Vuoskoski, Thompson, Clarke, & Spence, 2014). Therefore, in this study, we aimed to investigate emotion perception in music-induced movement. Increased knowledge on this topic is necessary to understand the multimodality of emotion perception—not only within dance and music contexts, but also more generally in human movement behavior.

The present investigation extends a previous study (Burger, Saarikallio et al., 2013) that investigated the relation between movement kinematics and emotions perceived in music (auditory domain) by exploring how subjective ratings of emotions displayed in video-only motion-capture animations (visual domain) related to movement kinematics. To this end, we drew on this earlier motion-capture study, in which participants freely moved to different musical stimuli, as a foundation for a perceptual experiment, in which a different set of participants evaluated motion-capture animations selected from the motion-capture dataset according to perceived emotions. For this analysis, we chose discrete emotion ratings (happiness, anger, sadness, and tenderness), as have been used in previous studies (in particular, movement-related ones), and set them in relation to the dimensional model. These discrete emotions were assumed to be more concrete and rather easily distinguishable and recognizable in dance movements, especially compared to the two axes of valence and arousal in the dimensional model or the more specific emotions in the GEMS model (Zenter et al., 2008). We chose video-only stimuli to prevent any influence of auditory input on participants' evaluations of movement. Moreover, we selected the same set of movement features used in Burger, Saarikallio et al. (2013) for comparability reasons.

We asked the following three research questions:

1. What are the relations between perceived emotions and the movement features of dance-like movement presented as motion-capture animations?
2. How do these movement features link to movement features found earlier related to emotion ratings of the audio stimuli?
3. How do the results for the discrete emotions relate to the dimensional model of valence and arousal (Russell, 1980)?

Based on the studies mentioned above, we predicted that each emotion would link to a specific set of movement characteristics, following and extending the results of Burger, Saarikallio et al. (2013). Thus, we expected that movements high on rotation and complexity, as well as having small foot distance, would be evaluated as happy, jerky movements without rotation as angry, movement of low complexity as sad, and smooth movement with low acceleration and forward tilted torso as tender. We further hypothesized that locating the discrete emotion results in a two-dimensional space would be consistent with the circumplex model of valence and arousal (Russell, 1980).

METHODS

The following section describes the methods that we used in this study. We drew on previously collected motion capture data that we subsequently utilized in a perceptual study to investigate relationships between embodiments of musical emotions and the perception thereof. The results of the motion capture study are published in Burger, Saarikallio et al. (2013) and will be used as references in the current analysis.

Motion Capture Data Collection

Participants and Stimuli

Sixty participants (43 female, 17 male; average age 23.92 years; *SD* = 3.27) took part in the underlying motion-capture study, recruited from the University of Jyväskylä’s general student population. Six participants had received formal music education, although 30 reported playing a musical instrument. Four participants had a formal background in dance, and 27 had received some kind of dance tuition during their lives. Fifty-one participants reported liking to dance to music. All participants expressed engaging in some kind of sports activity. Participation was rewarded with a cinema voucher. All participants gave their informed consent prior to their inclusion in the study and were free to discontinue the experiment at any point. Ethical permission for this study was not needed, according to the guidelines stated by the University of Jyväskylä’s (Finland) ethical board.

We presented participants with 30 musical excerpts of various popular music genres (techno, pop, rock, Latin, funk, and jazz). All stimuli were naturalistic (existing music) to make the experiment as ecologically valid as possible. The thirty stimuli provided participants with a wide range of music genres and emotional expression and catered to preferences and familiarities among the participants. A team of music researchers at the University of Jyväskylä selected the music based on these characteristics and the following: All stimuli were nonvocal, in 4/4 time, differed in their pulse clarity and rhythmic complexity (i.e., all with a perceivable beat), and had a tempo range of 82 to 140 BPM. We used a length of 30 seconds per song: This stimulus length kept the experiment sufficiently short while providing stimuli long enough to induce movement.

Apparatus and Procedure

Participants' movements were recorded using an eight-camera optical motion capture system (Qualisys ProReflex), tracking the three-dimensional positions of 28 reflective markers attached to each participant at a frame rate of 120 Hz. Figure 2 presents the locations of the markers. The location of the markers were as follows: (L = left, R = right, F = front, B = back): 1: LF head; 2: RF head; 3: LB head; 4: RB head; 5: L shoulder; 6: R shoulder; 7: sternum; 8: spine (T5); 9: LF hip; 10: RF hip; 11: LB hip; 12: RB hip; 13: L elbow; 14: R elbow; 15: L wrist/radius; 16: L wrist/ulna; 17: R wrist/radius; 18: R wrist/ulna; 19: L middle finger; 20: R middle finger; 21: L knee; 22: R knee; 23: L ankle; 24: R ankle; 25: L heel; 26: R heel; 27: L big toe; 28: R big toe. The musical stimuli were played via a pair of Genelec 8030A loudspeakers using a Max/MSP patch running on an Apple computer. We recorded the room sound with two overhead microphones positioned at a height of approximately 2.5 m. This microphone input, the direct audio signal of the playback, and the synchronization pulse transmitted by the Qualisys cameras when recording, were recorded using ProTools software in order to synchronize the motion capture data with the musical stimulus afterwards.

Participants were recorded individually and asked to move to the presented stimuli in a way that felt natural. Additionally, they were encouraged to dance if they wished but were requested to remain in the center of the capture space indicated by a 115 x 200 cm mat.

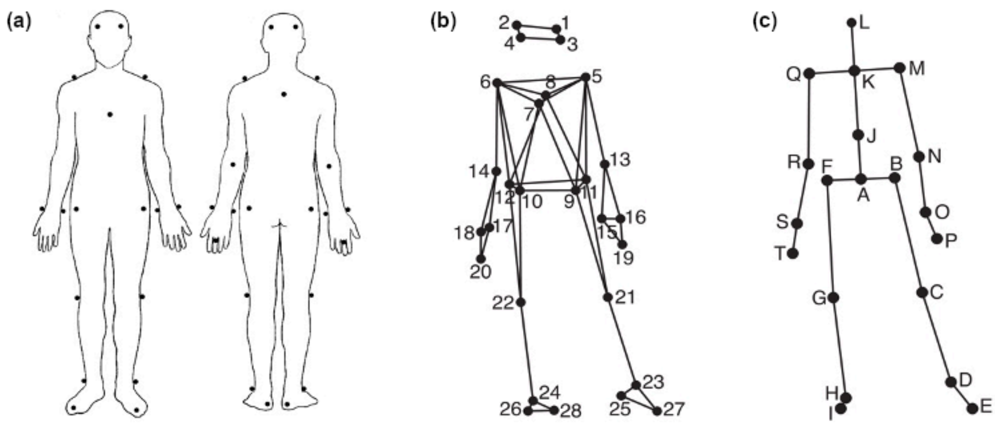


Figure 2. Marker and joint locations to track three-dimensional movement to music: (a) anterior and posterior view of the marker placement on the participants' bodies; (b) anterior view of the marker locations as stick figure illustration; (c) anterior view of the locations of the secondary markers/joints used in the animations and analysis.

Motion-Capture Data Preprocessing

After labeling the motion-capture (mocap) data in the Qualisys Track Manager software, we used the MATLAB Motion Capture Toolbox (Burger & Toiviainen, 2013) to further process the data in creating the stick figure animations and for analysis purposes. First, we edited the recordings to provide only the duration of each musical stimulus. Following this, we derived a set of 20 secondary markers—subsequently referred to as joints—from the original 28 markers by creating virtual joints (where one cannot place a mocap marker) and removing redundant markers as well as to receive an easily perceivable stick figure representation. The locations of these 20 joints are depicted in Figure 2c. The locations of joints C, D, E, G, H, I, M, N, P, Q, R, and T are identical to the locations of one of the original markers; the locations of the remaining joints were obtained by averaging the locations of two or more markers; joint A: midpoint of the four hip markers; B: midpoint of markers 9 and 11 (left hip); F: midpoint of markers 10 and 12 (right hip); J: midpoint of breastbone, spine, and the hip markers (mid-torso); K: midpoint of shoulder markers (manubrium), L: midpoint of the four head markers (head); O: midpoint of the two left wrist markers (left wrist); S: mid-point of the two right wrist markers (right wrist).

Perceptual Study

Stimulus Creation/Stimuli

The visual stimuli used in this experiment were selected from the pool of the above described motion capture study. Based on results of a rating experiment conducted within a previous study, in which 30 independent raters rated the audio stimuli regarding perceived happiness, anger, sadness, and tenderness (see Burger, Saarikallio et al., 2013), we identified four musical excerpts that most clearly conveyed one of the following emotions: happiness, anger, sadness, and tenderness (see Table 1). We chose four songs to keep the experiment sufficiently short. Subsequently, to present observers with as representative, yet variable, examples as possible, we chose two female and two male participants (from now on referred to as “dancers”). We selected these dancers based on their movement characteristics (i.e., dancers whose performances received

Table 1. Overview of the Stimuli for the Target Emotions, Including Artist, Song, Tempo, and Average Rating of 30 Independent Observers of the Respective Target Emotion (Rating Scale 1–7).

Target emotion	Artist	Song	Tempo	Average rating
Happiness	Dave Weckl (1994) (track 2, 0:10-0:41)	“Mercy, Mercy, Mercy”	105	5.15
Anger	In Flames (2006) (track 5, 0:00-0:30)	“Scream”	100	6.26
Sadness	Johanna Kurkela (2005) (track 2, 3:22-3:52)	“Hetki hiljaa”	122	5.18
Tenderness	The Rippingtons (1992) (track 1, 1:13-1:42)	“Weekend in Monaco”	113	4.53

the highest values regarding several movement features—such as speed, acceleration, area covered, or complexity—that Burger, Saarikallio et al., 2013, found significant for each emotion in their previous analysis). The four chosen dancers moderately liked the stimuli (ratings between 3 and 4.25 on a scale from 1 to 5) and were moderately familiar with them (ratings between 3 and 4 on a scale from 1 to 5). We then created stick figure animations of these movements via the MATLAB Motion Capture Toolbox using the joint configuration described above (see also Figure 2c). The animations were trimmed to the mid-20 s of each original performance, yielding 16 (4 dancers x 4 stimuli) combinations of stick-figure dance performances shown to the observing participants.

The four musical stimuli the dancers moved to were of different tempi (100 bpm, 105 bpm, 113 bpm, 122 bpm). In order to avoid issues that might arise from stimuli of different tempi, we time-shifted the movement data using the MATLAB Motion Capture Toolbox and adjusted their tempi to 113 bpm (corresponding approximately to the mean tempo, with keeping one at the original tempo). In addition, we rotated the movement data on the stick figures to be visible from the front with respect to the average locations of the hip markers. In all videos, the stick figures were plotted in black on a white background.

Participants

Eighty university students (from now on referred to as "observers"), aged 19–36 years, participated in the study (53 females; average age: 24.74; *SD* of age: 3.42). None of the observers participated in the preceding motion-capture experiment. We recruited observers from the University of Jyväskylä's general student population. In this group, 35 observers had not received any musical training, 23 had received between 1 and 5 years of musical training, and the remaining 22 had received more than 5 years of musical training. Fifty-two observers had taken some type of dance lessons; 75 observers reported liking movement-related activities, such as dance and sports. None of the participants was a professional musician, dancer, or athlete. Thirty-seven observers reported going out to dance more than once a month. Observers were rewarded with a movie ticket for participating in the study.

Apparatus

To gather the perceptual ratings, a custom patch was created in Max/MSP 5, a graphical programming environment, running on Max OS X on an Apple iMac computer (see Figure 3). The patch used QuickTime Player 7 to play back the video material in a separate window. The setup enabled the observers to repeat excerpts as often as they wished, to move forward at their own speed, and take breaks at any moment of the experiment.

Procedure

Observers accomplished the experiment individually in a soundproof room. They were instructed to rate the emotions expressed in the clips (perceived emotions) on seven-step scales (from *not at all* to *very much*) for Happiness, Anger, Sadness, and Tenderness. The 16 silent video clips were presented in randomized order. Prior to the experiment, observers completed a short questionnaire regarding their gender, age, musical and dance training, and movement and dance activity. We also provided observers with a practice example so that they became

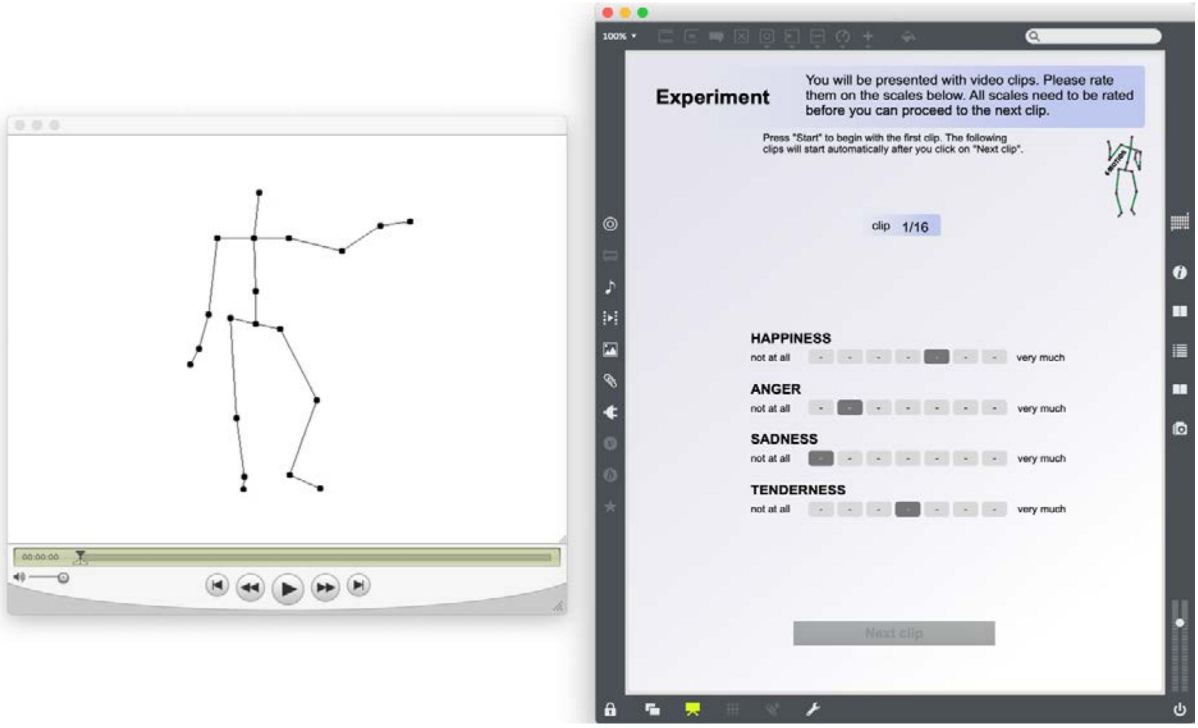


Figure 3. Screenshot of the Max/MSP patch used to collect the emotional perceptual ratings of the stick figure animations.

familiar with the interface, the type of stimuli, and the rating scales used. At the start of the experiment, we explicitly instructed participants to rate according to what emotion they thought the dancers wanted to express in the clip and not what they might feel when watching the clip. The participants were advised to take breaks in during the experiment if they felt like it. The experiment took between 10 and 25 minutes depending on the participants' rating speed.

Movement Feature Extraction and Analysis

Following the analysis process performed in Burger, Saarikallio et al. (2013), we extracted 10 movement features from the motion capture data. As a preparatory step, the instantaneous velocity and acceleration were estimated from the three-dimensional joint position data using numerical differentiation and a Savitzky-Golay smoothing FIR filter (Savitzky & Golay, 1964) with a window length of seven samples and a polynomial order of two. These values were found to provide an optimal combination of precision and smoothness in the time derivatives. Furthermore, to express the data relative to the center of the body with defined directions of the body axes, we transformed the data from a global to a local coordinate system by firstly rotating them on a frame-by-frame basis. As a result, the hip joints (A, B, and F) were aligned to be parallel to the x-axis and the origin of the coordinate system was shifted to the mid-point of the hip (Joint A). This procedure is visualized using position data in Figure 4. Most features, unless noted otherwise below, were calculated using the local coordinate system.

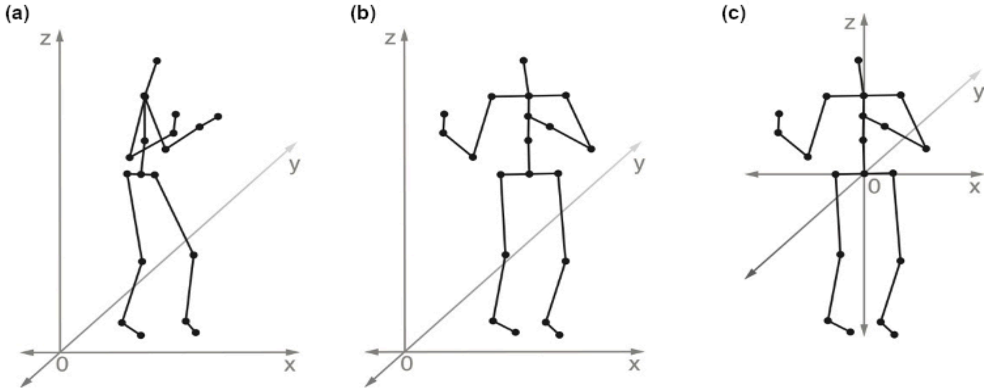


Figure 4. Transformation from global to local coordinate system exemplified with one frame of position data: (a) Original position data; (b) rotation on a frame-by-frame basis so that the hip markers are aligned parallel to the x-axis; (c) shift of the origin of the coordinate system to the midpoint of the hip to represent the data relative to this point.

The following features were used:

- Postural features
  - Torso Tilt: vertical tilt of the torso (Joints A–K), positive tilt related to bending forward.
  - Hand Distance: distance between hands (Joints P and T).
  - Foot Distance: distance between feet (Joints E and I).
- Local kinematics
  - magnitude of Head Acceleration (Joint L).
  - magnitude of Hand Acceleration (Joints P and T).
  - magnitude of Foot Acceleration (Joints E and I).
- Global kinematics
  - Area of Movement: The smallest rectangle that contains the projection of the trajectory of the Center of Mass (Joint A) on the horizontal plane (i.e., floor), averaged across a 4 s analysis windows with a 2 s overlap. This feature was calculated based on the global coordinate system.
  - Fluidity: Overall movement, fluidity/smoothness measure based on the ratio of velocity to acceleration. The combination of high velocity and low acceleration reflects fluid movement, whereas the combination of low velocity and high acceleration reflects nonfluid movement.
  - Rotation Range: Amount of rotation of the body (Joints M and Q) around the vertical axis.
  - Movement Complexity: Based on Principal Components Analysis (PCA) of the position data (all joints), in this case, can be understood as a decomposition into distinct (independent) movement components. The more principal components required to explain the movement, the more dimensions are present in the movement and thus the more complex the movement has been. The feature is quantified by taking the cumulative sum of the variance that could be explained by the first five principal components. A more extensive description of this feature can be found in Burger, Saarikallio et al. (2013).

Subsequently, we averaged the instantaneous time–series values of each variable across the dancers for each stimulus presentation to receive one value per observer per emotion stimulus. These values were used in the statistical analysis presented below.

RESULTS

Movement Features

Figure 5 shows the distribution of movement features per emotion/stimulus averaged across the four dancers. Dancers altered their movements considerably when dancing to the different stimuli. We did not perform inferential statistics on these data, as there were only four dancers.

Next, we correlated the movement features with each other to assess relationships among them. In our aim of using unique and independent features in the analysis, an uncorrelated feature

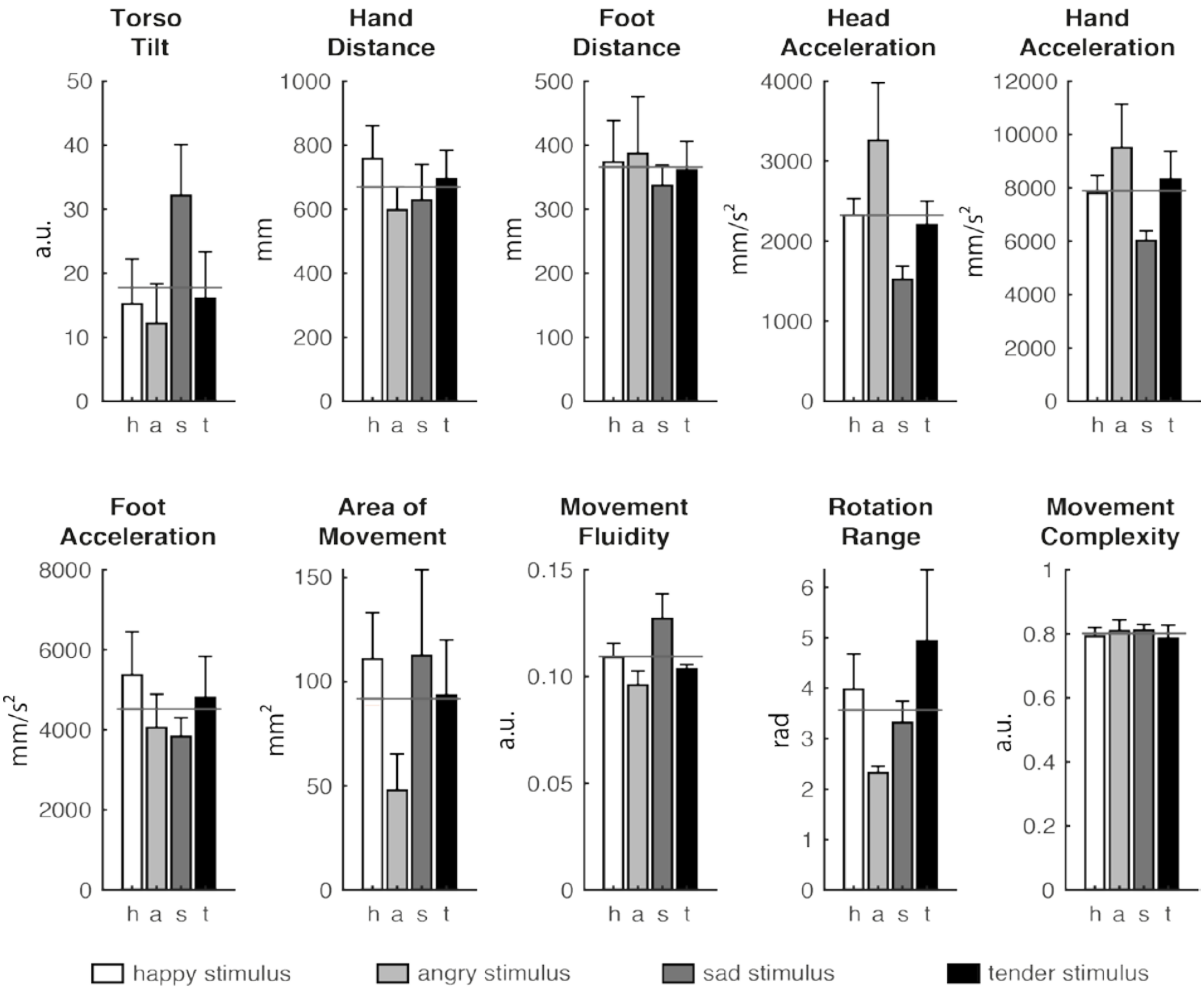


Figure 5. Distribution of movement features per emotion/stimulus averaged across the four dancers. The horizontal line in each subplot indicates the mean across the stimuli. Error bars indicate the standard error. *Mm* refers to millimeters, *s* to seconds, *rad* to radian, and *a.u.* to arbitrary unit.

set is desired; however, because the features are based on markers placed on the same person, it is likely that the selected features displayed some degree of dependence. Indeed, features based on similar markers or calculations correlated significantly, such as Foot Acceleration, Complexity and Area of Movement, as well as Head and Hand Acceleration. Table 2 provides the full correlation matrix.

Perceptual Ratings

We correlated the perceptual ratings with each other using Spearman correlations (as the rating data were considered ordinal) and found significant negative correlations between Happiness and Sadness ratings and between Anger and Tenderness ratings. The full correlation results are displayed in Table 3. The results suggest that observers were able to distinguish meaningfully between the emotions, aligning them as suggested in Russell's dimensional model of valence and arousal (Russell, 1980).

In order to verify whether the 80 observers could recognize the target emotions when watching the movements, we conducted repeated measures ANOVAs. For the Happiness ratings,

Table 2. Pearson Correlations (N = 16 video stimuli) of Movement Features; Significant Correlations are Indicated in Bold (*** $p < .001$, ** $p < .01$, and * $p < .05$).

	To Ti	Ha Di	Fo Di	He Ac	He Ac	Fo Ac	Area	Fluid	Rot
Ha Di	-.29								
Fo Di	.23	-.44							
He Ac	**.75*** **	-.16	**.65** **						
Ha Ac	**.61** **	-.21	.38	**.75*** **					
Fo Ac	.44	-.43	.31	.33	.48				
Area	-.01	-.26	.14	-.09	-.04	**.70** **			
Fluid	-.33	**.51***	-.23	-.37	**-.54***	-.34	.15		
Rot	-.08	.05	.10	-.03	.17	.43	.33	.17	
Comp	-.24	**.61***	-.45	-.24	-.41	**-.71** **	**-.53***	.23	-.47

Note. To Ti–Torso Tilt; Ha Di–Hand Distance; Fo Di–Foot Distance; He Ac–Head Acceleration; Ha Ac–Hand Acceleration; Fo Ac–Foot Acceleration; Area–Area of Movement; Fluid–Fluidity; Rot–Rotation Range; Comp–Complexity.

Table 3. Spearman Correlations (N = 16 video stimuli) of Emotion Ratings; Significant Correlations are Indicated in Bold (*** $p < .001$).

	Happiness	Anger	Sadness
Anger	-.38		
Sadness	**-.84*** **	-.14	
Tenderness	.26	**-.83*** **	.43

the happy stimuli received the highest ratings on average (*M* = 4.51, *SD* = 0.88), followed by tender (*M* = 4.23, *SD* = 0.97), angry (*M* = 3.52, *SD* = 0.94), and sad (*M* = 3.11, *SD* = 0.86). The repeated measures ANOVA showed a significant main effect, $F(3, 237) = 81.61$, $p < .001$, $\eta_{\mathrm{p}}^2 = .51$. Follow-up pairwise comparisons using Bonferroni correction for multiple comparisons revealed significant differences between the happy and the other three stimuli (*p* between .000 and .01, levels indicated in Figure 6), indicating that observers rated the happy stimuli significantly higher on the Happiness scale than the other stimuli. For the Anger ratings, the angry stimuli received the highest ratings on average (*M* = 2.54, *SD* = 0.97), followed by happy (*M* = 1.48, *SD* = 0.69), tender (*M* = 1.38, *SD* = 0.62), and sad (*M* = 1.33, *SD* = 0.60). The repeated measures ANOVA showed a significant main effect, $F(1.54, 121.64) = 99.15$, $p < .001$, $\eta_p^2 = .56$ (Greenhouse-Geisser correction due to violation of sphericity). Follow-up pairwise comparisons revealed significant differences between the angry and the other three stimuli (all $p < .001$, see Figure 6), indicating that observers rated the angry stimuli significantly higher on the Anger scale than the other stimuli. Noteworthy though are the lower averages overall compared to the Happiness ratings. For the Sadness ratings, the sad stimuli received the highest ratings on average (*M* = 3.03, *SD* = 1.13), followed by tender (*M* = 1.95, *SD* = 0.81), angry (*M* = 1.90, *SD* = 0.76), and happy (*M* = 1.75, *SD* = 0.71). The repeated measures ANOVA showed a significant main effect, $F(2.12, 167.28) = 95.94$, $p < .001$, $\eta_p^2 = .55$ (Greenhouse-Geisser correction due to violation of sphericity). Follow-up pairwise comparisons revealed significant differences between the sad and the other three stimuli (all $p < .001$, see Figure 6), indicating that observers rated the sad stimuli significantly higher on the Sadness scale than the other stimuli. For the Tenderness ratings, the sad stimuli received the highest ratings on average (*M* = 4.09, *SD* = 1.01),

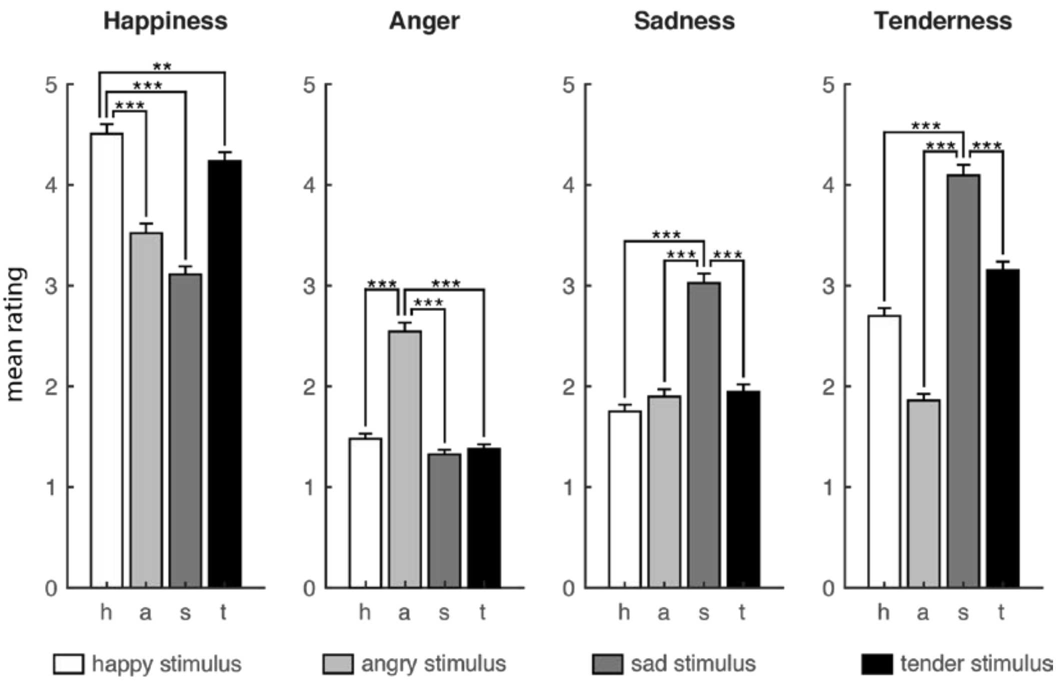


Figure 6. Overview of the emotion ratings. Significant differences only indicated related to the emotion that received highest rating (*** $p < .001$ and ** $p < .01$).

followed by tender (M = 3.15, SD = 1.12), happy (M = 2.70, SD = 1.08), and angry (M = 0.86, SD = 0.87). The repeated measures ANOVA showed a significant main effect, $F(2.36, 186.31) = 147.70$, $p < .001$, $\eta_p^2 = .69$ (Greenhouse-Geisser correction due to violation of sphericity). Follow-up pairwise comparisons revealed significant differences between all stimuli ($p < .001$, see Figure 6); however, the tender stimuli were not rated highest on the Tenderness but on the Sadness scale, indicating a confusion between these target emotions.

Linking Movement Features and Perceptual Ratings

In order to identify relationships between movement characteristics and the emotion ratings of the video animations, we conducted Spearman correlations. Due to the large number of correlations (40), we controlled the false discovery rate by applying the Benjamini-Hochberg procedure (see Benjamini & Hochberg, 1995) at a significance level of $p < .05$. The full results can be found in Table 4.

Happiness ratings correlated significantly with Head and Foot Acceleration, Area of Movement, and Complexity (all positive apart from Complexity); thus observers rated stick figure videos high on happiness when dancers had elevated head and foot acceleration, covered a large area, and had low-dimensional movement complexity. Anger ratings showed significant negative correlations with Area of Movement and Rotation Range, suggesting that observers rated stick figure clips as angry when dancers used a small area and did not rotate. Sadness ratings showed significant negative correlations with Head, Hand, and Foot Acceleration and positive correlations with Torso Tilt, Fluidity, and Complexity. Hence, observers rated stick figure videos high on sadness when dancers showed limited head, hand, and foot acceleration; were upright;

Table 4. Spearman Correlations (N = 16 video stimuli) Between Movement Features and Emotion Ratings; Significant Correlations are Indicated in Bold (*** $p < .001$, ** $p < .01$, and * $p < .05$ after FDR correction).

		Perceptual ratings			
		Happiness	**Anger**	**Sadness**	**Tenderness**
Movement features	**Torso Tilt**	.38	.24	**-.60***	-.27
	Hand Distance	-.49	.15	.39	-.21
	Foot Distance	.55	-.12	-.54	.18
	Head Acceleration	**.60***	.20	**-.78*****	-.32
	Hand Acceleration	.47	.23	**-.69****	-.46
	Foot Acceleration	**.84*****	-.41	**-.71****	.19
	Area of Movement	**.67****	**-.76*****	-.29	**.62****
	Fluidity	-.46	-.35	**.68****	.32
	Rotation Range	.53	**-.67****	-.19	.44
	Complexity	**-.83*****	.47	**.56***	-.39

and presented highly fluid and complex movement. Tenderness ratings correlated significantly with Area of Movement, suggesting that observers rated stick figure clips as tender when dancers used a large amount of space.

In order to relate our correlational results to Russell's (1980) circumplex model of valence and arousal, we followed Russell's approach, using multidimensional scaling (MDS) to achieve a representation of the data in a two-dimensional space. First, we correlated the columns of the correlation matrix of Table 4 (subsequently referred to as the kinematic profiles of the respective emotions) with each other using Spearman correlation to obtain an emotion similarity matrix based on their dependence on the kinematics. Next, we created a full dissimilarity matrix as a distance matrix required for the multidimensional scaling by subtracting the correlation values from 1 (to obtain zeros along the diagonal and otherwise positive elements only). Subsequently, MDS was performed on this dissimilarity matrix (using *cmdscale* in MATLAB). As the rotational configuration of the MDS solution is arbitrary, the initial solution was rotated by 65° counterclockwise in order to better line up with the placement of discrete emotions in the circumplex model (Russell, 1980). The graphical depiction is shown in Figure 7. For reference, we added Russell's discrete emotion locations to the plot. As can be seen, the locations align well, with an amplification of the arousal axis, suggesting that relationships between movement features and perceived emotions are consistent and fit well with this model.

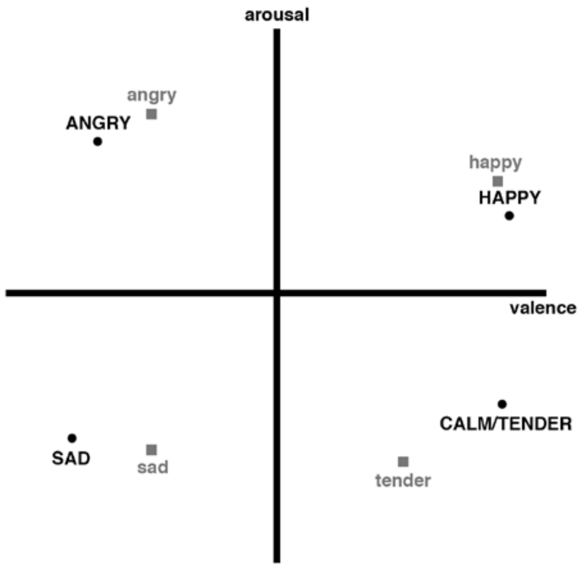


Figure 7. Multidimensional scaling solution of the emotion rating/movement feature correlations (indicated by the squares, gray font, and lowercase letters), together with Russell's (1980) multidimensional scaling solution for emotion terms according to the valence/arousal axes (indicated by the circles, black font, and uppercase letters).

DISCUSSION

In this study, we related the kinematics of music-induced movement to visual perception of emotions in corresponding silent motion capture animations, aiming to investigate the relationships between embodied responses to music and human perceptual evaluations of these responses. Each emotion could be linked to a specific set of movement characteristics, resulting in kinematic profiles, suggesting that observers rated the emotions consistently according to distinct movement features prevalent in the animations. Furthermore, the movement–emotion links conformed well to the location of discrete emotions in the circumplex model of emotions (Russell, 1980), indicating high consistency and unambiguousness in how observers perceived emotional expression in music-induced dance movements. This investigation forms an illustrative example of embodied music cognition (Leman, 2007) and provides support for Leman's notion of empathy, where music conveys emotional qualities that can be embodied and expressed via movement. The results furthermore provide insights into the multimodality of emotion perception of movement.

Earlier studies often employed professional actors or dancers asked to portray specific emotions (e.g., Camurri et al., 2003; Castellano et al., 2007). Such an approach yields stylized movements that might be very representative for a certain emotion, but might lack naturalness. However, the stimuli used in the current study derived from quasispontaneous movement to music, which potentially lacked emotion-specific qualities because the dancers were not instructed specifically to express certain emotions via their movements. Despite the dancers being unaware of any connections to emotions, they still moved in ways that observers linked to emotional expressions. Furthermore, our choosing such quasispontaneous movements, as opposed to using staged movements, in combination with existing music and nonprofessional dancers, increased the ecological validity of the study, in that such movements occurred naturally.

The musical stimuli corresponding to the four discrete emotions showed mutual differences in the ways the four dancers, used later as stick figure animations, expressed them. Noticeably, happy-sounding music let dancers move with higher foot acceleration and using a larger area, whereas moving to an angry-sounding musical stimulus increased head acceleration and decreased fluidity (e.g., head banging). Fluidity and head acceleration were found highest for sad-sounding excerpts; tender-sounding stimuli showed the highest amount of rotation. Although the analysis of these characteristics is not the focus of this paper (as our research aim was to connect these features with perceptual evaluations), the results are in line with the results of the larger data set to which the four dancers selected for this study belong (see Burger, Saarikallio et al., 2013). Thus, this research provides an adequate representation of movement features communicating emotional expressions useful in perceptual experiments. The results furthermore are in line with the characteristics found in studies on dance movement (e.g., Camurri et al., 2003; Castellano et al., 2007), as well as general movement studies (e.g., Coulson, 2004; de Meijer, 1989; Wallbott, 1998).

As expected, the selected movement features correlated to some extent with each other. Because the human body is coupled, and thus also are the related motion-capture data, movement features of the same or adjacent body parts (e.g., feet or hand and elbow), as well as movement features of the same kind (e.g., acceleration features of various body parts), often correlate significantly. This is the case in the present findings as well, such as Acceleration of Head and Hands, Fluidity and Acceleration of Hands, or Foot Acceleration and Area of Movement. However, the features still cover different movement qualities and characteristics (e.g., Area of Movement and Complexity or Head Acceleration and Foot Distance), thus we decided to keep these variables in further analyses.

The emotion ratings provided by the observers in the perceptual experiment resulted in significant negative correlations for Happiness and Sadness as well as for Anger and Tenderness. These findings align with research on the relations between discrete emotions

(Eerola & Vuoskoski, 2011; Russell, 1980; Vieillard et al., 2008), and indicate both an opposite/adverse perception of these emotion pairs and a coherent and consistent rating behavior of the observers overall.

Observers were able to identify successfully the target emotions in case of Happiness, Anger, and Sadness (i.e., rated the respective stimuli significantly higher on the intended emotion scale than the other stimuli). Interesting to note is that ratings for Anger and Sadness were much lower than for the other two emotions (see Figure 6). This could indicate difficulties in portraying negative emotions via spontaneous dance movements, as observers would more commonly connect dance with positive feelings, pleasure, and fun. Acted portrayal of emotions, as used in previous studies (e.g., Camurri et al., 2003), might have resulted in clearer outcomes.

In case of Tenderness, however, observers rated Sadness highest, indicating a confusion between the two low arousal emotions. Music expressing either sadness or tenderness could have elicited similar movements, for instance, fluid movement as indicated in Burger, Saarikallio et al. (2013) and Dahl and Friberg (2007). Thus, when watching the animations without sound, movements could have been similarly mistaken for sadness rather than tenderness. Such confusion relates to the communication of tenderness that has been found previously in both music and movement-related research (e.g., Eerola & Vuoskoski, 2011; Gross, Crane, & Fredrickson, 2010; Pollick, Paterson, Bruderlin, & Sanford, 2001). Interestingly, the Tender stimuli rated as Sad happened much more often than the Sad stimuli rated as Tender. It could be posited that the nature of the stimuli and their acoustic features serve as explanatory factors, in that the prerequisite for the stimuli selection was having some regular beat structure to induce movement responses. However, it is difficult—if not impossible—to find musical stimuli that express either tenderness or sadness and simultaneously feature a perceivable beat, as a stronger beat and higher onset rate would increase the arousal level of the stimuli and thus be rated towards happiness or anger/fear respectively (Grekow, 2018; Laurier, Lartillot, Eerola, & Toiviainen, 2009). Consequently, movements to Tender stimuli might exhibit similar characteristics as movements to sad-sounding music, both being on a low arousal level. As the observers rated the movements without being able to listen to the music perceived by the dancers, the movements might have appeared more towards the sad than the tender side.

By bringing movement features and emotion ratings together, we were able to identify several relationships—kinematic profiles—between kinematics and emotion evaluations, illuminating how observers might base emotional judgments on movement characteristics when viewing silent dance animations. Observers rated clips as happy when dancers used high acceleration of head and feet, showed low-dimensional movement, and covered a large area, thus overall active movement using a large area. Moreover, we found a moderate correlation for rotation that was close to being significant (i.e., nonsignificant after adjusting the significance levels). This result is in line with the result obtained in Burger, Saarikallio et al. (2013) that found rotation related to happiness presented in the audio stimuli. The correlation values are indeed very close to each other (.53 in this analysis and .55 in Burger, Saarikallio et al., 2013), suggesting that participants also linked rotating movements with happiness when perceiving the videos. Due to the different sample size—30 in the previous analysis and 16 in the current one—however, the threshold of the correlation values to be flagged significant was higher in this analysis, marking the correlation as nonsignificant. Yet, we could not replicate the significant positive correlation between movement complexity and Happiness ratings of the audio examples. This result seems counterintuitive, as it could be expected that happiness—as an active emotion—would be embodied using complex movement as well. Moreover, combined with the positive values for acceleration and rotation, one could assume a positive relationship with complexity, too. However, this result could have emerged due to the larger sample size (60 dancers, 30 stimuli) in Burger, Saarikallio et al. (2013), so that the reduced sample (four dancers, four musical stimuli) used in the current study might not have offered the same variability. Thus, the four dancers might have moved in complex ways for all four stimuli, explaining the rather similar values of this movement feature (see Figure 5—the average for the Happy stimulus is slightly lower as compared to the other three stimuli). Furthermore, this analysis found significant relationships that the previous analysis did not find, although these correlations (acceleration features and area of movement) were in the same direction and highlight the relation between active movement and happiness. These results are in line as well with previous findings, for instance, by Burger and Bresin (2010), Dahl and Friberg (2007), and De Meijer (1989).

Observers perceived anger in movements covering small areas and without rotation, which means dancers were standing mainly in one spot and were neither moving nor turning around. This is very much in line with results described in Burger, Saarikallio et al. (2013), where significant negative correlations were found between fluidity and rotation and the angry-sounding music stimuli. Fluidity also correlated negatively (not significant) in the current analysis, while area correlated negatively (not significant) in the previous analysis, corroborating the results. Previous research, such as Dahl and Friberg (2007), Camurri et al. (2003), and Boone and Cunningham (1998), found similar relationships between nonfluent, nonrotational movement and anger.

Observers rated clips as sad when dancers' head, hand, and foot acceleration was low, and movements were bent forward, fluid, and complex. This pattern is opposite to the results for the Happiness rating, further indicating consistent rating behavior of the participants. This result pattern overall suggests that observers rated a video as sad when movements were smooth and of low activity (low acceleration). The results are somewhat in line with Burger, Saarikallio et al. (2013), particularly when comparing the correlations for acceleration and fluidity (same direction, but not significant). Again surprising is the result for complexity, as one would expect sad movements to be less complex (e.g., Camurri et al., 2003; Dahl & Friberg, 2007), which would fit to the negative correlations with the acceleration features. However, it again could be due to the dancer and stimulus reduction, choice of one specific stimulus representing sadness, and the low variability in the complexity values.

Observers associated tenderness with dancers that covered a large area, probably dancers stepping or walking. This result is to some extent in line with Burger, Saarikallio et al. (2013). In this previous analysis, tender music was embodied with a forward-bent body, low acceleration of the head and hands, and fluid movement. In the current analysis, all these features correlated in the same direction, although not significantly. It could have been that observers focused primarily on the space used by the dancers and only marginally paid attention to other features.

In order to connect the current results to emotion models, we used multidimensional scaling to locate the emotions in a two-dimensional space. These locations match up very well with locations suggested by the Burger, Saarikallio et al. (2013) study that displayed the emotion audio-ratings based on polar angles in a two-dimensional space. Furthermore, the solution fits well with the locations of discrete emotions in the circumplex model of emotions (Russell, 1980), indicating high consistency and unambiguousness in how observers perceived emotional expression in music-induced dance movements. Compared to Russell's solution, the locations of all four emotion ratings appear amplified on the arousal axis. This could suggest that arousal or activity is perceptually more salient in the movement than valence. Sadness and tenderness were located closer to each other than to anger or happiness, which could explain the confusion between sadness and tenderness.

This study gathered only observers' ratings, which could be conceptualized as a conscious, overt decision of their underlying perception, expressed in and potentially restricted to given categories (rating scales) provided by the researcher. Thus, adding physiological measurements, such as galvanic skin conductance, could reveal insights into observers' own emotional states and provide access to a covert perception of emotional content. Furthermore, eye-tracking technology could be used to measure where observers visually focused when making the rating decision. This might yield a more precise identification of movement features and body parts being characteristic for expressing emotions in dance movement. Additionally, in future studies, researchers could ask observers to name specific body parts or movement characteristics they paid attention to and found relevant for providing the rating.

This study involved only video excerpts without sound, as we were interested in the communication of emotional expression solely via movement. However, it would likewise be interesting to investigate audio–video stimuli in studying the multimodal integration that occurs when audio presents a second means of communicating emotional expression. Such research could involve both congruent and incongruent combinations (i.e., clips that show either a matching combination of stimulus/emotion and exhibited movements or a mismatched combination of both). Interesting questions that could arise, then, are whether audio–video clips would be evaluated differently than when audio is not present, whether different movement features would turn out significant, and whether one of the domains would influence the perception of observers more than the other. Moreover, we used only four stimuli—one per emotion (resulting in 16 clips overall)—to keep the experiment sufficiently short. However, future study could include more or different stimuli to validate further these results.

It could be argued that time-shifting the movement data could have changed some of the emotional characteristics of the dance movements. However, taken that the time-shifting was relatively moderate (7% for the happy and sad stimulus, 13% for the angry one), we assume that these changes did not affect emotion perception. Furthermore, body movement could have been impacted not only by the emotional connotation of the music, but also by the underlying musical characteristics of the chosen stimuli, and dancers could have moved simply in relation to the musical characteristics without taking emotion into account, particularly because they were instructed to move only as they felt natural. However, given that emotional connotations arise from certain combinations of musical characteristics (e.g., Grekow, 2018; Laurier et al., 2009; Saari, Eerola, & Lartillot, 2011), those cannot be disentangled fully, which is evident also from the high prediction rate of music emotion recognition models that predict perceived emotion from musical/audio features (e.g., Kim et al., 2010; Vempala & Russo, 2018). That the movement analysis presented in Burger, Saarikallio et al. (2013) aligns well with the perceptual results presented here, we feel confident that the dancers were able to convey emotions in their movement. Furthermore, the approach of asking people to dance without a concrete emotional instruction is a more ecologically valid approach than asking them, for instance, to portray specific emotions, and thus such an approach might say more about spontaneous human behavior.

The perceptual data collection involved a larger percentage of female than male observers. It could be that women are better at recognizing emotions in others, as research suggests that women tend to be higher in empathy than men are (e.g., Lennon & Eisenberg, 1987; Schieman & van Gundy, 2000). However, due to the skewed gender distribution in the data set, we did not include an analysis of gender here. Thus, future data collections could aim at a more balanced gender distribution to investigate whether gender affects emotion recognition in dance movement. Furthermore, we did not analyze our data with respect to music or dance expertise. Although previous research, for instance, suggests that professional musicians rate emotions with higher accuracy (Castro & Lima, 2014), our sample did not include professional musicians or dancers. Thus, further research could investigate the influence of professional expertise by specific participant recruiting.

The current study used animations in two dimensions that might have lacked characteristics visible when seen with more depth information (i.e., using three-dimensional projection techniques or presenting the two-dimensional animations from various angles). Future perceptual studies could therefore include animations with more depth information or use, instead of stick figure, a more naturalistic-looking avatar to provide an increased set of human features.

This study used animations of dancers who had been instructed to dance or move as it felt natural to them. A future motion-capture study could explicitly instruct dancers to portray the emotion they perceive or feel in the music or portray an emotion specified by the experimenter. Additionally, the musical stimuli could be selected based on their emotional characters to express the target emotions more clearly than the current set of stimuli. This might provide a more controlled, coherent, and systematic dataset, containing more emotion-stereotypical movement, which might be more readily perceivable by observers. Such stimuli might be useful even for target groups having challenges understanding and expressing emotions (e.g., people on the alexithymia or autism spectrum).

The present study provided profound insights into the perception of dance movements, establishing links between kinematics of music-induced movement responses and the evaluations of independent observers regarding emotions expressed in these movements (visual domain). This investigation follows up a motion-capture study (Burger, Saarikallio et al., 2013) that connected kinematics of music-induced movement with emotional characteristics of music (auditory domain). Based on the current and 2013 studies, we have found significant congruence in auditory and visual perception of emotions in music-related activities. Significant kinematic profiles for each emotion indicate commonalities and differences between both studies, providing insights into how emotional expression of music-induced movement could be conveyed and understood through auditory and visual channels.

IMPLICATIONS FOR RESEARCH, APPLICATION, OR POLICY

This investigation forms a contribution to the theory of embodied music cognition and provides support for Leman's (2007) notion of empathy. The results additionally provide insights into emotion perception of movement as a multimodal phenomenon that not only contributes to dance and music contexts, but also to understanding expression and expressivity in general human movement. Furthermore, this research could have implications on brain functionality studies, in particular to notions regarding the common coding theory (Prinz, 1984) that claims that perceiving an action activates brain areas associated with producing the action and vice versa. Specifically, our research could be viewed as support for arguments that the common coding theory would apply not only to overt actions but also to covert capacities like emotion, suggesting observers understand others' emotions in movements by internally mirroring the movement themselves. In addition to these theoretical implications, our research might have implications on practical level, for instance in a (music) therapy context, using music and movement to express and better understand one's own emotions.

REFERENCES

Balkwill, L.-L., & Thompson, W. F. (1999). A cross-cultural investigation of the perception of emotion in music: Psychophysical and cultural cues. *Music Perception, 17,* 43–64.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B (Methodological), 57,* 289–300.

Boone, R. T., & Cunningham, J. G. (1998). Children's decoding of emotion in expressive body movement: The development of cue attunement. *Developmental Psychology, 34*, 1007–1016.

Burger, B., & Bresin, R. (2010). Communication of musical expression by means of mobile robot gestures. *Journal on Multimodal User Interfaces, 3*, 109–118.

Burger, B., Saarikallio, S., Luck, G., Thompson, M. R., & Toiviainen, P. (2013). Relationships between perceived emotions in music and music-induced movement. *Music Perception, 30*, 517–533.

Burger, B., Thompson, M. R., Luck, G., Saarikallio, S., & Toiviainen, P. (2013). Influences of rhythm- and timbre-related musical features on characteristics of music-induced movement. *Frontiers in Psychology, 4,* Article 183. https://doi.org/10.3389/fpsyg.2013.00183

Burger, B., & Toiviainen, P. (2013). MoCap Toolbox: A Matlab toolbox for computational analysis of movement data. In R. Bresin (Ed.), *Proceedings of the 10th Sound and Music Computing Conference* (pp. 172–178). Stockholm, Sweden: KTH Royal Institute of Technology.

Camurri, A., Lagerlöf, I., & Volpe, G. (2003). Recognizing emotion from dance movement: Comparison of spectator recognition and automated techniques. *International Journal of Human-Computer Studies, 59*, 213–225.

Castellano, G., Villalba, S. D., & Camurri, A. (2007). Recognising human emotions from body movement and gesture dynamics. In A. Paiva, R. Prada, & R. W. Picard (Eds.), *Proceedings of the 2nd International Conference in Affective Computing and Intelligent Interaction* (ACII; pp. 71–82). Berlin, Germany: Springer.

Castro, S. L., & Lima, C. F. (2014). Age and musical expertise influence emotion recognition in music. *Music Perception, 32*, 125–142.

Coulson, M. (2004). Attributing emotion to static body pictures: Recognition accuracy, confusions, and viewpoint dependence. *Journal of Nonverbal Behavior, 28,* 117–139.

Dahl, S., & Friberg, A. (2007). Visual perception of expressiveness in musicians' body movements. *Music Perception, 24,* 433–454.

Darwin, C. (1872). *The expression of the emotions in man and animals.* London, UK: John Murray.

De Meijer, M. (1989). The contribution of general features of body movement to the attribution of emotions. *Journal of Nonverbal Behavior, 13,* 247–268.

Eerola, T., & Vuoskoski, J. K. (2011). A comparison of the discrete and dimensional models of emotion in music. *Psychology of Music, 39,* 18–49.

Gabrielsson, A. (2010). Strong experiences with music. In P. N. Juslin & J. A. Sloboda (Eds.), *Handbook of music and emotion* (pp. 547–574). Oxford. UK: Oxford University Press.

Grekow, J. (2018). Audio features dedicated to the detection and tracking of arousal and valence in musical compositions. *Journal of Information and Telecommunication, 2*(3), 322–333.

Gross, M. M., Crane, E. A., & Fredrickson, B. L. (2010). Methodology for assessing bodily expression of emotion. *Journal of Nonverbal Behavior, 34*(4), 223–248.

Halovic, S., & Kroos, C. (2018). Not all is noticed: Kinematic cues of emotion-specific gait. *Human Movement Science, 57,* 478–488.

Ilie, G., & Thompson, W. F. (2006). A comparison of acoustic cues in music and speech for three dimensions of affect. *Music Perception, 23,* 319–330.

In Flames (2006). Scream. On *Come clarity* [CD]. Donzdorf, Germany: Nuclear Blast.

Juslin, P. N., & Sloboda, J. A. (Eds.). (2010). *Handbook of music and emotion.* Oxford, UK: Oxford University Press.

Juslin, P. N., & Västfjäll, D. (2008). Emotional responses to music: The need to consider underlying mechanisms. *Behavioral and Brain Sciences, 31,* 559–575.

Keller, P. E., & Rieger, M. (2009). Musical movement and synchronization. *Music Perception, 26,* 397–400.

Kim, Y. E., Schmidt, E. M., Migneco, R., Morton B. G., Richardson, P., Scott, J., Speck, J. A., & Turnbull, D. (2010). Music emotion recognition: A state of the art review. In *Proceedings of the 11th International Society for Music Information Retrieval Conference* (ISMIR 2010; pp. 255–266). Utrecht, the Netherlands: International Society for Music Information Retrieval.

Krumhansl, C. L. (2002). Music: A link between cognition and emotion. *Current Directions in Psychological Science, 11,* 45–50.

Kurkela, J. (2005). Hetki hiljaa. On *Hetki hiljaa* [CD]. New York, NY, USA: Warner Music Group.

Laurier, C., Lartillot, O., Eerola, T., & Toiviainen, P. (2009). Exploring relationships between audio features and emotion in music. In J. Louhivuori, T. Eerola, S. Saarikallio, T. Himberg, & P.-S. Eerola (Eds.), *Proceedings of the 7th Triennial Conference of European Society for the Cognitive Sciences of Music* (ESCOM 2009; pp. 260–264). Jyväskylä, Finland: University of Jyväskylä.

Leman, M. (2007). *Embodied music cognition and mediation technology.* Cambridge, MA, USA: MIT Press.

Leman, M., & Godøy, R. I. (2010). Why study musical gesture? In R. I. Godøy & M. Leman (Eds.), *Musical gestures: Sound, movement, and meaning* (pp. 3–11). New York, NY, USA: Routledge.

Lennon, R., & Eisenberg, N. (1987). Gender and age differences in empathy and sympathy. In N. Eisenberg & J. Strayer (Eds.), *Empathy and its development* (pp. 195–217). Cambridge, UK: Cambridge University Press.

Lesaffre, M., De Voogdt, L., Leman, M., De Baets, B., De Meyer, H., & Martens, J.-P. (2008). How potential users of music search and retrieval systems describe the semantic quality of music. *Journal of the American Society for Information Science and Technology, 59,* 695–707.

Luck, G., Saarikallio, S., Burger, B., Thompson, M. R., & Toiviainen, P. (2010). Effects of the Big Five and musical genre on music-induced movement. *Journal of Research in Personality, 44,* 714–720.

Pollick, F. E., Paterson, H. M., Bruderlin, A., & Sanford, A. J. (2001). Perceiving affect from arm movement. *Cognition, 82,* B51–B61.

Prinz, W. (1984). Modes of linkage between perception and action. In W. Prinz & A.-F. Sanders (Eds.), *Cognition and motor processes* (pp. 185–193). Berlin, Germany: Springer.

Ripplingtons, The (1992). Weekend in Monaco. On *Weekend in Monaco* [CD]. New York, NY, USA: GRP Records.

Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology, 39,* 1161–1178.

Saari, P., Eerola, T., & Lartillot, O. (2011). Generalizability and simplicity as criteria in feature selection: Application to mood classification in music. *IEEE Transactions in on Audio, Speech, and Language Processing, 19,* 1802–1812.

Saarikallio, S. (2011). Music as emotional self-regulation throughout adulthood. *Psychology of Music, 39,* 307–327.

Saarikallio, S., Luck, G., Burger, B., Thompson, M. R., & Toiviainen, P. (2013). Dance moves reflect current affective state illustrative of approach-avoidance motivation. *Psychology of Aesthetics, Creativity, and the Arts, 7*(3), 296–305.

Savitzky, A., & Golay, M. J. E. (1964). Smoothing and differentiation of data by simplified least squares procedures. *Analytical Chemistry, 36,* 1627–1639.

Schieman, S., & van Gundy, K. (2000). The personal and social links between age and self-reported empathy. *Social Psychology Quarterly, 63,* 152–174.

Schlossberg, H. (1954). Three dimensions of emotion. *Psychological Review, 61,* 81–88.

Tsay, C.-J. (2013). Sight over sound in the judgment of music performance. *Proceedings of the National Academy of Science of the United States of America, 110,* 14580–14585.

Van Dyck, E., Burger, B., & Orlandatou, K. (2017). The communication of emotions in dance. In M. Lesaffre, P.-J. Maes, & M. Leman (Eds.), *The Routledge companion to embodied music interaction* (pp. 122–130). London, UK: Routledge.

Van Dyck, E., Maes, P.-J., Hargreaves, J., Lesaffre, M., & Leman M. (2013). Expressing induced emotions through free dance movement. *Journal of Nonverbal Behavior, 37,* 175–190.

Van Dyck, E., Moelants, D., Demey, M., Deweppe, A., Coussement, P., & Leman, M. (2013). The impact of the bass drum on human dance movement. *Music Perception, 30*(4), 349–359.

Van Dyck, E., Vansteenkiste, P., Lenoir, M., Lesaffre, M., & Leman, M. (2014). Recognizing induced emotions of happiness and sadness from dance movement. *PLoS ONE, 9*(2), e89773.

Varela, F. J., Thompson, E., & Rosch, E. (1991). *The embodied mind: Cognitive science and human experience.* Cambridge, MA, USA: MIT Press.

Vempala, N. N., & Russo, F. A. (2018). Modeling music emotion judgments using machine learning methods. *Frontiers in Psychology, 8,* Article 2239. https://doi.org/10.3389/fpsyg.2017.02239

Vieillard, S., Peretz, I., Gosselin, N., Khalfa, S., Gagnon, L., & Bouchard, B. (2008). Happy, sad, scary and peaceful musical excerpts for research on emotions. *Cognition & Emotion, 22*(4), 720–752.

Vuoskoski, J. K., Thompson, M. R., Clarke, E. F., & Spence, C. (2014). Crossmodal interactions in the perception of expressivity in musical performance. *Attention, Perception, & Psychophysics, 76*, 591–604.

Wallbott, H. G. (1998). Bodily expression of emotion. *European Journal of Social Psychology, 28,* 879–896.

Weckl, D., & The Buddy Rich Big Band (1994). Mercy, Mercy, Mercy.On *Burning for Buddy* Vol. 1 [CD]. New York, NY, USA: Atlantic Records.

Zentner, M., Grandjean, D., & Scherer, K. R. (2008). Emotions evoked by the sound of music: Characterization, classification, and measurement. *Emotion, 8,* 494–521.

Authors’ Note

This research was supported by Academy of Finland grants (Project Numbers 272250, 274037, and 299067).

All correspondence should be addressed to
Birgitta Burger
Institute for Systematic Musicology
University of Hamburg, Germany
Neue Rabenstraße 13
20354 Hamburg
Germany
Birgitta.burger@uni-hamburg.de

Human Technology
ISSN 1795-6889
www.humantechnology.jyu.fi

HUMAN TECHNOLOGY

ISSN: 1795-6889

www.humantechnology.jyu.fi Volume 16(3), November 2020, 257–282

COMMUNICATING THROUGH ANCILLARY GESTURES: EXPLORING EFFECTS ON COPERFORMERS AND AUDIENCES

Anna Siminoski
Department of Psychology, Neuroscience & Behaviour
McMaster University
Hamilton, ON
Canada

Erica Huynh
Schulich School of Music
McGill University
Montréal, QC
and
Department of Psychology, Neuroscience & Behaviour
McMaster University
Hamilton, ON
Canada

Michael Schutz
School of the Arts and
Department of Psychology, Neuroscience & Behaviour
McMaster University
Hamilton, ON
Canada

Abstract: *Musicians make elaborate movements while performing, often using gestures that might seem extraneous. To explore these movements, we motion-captured and audio-recorded different pairings of clarinetists and pianists performing Brahms' Clarinet Sonata No. 1 with two manipulations: (a) allowing the performers full vs. no visual feedback, and (b) allowing the performers full vs. partial auditory feedback (i.e., the clarinetist could not hear the pianist). We found that observer ratings of audio–visual point-light renditions discriminated between manipulations and refined this insight through subsequent audio-alone and visual-alone experiments, providing an understanding of each modality's contribution. This novel approach of evaluating point-light displays of performances under systematically manipulated conditions provides new perspective on the ways in which ancillary gestures contribute to both performer communication and audience reception of live performances.*

Keywords: *music performance, ancillary gestures, expression, performer cohesion, audio–visual, point-light displays.*

DOI: https://doi.org/10.17011/ht/urn.202011256765

VERTAISARVIOITU KOLLEGIALT GRANSKAD PEER-REVIEWED www.tsv.fi/tunnus

INTRODUCTION

Performing musicians often move dramatically: swaying in synchronization with the music, raising their hands with a flourish, or subtly nodding to one another. Are these extraneous movements superfluous or do they influence coperformer communication or audience perception? Not all movements made by musicians are mandatory for sound production, so how are these ancillary gestures affecting performances (Wanderley, 2002)? Musical performance entails a rich exchange of nonverbal social interactions. Musicians must execute fine motor control under multimodal stimulation from the auditory, visual, and tactile domains to create a cohesive performance with another musician. Musicians must adjust their playing dynamically in order to compensate for differences in performer timing, timbre, expression, and many other collaborative aspects. Audience members simultaneously receive sound and visual information that they process to create a coherent perception of the performance.

Ancillary Gestures Shape Audience "Listening"

Previous researchers have examined which domain—visual or auditory—takes precedence when shaping viewer perception in different situations or conditions (Broughton & Stevens, 2009; Platz & Kopiez, 2012; Schutz, 2008; Thompson, Graham, & Russo, 2005; Vines, Krumhansl, Wanderley, & Levitin, 2006; Wanderley, Vines, Middleton, McKay, & Hatch, 2005). In one such study, Vines et al. (2006) investigated the cross-modal interactions of sound and vision during clarinet performances. Research participants were played auditory, visual, or both auditory and visual recordings from the performances and were asked to judge the emotional and structural content of the music. The authors found that vision could either strengthen emotional responses of the research participants when visual information was consistent with the auditory information or dampen emotional responses when sensory information did not match. They concluded that vision and sound could communicate different emotional information, but both are integrated into overall perceived emotion, thus creating an emergent experience. Conversely, Vines et al. (2006) also found that vision and sound conveyed similar structural information as indicated by participants' judgments of the phrasing. Platz and Kopiez (2012) conducted a meta-analysis of the effect audio–visual manipulations have on perceived quality, expressiveness, and preferences for music. Fifteen studies were surveyed, including Vines et al. (2006), and an effect size of d = 0.51, Cohen's d; 95% CI [0.42, 0.59], was found for the influence of the visual component. Given that this is a medium effect size, it suggests that vision is an important aspect of a musical performance.

Additional support for the influence that vision can have on an audience's evaluation of musical performance was presented by Tsay (2013): Both novice and professional musicians were more accurate at selecting the winner of a competition between highly skilled pianists when presented with only visual clips of the musicians rather than with audio alone or both audio and visual presented together. Because highly expert musicians play extremely well, auditory output would be similar across pianists. Their movements, however, were more likely to vary. Therefore, the differences in participant ratings may reflect most strongly the variability of competitors' gestures. Mehr, Scannell, and Winner (2018) expanded on the Tsay (2013) study and tested the generalizability of their findings. When using the exact stimuli from Tsay (2013), the results were replicable. However, when Mehr et al. (2018) used other stimuli, even video clips that presented a greater distinction of skill, the sight-over-sound conclusions did not hold and selecting a piano competition winner with visuals were at or below chance. This research suggests that the amount of variability and information in the auditory and visual modalities determines which is most useful in any given situation. This confusion regarding the role of visual information in evaluating musical performances illustrates challenges with understanding gestures' role in naturalistic musical performances. Additionally, it shows the value of exploring new approaches to manipulating performance conditions to clarify the precise contributions of sight and sound to musical evaluations.

Interest in the importance of visual information in the assessment and perception of musical performances is longstanding for both practical and theoretical reasons. For example, Davidson (1993) suggested that vision plays an even more important role than sound in certain musical conditions. In that study, violinists performed a musical excerpt in three different manners that varied in degree of expression—deadpan, standard, or exaggerated. Study participants rated the performances when presented with audio–visual, audio, or visual recordings, and their ratings showed that differentiating the degree of expressiveness was most accurate with vision alone. When audio was presented alone, participants had difficulty distinguishing between the expressiveness of the performances. Vuoskoski, Thompson, Clarke, and Spence (2014) used a similar design, but also created mismatching audio–visual stimuli to examine cross-modal interactions. They found that auditory and visual cues both contribute important information to the perception of expressivity in a musical performance, but visual kinematic cues may be more important depending on individual performer's success at communicating through gestures. Furthermore, they observed cross-modal interactions when sensory information could be integrated, but extreme mismatched stimuli did not show cross-modal effects. When discussing music, sound is usually the main focus. However, these studies demonstrate the importance of considering vision.

In the current study, we ran three experiments that used either audio–visual (Experiment 1), audio-only (Experiment 2), and visual-only (Experiment 3) stimuli to examine the influence of auditory and visual information on study participants' perception of musical performances. We tried to maintain as much ecological validity as possible when designing our experiments, aside from manipulating auditory and visual feedback during performer recordings. We presented participants with stimuli that preserved the musicians' original performances. For instance, we did not cross visual recordings with different audio recordings to create our stimuli; rather, we used corresponding audio–visual material. We also balanced our musician pairings, having three clarinetists perform with each of three pianists. This allowed for performers' individual variety and magnitude of movements to be presented multiple times within unique performer pairs.

Ancillary Gestures Shape Intermusician Communication and Expression

Movement in musical performances can be categorized for the purpose they serve: to produce sound, to coordinate an ensemble, or to present expressive ideas (Bishop & Goebl, 2018). The latter two categories can be grouped under the term ancillary gestures. Ancillary gestures—gestures that do not directly influence sound production on an instrument—can be thought of as a form of nonverbal communication (Dahl & Friberg, 2007; Wanderley et al., 2005). In speech, people punctuate and emphasize certain aspects of their dialogue through body language, such as hand movements or shrugging the shoulders. Body language used in speech is analogous to ancillary gestures utilized in a musical performance, both of which have the ability to convey additional information to the viewer. The current study focuses on the communicative and expressive quality of ancillary gestures when analyzing visual-cue contributions in a musical performance. This complements previous research on the degree to which performers' visual communication affects the precision of their synchronization (D'Amario, Daffern, & Bailes, 2018) by exploring whether this communication can shape audience evaluation of their movements and sound.

Communication between performers has been shown to occur through visual information in the form of head movements and body sway, both of which are types of ancillary gestures (Badino, D'Ausilio, Glowinski, Camurri, & Fadiga, 2014; Chang, Livingstone, Bosnyak, & Trainor, 2017; Volpe, D'Ausilio, Badino, Camurri, & Fadiga, 2016). Ancillary gestures also can influence audiences in the way they perceive, understand, and interpret a musical piece (Vines et al., 2006). It is apparent that gestures possess expressive content that is intrinsically recognized by audiences (Davidson, 1993). Dahl and Friberg (2007) took videos of musicians expressing various emotions when playing a piece and asked participants to rate the expressive content when presented with different views of a silent video. Viewing conditions varied across videos in the amount of the body visible in the frame. Participants correctly identified the performers' intent of conveying happiness, sadness, and anger in all viewing conditions, suggesting that movement alone is enough to impart intended emotions. Furthermore, other studies have shown that point-light displays, which present only physical movements in the form of stick figure videos, convey enough information to discern emotional intent and other salient features of a musical performance (Davidson, 1993; Schutz & Kubovy, 2009; Sevdalis & Keller, 2011, 2012; Vuoskoski et al., 2014). In the current study we used point-light display videos in the audio–visual and visual-only experiments to analyze the impact of biological motion isolated from the facial expressions, physical appearances, and other noticeable features of performers (Figure 1).

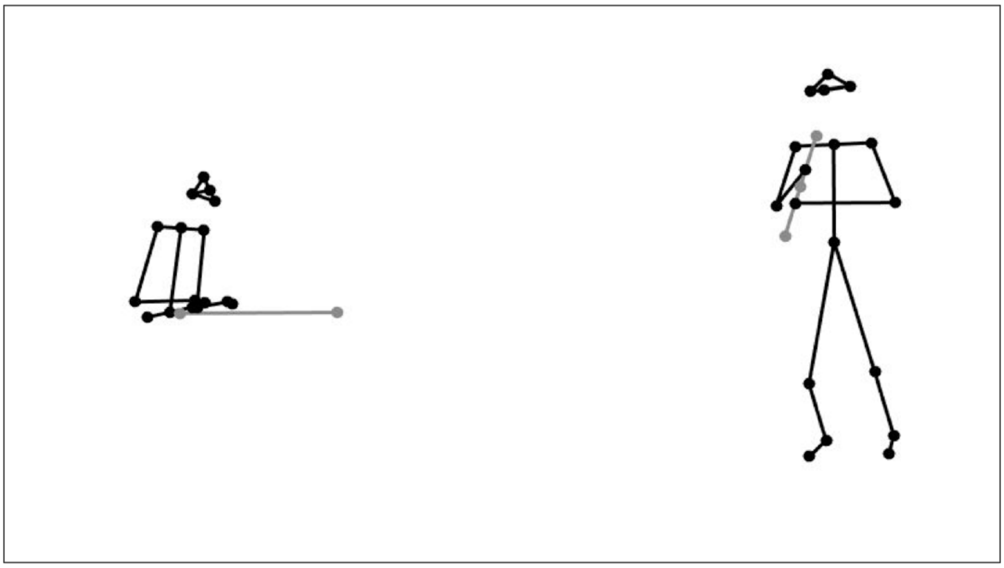


Figure 1. Screenshot of a point-light display video. Representations show the torso of the pianist and the full clarinetist in black. The clarinet appears in grey, and two points from the keyboard connected with a line provide a visual reference point for the pianist's instrument.

In this paper, we present three separate experiments that investigated audience perception of a musical performance when experiencing different types of sensory information. The three experiments were based on manipulated sensory feedback available to the performers during the recorded initial musical performances to generate our stimuli to be used in the experiments. To achieve that, we motion-captured and audio-recorded professional clarinetists and pianists performing a duet under distinct experimental conditions to see how reduced sensory feedback influenced interperformer communication abilities, as rated later by naïve participants. Musicians performed under four different conditions: (a) full auditory and visual feedback between performers; (b) no visual but full auditory feedback between performers; (c) full visual but partial auditory feedback; and (d) no visual and partial auditory feedback. In conditions (c) and (d) with partial auditory feedback, the pianist could hear both instruments, whereas the clarinetist could hear only themselves. We created point-light display videos and audio recordings from the musician data. Then, in experiments to examine research participant sensitivity to the performer manipulations, participants were exposed to audio–visual, auditory-only, or visual-only stimuli and were instructed to rate each set of stimuli on how expressive and cohesive they thought the performances were and how much they liked each performance. Participants were not informed of the conditions under which the musicians were initially recorded.

Our study had two principal aims. First, we were interested in whether participant ratings could distinguish between the different conditions under which the musicians performed—whether the musicians could fully see and hear one another. Second, we were interested in how differences in these ratings varied as a function of stimuli (i.e., audio–visual, auditory-only, or visual-only presentations of the performances). To achieve this goal we asked participants to rate three aspects of the performances—performer expression, performer cohesion, and the degree to which participants liked the performance. We predicted that visual information would play a more important role in all ratings than auditory information, based on previous literature (Davidson, 1993; Tsay, 2014; Vines et al., 2006; Vuoskoski et al., 2014). Specifically, we predicted that participant ratings of visual-only and audio–visual stimuli would be more varied across conditions than the audio-only stimuli, indicating that participants were better able to distinguish nuances in performances from visual rather than auditory information. We predicted that participants would be less able to discern differences in expressiveness and cohesion between performance conditions when listening to music without any visual information in comparison to the conditions where visuals were available. As to the initial conditions during the performances, we expected the normal performance where the musicians could both see and hear each other to yield higher ratings of expressiveness, cohesion, and likability than all other conditions because the feedback between the performers was not limited visually or auditorily, as in the other conditions that created fewer means of communication. This prediction is consistent with previous research showing that compensatory methods of communicating with a coperformer are used when visual and/or auditory feedback is diminished (Fulford, Hopkins, Seiffert, & Ginsborg, 2018; Goebl & Palmer, 2009). Goebl and Palmer (2009) found pianists' head movements to have greater synchronization as auditory feedback was reduced, suggesting that visual cuing may function as an alternate means of communication distinct from communicating through sound. Furthermore, limited sensory feedback may make musicians more focused on coperformer communication and synchronization rather than expressing emotions, which could lead to lower expression ratings.

In the next section, we outline how this performer data were collected and our process for creating the point-light stimuli. Then, in subsequent sections, we present the three experiments—each conducted with a new group of participants—based on audio–visual (Experiment 1), audio-alone (Experiment 2), and visual-alone (Experiment 3) presentations of the same performances. All studies met the criteria set by the McMaster University Research Ethics Board.

STIMULI

Performer Participants

We motion-captured and audio-recorded professional clarinetists and pianists performing under four experimental manipulations. Three professional clarinetists (1 female) and three professional pianists (2 female) from the Greater Toronto Area participated in the study for monetary compensation. The clarinetists (mean age = 51.3 years, *SD* = 16.1) had an average of 39.0 years (*SD* = 14.9) performance experience and 12.0 years (*SD* = 2.00) of lessons. They spent an average of 17.7 hours (*SD* = 4.93) a week playing the clarinet. The pianists (mean age = 41.3 years, *SD* = 3.06) had played the piano for an average of 35.0 years (*SD* = 1.00) and had an average of 31.0 years (*SD* = 3.61) of lessons. They played the piano for an average of 14.3 hours (*SD* = 8.14) a week. All musicians were highly trained, performed regularly in a number of ensembles, and most have taught at a university level. All musicians reported normal hearing and right-handedness.

The performer data collection resulted in 108 duet trials (27 per condition). From these recordings, we selected one trial for each condition and pairing, totaling 36 trials (9 per condition). We selected the trials based on the best audio quality. We then edited the audio recordings of these 36 trials using Reaper software. The motion capture data from the selected trials were cleaned using Qualisys Track Manager software.[1] We used MATLAB (Math Works, Inc.) to create point-light display videos using the Mocap Toolbox (Burger & Toiviainen, 2013; see Figure 1). In one of the pairings, a marker from the pianist was inconsistently rendered in the animations. Therefore, we eliminated that stimulus from experiments with a visual component, leaving 32 point-light display videos for Experiments 1 (audio–visual) and 3 (visual-only) respectively. The corresponding point-light display videos and audio recordings were combined using iMovie software,[2] creating 32 audio–visual clips, 32 visual-only animations, and 36 audio files approximately 40 seconds in length. We used PsychoPy v1.85[3] for all three experiments.

Materials

Clarinetists brought their personal professional-model clarinets and the pianists were provided a Roland FP-80 MIDI keyboard. A directional microphone (AKG C414 XLS) placed in a shield recorded the clarinet to a computer running Reaper software.[4] The clarinetists wore earplugs (noise reduction rating of 32 dB) along with Seinnheiser HDA 200 closed-back headphones. The piano's MIDI output was recorded using the same Reaper program as the clarinet. The pianist wore NADA QH560 open-back headphones. The audio setup allowed for auditory feedback to be adjusted throughout the experiment, depending on the condition.

A Qualisys motion-capture system[5] recorded participants' movements when performing in the LIVE Lab at McMaster University. Clarinet players wore 18 reflective markers to allow full body movements to be captured. Markers were placed bilaterally at the ankle, knee, hip, shoulder, elbow, and wrist; one marker was placed centrally on the nape of the neck; and a solid cap was worn containing four markers: one on top of the head, one centrally on the forehead, and two on the temples. The clarinet had two markers: one on the bell and another on the barrel. Piano players wore 14 reflective markers to capture the movements of the upper half of the body: bilaterally on the hip, shoulder, elbow, and wrist; one centrally on the nape of the neck; and a cap with four markers: one on top of the head, one centrally on the forehead, and two on the temples. The piano had two markers on either side of the keyboard. Eighteen Qualisys cameras recorded the infrared signals reflected off the markers.

Musicians performed an excerpt from the first movement of Brahms' *Clarinet Sonata No. 1* in f minor (bars 1–28; Brahms, 1951). This composition is from the classical–romantic period and allows performers to add emotive expression and timing fluctuations. A copy of the sheet music was provided for musicians with the editor's musical nuances indicated on the score.

Recording Procedure

Each clarinetist performed with each pianist, forming nine pairings of musicians. On the day of the experiment, each musician first played the excerpt solo three times before playing duets. Each duo performed the excerpt three times under the four conditions. To make our results easier to follow, we used abbreviations alongside condition numbers, as outlined in Figure 2. In Condition 1 (FvFa), participants could both hear and see each other, simulating a normal performance setting. In Condition 2 (NvFa), an acoustically transparent screen placed between the musicians blocked visual feedback so they could not see their coperformer. In Condition 3 (FvPa), the pianists could hear both instruments, the clarinetists could hear only themselves, but both musicians could see one another. Condition 4 (NvPa) combined the visual restriction of Condition 2 with the auditory restriction of Condition 3 such that performers were unable to see each other and the clarinetists could hear only themselves while the pianists could hear both instruments. We randomized order of the four performance conditions for each duo in a 2 (visual feedback) x 2 (auditory feedback) design. We instructed the musicians to play as if they were performing for an audience. The experiment was conducted over three consecutive days.

	Full Visual Feedback	No Visual Feedback
Full Audio Feedback	**FvFa** Condition 1	**NvFa** Condition 2
Partial Audio Feedback	**FvPa** Condition 3	**NvPa** Condition 4

Figure 2. Summary and abbreviations of the four performance conditions. In conditions involving partial auditory feedback, the clarinetist could not hear the pianist.

EXPERIMENT 1: AUDIO–VISUAL

Participants observed audio–visual renderings of the professional musicians performing a duet. After each trial, they made ratings of what they had seen and heard, providing the basis for comparison with the audio-only (Experiment 2) and visual-only (Experiment 3) presentations.

Method

Participants and Stimuli

Sixty-three undergraduate students from McMaster University completed the study for course credit or monetary compensation. We presented them with the 32 audio-visual stimuli previously described. Eight participants were excluded from analysis due to incorrectly completing the task.

Evaluation Procedure

Participants watched and listened to all audio–visual stimuli in a sound attenuated booth. The audio was presented through Seinnheiser HDA 300 closed-back headphones and video was presented on a MacBook. After each audio–video clip, participants rated the performance regarding (a) expression, (b) cohesion, and (c) likability. We defined expression as how well the musicians conveyed emotion during the performance. For cohesion ratings, participants evaluated how well the musicians worked together during the performance. Likability was defined as how much participants liked the performance. Each of these three ratings were measured using a continuous scale from 1 (*not expressive/cohesive/likeable*) to 100 (*very expressive/cohesive/likeable*). Participants were asked to consider the entire performance when assigning ratings. After completing each expression and cohesion rating, participants indicated their confidence in their rating on a 7-point Likert scale, from 1 (*not confident*) to 7 (*very confident*). Participants completed ratings for each stimulus in the following order: expression, confidence (of expression rating), cohesion, confidence (of cohesion rating), and likability. The stimuli were presented in a random order for each participant. Participants completed two practice trials consisting of audio–visual recordings of a different Brahms excerpt before starting the experimental trials.

Analyses

Due to some technical difficulties with the PsychoPy interface, 166 of the expression and cohesion ratings were not saved, representing approximately 3.1% of the 5,280 evaluations. This included 61 expression ratings (of 1,760 in total), and 105 cohesion ratings (of 1,760 in total). All participant ratings of likability were successfully recorded. To account for these missing expression and cohesion ratings, we conducted multiple imputations (MI) using the mice (van Buuren & Groothuis-Oudshoorn, 2011) and mitml (Grund, Robitzsch, & Luedtke, 2019) packages on R. MI replaces missing data plausibly based on the observations themselves and a specified statistical model. We ran MI using five imputations, which is the default, and 10 iterations, the number of iterations for which the imputed data reach considerable convergence. In other words, the convergence of imputed data does not improve beyond 10 iterations. We did not run MI on the likability ratings because none of them were missing.

Our experimental design consisted of repeated-measures, therefore we used the lme4 package (Bates, Maechler, Bolker, & Walker, 2015) to fit our imputed expression and cohesion ratings on two separate multilevel models with two predictors as fixed effects: (a) the availability of visual feedback between performers and (b) the availability of auditory feedback between performers. Using a multilevel model allowed us to pool ANOVA-like estimates with imputed data (Grund, 2018). To maintain consistency with the analyses for the imputed expression and cohesion ratings, we also fit the likability ratings on a multilevel model with the same two predictors, even though we did not impute them. Thus, our analyses assessed participants' sensitivity to performance conditions based on audio–visual information (i.e., sound and gestures). In Conditions 1 (FvFa) and 3 (FvPa) musicians could see each other, but in Conditions 2 (NvFa) and 4 (NvPa) vision was blocked. In Conditions 1 (FvFa) and 2 (NvFa) auditory feedback was intact, but in Conditions 3 (FvPa) and 4 (NvPa) auditory feedback was only partial. Because we collected data (i.e., ratings) from all levels of interest for each independent variable (e.g., visual feedback between performers: full vs. none; auditory feedback between performers: full vs. partial) and our manipulations pertaining to each level were consistent for each participant, we did not add random slopes in the model. However, to control for random individual differences, we included participant ID as a random intercept in each of our models. This informs the models that there are multiple responses from participants that vary according to their respective baseline levels. Thus, each model assumes an intercept that varies between participants. For each type of rating, we also report the estimated intraclass correlation (ICC) for the intercept-only model with no predictors (i.e., intercept model). ICC indicates the extent to which ratings from the same participant are more similar than those between participants. Then, we report the estimates of the multilevel model with the availability of visual and auditory feedback as predictors.

Results

Expression Ratings

The intercept-only model of the imputed expression ratings with no predictors estimated an ICC of 0.32, which indicates fair individual differences between participants. Consequently, participant ID was added as a random intercept to our multilevel model to control for individual differences. We fit the multilevel model to the imputed expression ratings with two repeated-measures predictors as fixed effects: the availability of visual feedback (full vs. none) and auditory feedback (full vs. partial) between performers. The estimated intercept, which represents the mean expression rating of the FvFa condition, is $\hat{\gamma}_0$=70.66, 95% CI [67.39, 73.92]. The model estimated a significant effect of having no visual feedback between performers, $\hat{\gamma}_1$=–3.78, t(28907.01[6])=–3.54, p<.001, 95% CI [–5.87, –1.68]: When performers were unable to see each other, expression ratings decreased by 3.78 units, on average, controlling for the effect of the availability of auditory feedback. A significant effect of partial auditory feedback was estimated also, $\hat{\gamma}_2$ = –3.61, t(22438.62) = –3.38, p = .001, 95% CI [–5.70, –1.52]: If the clarinetists were able to hear themselves only while the pianists heard both instruments, then on average the expression ratings decreased by 3.61 units, controlling for the availability of visual feedback. The estimated effect of the interaction between sight and sound was nonsignificant, $\hat{\gamma}_{1\times2}$ = 2.35, t(22905.67) = 1.56, p = .120, 95% CI [–0.61, 5.31].

Post hoc multiple comparisons using Tukey adjustments showed the mean expression rating of FvFa to be significantly higher than in all other conditions: NvFa ($p < .001$, 95% CI [–5.87, –1.68]), FvPa; ($p = .001$, 95% CI [–5.70, –1.52]), and NvPa ($p < .001$, 95% CI [–7.19, –2.94]; Figure 3). Table 1 shows mean expression ratings for each condition.

Cohesion Ratings

In the intercept-only model of the imputed cohesion ratings, the estimated ICC was 0.29. Given that the ICC value is quite high, we controlled for individual differences by including participant ID as a random intercept in our multilevel model. The imputed cohesion ratings were fit to the multilevel model with the same two predictors as fixed effects. The model estimated an intercept of $\hat{\gamma}_0 = 72.50$, 95% CI [69.70, 75.30], which represents the mean cohesion rating for the FvFa condition. Significant effects of both no visual feedback and partial auditory feedback were estimated: $\hat{\gamma}_1 = –2.58$, $t(1197.46) = –2.59$, $p = .010$, 95% CI [–4.54, –0.63] and $\hat{\gamma}_2 = –3.19$, $t(3176.92) = –3.42$, $p = .001$, 95% CI [–5.12, –1.26], respectively. Cohesion ratings decreased by 2.58 units on average when the clarinetist and pianist could not see each other, controlling for the effect of auditory feedback. Furthermore, cohesion ratings decreased by an average of 3.19 units when there was partial auditory feedback between performers, controlling for the effect of visual feedback. No significant effect was found for an interaction between the availability of visual and auditory feedback, $\hat{\gamma}_{1\times2} = 0.73$, $t(2677.90) = 0.53$, $p = .599$, 95% CI [–2.00, 3.47].

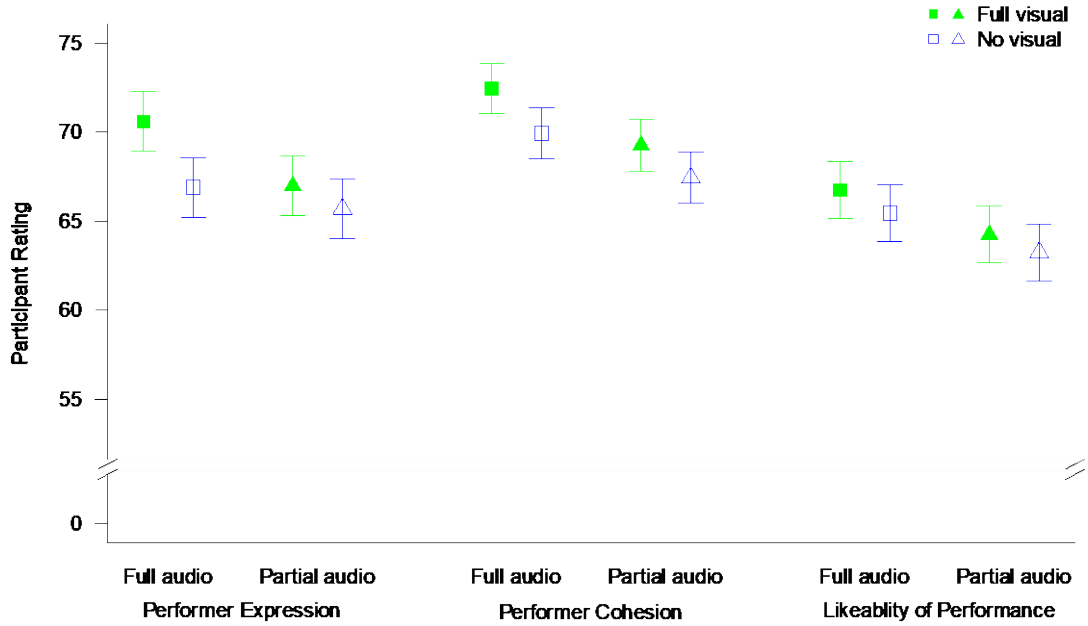


Figure 3. Participant ratings of audio–visual stimuli. Error bars represent standard error about the mean for evaluations of performer expression (left), performer cohesion (middle), and participant liking of performance (right).

Table 1. Descriptive Statistics for Audio–Visual Stimuli.

Parameter	Condition	Description	*M*	*SD*
Expression	1	FvFa	70.61	12.40
	2	NvFa	66.90	12.38
	3	FvPa	66.98	12.40
	4	NvPa	65.69	12.39
Cohesion	1	FvFa	72.46	10.56
	2	NvFa	69.94	10.56
	3	FvPa	69.27	10.66
	4	NvPa	67.44	10.54
Likability	1	FvFa	66.74	11.92
	2	NvFa	65.45	11.92
	3	FvPa	64.25	11.92
	4	NvPa	63.25	11.92

Note. $n = 55$

Post hoc comparisons using Tukey adjustments found that the mean cohesion rating for FvFa was significantly higher than NvFa (p =.010, 95% CI [–4.54, –0.63]), FvPa (p = .001, 95% CI [–5.12, –1.26]), and NvPa (p < .001, 95% CI [–6.97, –3.11]). The mean cohesion rating of NvFa was also significantly higher than that of NvPa (p=.013, *95% CI* [–4.40, –0.52]; Figure 3). No other conditions were significantly different from each other. See Table 1 for means of cohesion ratings for each condition.

Likability Ratings

Because none of the likability ratings were missing, we did not run MI on them. The intercept-only model of the likability ratings estimated an ICC of 0.33, which indicates considerable individual differences between participants. To be consistent with the analyses presented for the expression and cohesion ratings, we fit the likability ratings to a multilevel model with the same two predictors as fixed effects and included participant ID as a random intercept to control for individual differences. The estimated intercept of the model (i.e., the mean likability rating for FvFa) was $\hat{\gamma}_0 = 66.74$, 95% CI [63.57, 69.90]. There was an estimated significant effect of partial auditory feedback, $\hat{\gamma}_2 = -2.49$, $t(1702) = -2.42$, p = .016, 95% CI [–4.50, –0.48], which meant that controlling for the effect of visual feedback availability, when there was partial auditory feedback between performers, likability ratings decreased by 2.49 units on average. The effect of no visual feedback between performers and the interaction between visual and auditory feedback were nonsignificant: $\hat{\gamma}_1 = -1.29$, $t(1702) = -1.26$, p = .208, 95% CI [–3.31, 0.72], and $\hat{\gamma}_{1\times 2} = 0.29$, $t(1702) = 0.20$, p = .840, 95% CI [–2.55, 3.14], respectively.

Tukey adjusted post hoc comparisons showed FvFa to have higher likability ratings than NvPa, $p < .004$, 95% CI [–6.13, –0.85; Figure 3]. No other conditions were significantly different from each other (see Table 1 for condition means).

Discussion

Participants consistently rated performances where musicians could see and hear each other (FvFa) as being most expressive, cohesive, and likable. When musicians could not see each other and the clarinetist could not hear the pianist (NvPa), participants rated these performances as less cohesive, expressive, and likable. Although this follows our predictions and provides support for past research (Vines, Krumhansl, Wanderley, Dalca, & Levitin, 2011), the ratings decreased only by two to six units on a scale of 1 to 100. Moreover, some of the confidence intervals were close to zero, indicating that the effects were small.

Overall, participants' expression and cohesion ratings were sensitive to the lack of visual and auditory feedback between performers. Likability ratings decreased when there was partial auditory feedback between performers, meaning participants were sensitive to whether the musicians could fully hear each other or not when rating how much they liked the performance. Having audio–visual information available to participants led to distinctions between expression, cohesion, and likability ratings. Specifically, participants rated performances in which the musicians could not see one another as lower in expression and cohesion. Similarly, participants rated performances in which the clarinetist could not hear the pianist as lower in expression, cohesion, and likability.

EXPERIMENT 2: AUDIO ONLY

Experiment 2 assessed the auditory component of the audio–visual stimuli used in the first experiment. Participants listened to audio-only recordings of the clarinetists and pianists performing a duet using the same procedure and instructions as previously described.

Method

Participants

A new group of 63 undergraduate students from McMaster University completed the study for course credit. We removed eight participants from the analysis due to technical difficulties with PsychoPy or participants incorrectly completing task instructions.

Stimuli and Evaluation Procedure

Participants followed a similar procedure as Experiment 1, but instead of watching videos, they listened to the auditory component of the audio–visual stimuli described previously. As the audio-only recordings were not affected by issues with motion capture markers described previously, this experiment contained the audio from all 36 performances, rather than the 32 used for Experiments 1 (audio–visual) and 3 (visual-only). Participants completed two practice trials consisting of audio-only recordings of a different Brahms excerpt before starting the experimental trials. The experiment was programmed and run through PsychoPy v1.85 on a MacBook.

Analyses

As in Experiment 1, some of the expression and cohesion ratings were missing due to technical difficulties. Specifically, 29 of the 1,980 (1.5%) of expression ratings and 83 (4.2%) of the 1,980 cohesion ratings were missing, with all ratings of likability successfully captured. We used MI to estimate the missing expression and cohesion ratings with five imputations and 10 iterations. Similar to the analyses for the previous experiment, five imputations are the default in the mice package and the imputed data converge well with 10 iterations. Then, we fit the imputed expression and cohesion ratings into separate multilevel models with the availability of visual feedback and auditory feedback between performers as fixed effect predictors. None of the likability ratings were missing, so we did not run MI on them. We also fit the likability ratings into a multilevel model with the same predictors. The three multilevel models mentioned included participant ID as a random intercept to control for individual differences. We did not include random slopes in the models because we intend to generalize our findings to the population and ratings were collected from all levels of interest for each predictor. Similar to Experiment 1, we report the ICC for the intercept-only model of the expression, cohesion, and likability ratings in addition to the estimates of the multilevel models with the availability of visual and auditory feedback as fixed effect predictors.

Results

Expression Ratings

The intercept-only model of the imputed expression ratings estimated an ICC of 0.38, which means there are fair individual differences between participants. Thus, we added participant ID as a random intercept in our multilevel model to control for random individual differences. Additionally, we fit the imputed expression data to our multilevel model to examine if the availability of visual (full vs. none) and auditory (full vs. partial) feedback between performers can predict participants' expression ratings. The estimated intercept, which represents the mean expression rating of the FvFa condition, was $\hat{\gamma}_0$ = 72.34, 95% CI [69.40, 75.28]. The model estimated no significant effects of no visual feedback between performers, $\hat{\gamma}_1$ = –0.27, t(182956.72) = -0.33, p = .745, 95% CI [–1.91, 1.37], or partial auditory feedback between them, $\hat{\gamma}_2$ = –0.84, t(89237.19) = –1.00, p = .316, 95% CI [–2.48, 0.80]. The estimated effect of the interaction between visual and auditory feedback was nonsignificant, $\hat{\gamma}_{1\times2}$ = –0.11, t(15459.89) = –0.10, p = .925, 95% CI [–2.44, 2.22; Figure 4]. Mean expression ratings can be seen in Table 2.

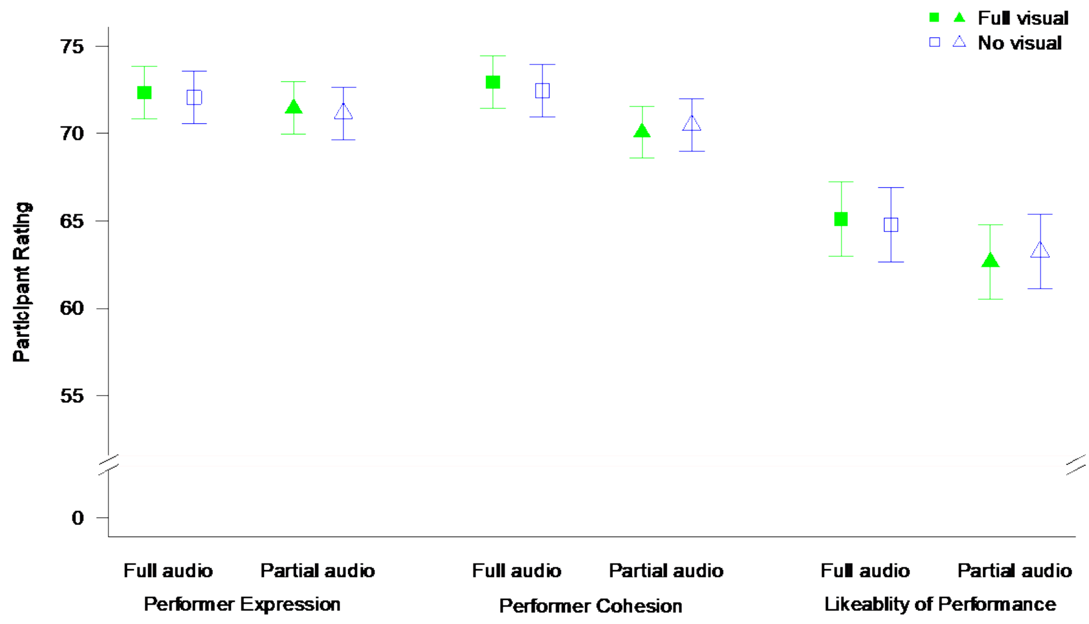


Figure 4. Participant ratings of audio-only stimuli. Error bars represent standard error about the mean for evaluations of performer expression (left), performer cohesion (middle), and participant liking of performance (right).

Table 2. Descriptive Statistics for Audio-Only Stimuli.

Parameter	Condition	Description	*M*	*SD*
Expression	1	FvFa	72.33	11.13
	2	NvFa	72.08	11.13
	3	FvPa	71.45	11.15
	4	*NvPa*	71.17	11.13
Cohesion	1	FvFa	72.95	11.09
	2	NvFa	72.46	11.10
	3	FvPa	70.07	11.07
	4	NvPa	70.48	11.09
Likability	1	FvFa	65.11	15.80
	2	NvFa	64.78	15.80
	3	FvPa	62.67	15.80
	4	NvPa	63.25	15.80

Note. *n* = 55

Cohesion Ratings

The ICC of the intercept-only model of the imputed cohesion ratings was estimated as 0.31. Thus, we added participant ID as a random intercept in our multilevel model to control for random individual differences. The results of fitting the multilevel model to the imputed cohesion ratings with the same two predictors revealed an estimated intercept (i.e., mean cohesion rating for FvFa) of $\hat{\gamma}_0$ = 72.98, 95% CI [70.03, 75.93]. The estimated effect of partial auditory feedback between performers was significant, $\hat{\gamma}_2$ = –2.87, *t*(10063.63) = –3.06, *p* = .001, 95% CI [–4.71, –1.03]. In other words, when the clarinetists could hear themselves only and the pianists could hear both instruments, cohesion ratings decreased by an average 2.87 units, controlling for the effect of the availability of visual feedback. The estimated effect of no visual feedback was nonsignificant, $\hat{\gamma}_1$ = –0.52, *t*(1298.83) = –0.54, *p* = .589, 95% CI [–2.39, 1.36]. The estimated effect of the interaction between visual and auditory feedback was nonsignificant also, $\hat{\gamma}_{1\times2}$ = 0.91, *t*(43570.08) = 0.69, *p* = .489, 95% CI [–1.68, 3.51].

We conducted post hoc comparisons using Tukey adjustments, which found that cohesion ratings were higher for FvFa in comparison to FvPa, *p* = .002, 95% CI [–4.71, –1.03], and NvPa, *p* = .010, 95% CI [–4.36, –0.59]. Moreover, cohesion ratings for NvFa were significantly higher than those for FvPa, *p* = .013, 95% CI [–4.21, –0.50], and NvPa, *p* = .038, 95% CI [–3.80, –0.11; Figure 4]. See Table 2 for mean cohesion ratings for each condition.

Likability Ratings

The intercept-only model of the likability ratings with no predictors revealed an estimated ICC of 0.51. This indicates that ratings from the same participant were more similar than those between participants. To control for these individual differences, we added participant ID as a random intercept in our multilevel model. To be consistent with the analyses conducted for the expression and cohesion ratings, our multilevel model for the likability ratings included the same two repeated-measures predictors as those in the multilevel models of the expression and cohesion ratings. The model revealed an estimated intercept of $\hat{\gamma}_0$ = 65.11, 95% CI [60.91, 69.31] (i.e., the mean likability rating for FvFa). A significant estimated effect of partial auditory feedback was revealed, $\hat{\gamma}_2$ = –2.44, *t*(1922) = –2.60, *p* = .010, 95% CI [–4.28, –0.60], indicating an average decrease of 2.60 units in likability ratings when there was partial auditory feedback between performers, controlling for the effect of visual feedback. The model did not estimate significant effects of no visual feedback, ($\hat{\gamma}_1$ = –0.33, *t*(1922) = –0.35, *p* = .728, 95% CI [–2.17, 1.51], or an interaction between auditory and visual feedback $\hat{\gamma}_{1\times2}$ = 0.91, *t*(1922) = 0.68, *p* = .496, 95% CI [–1.70, 3.51]. With Tukey-corrected post hoc comparisons, we found that the likability ratings were significantly higher for FvFa than for FvPa, *p* = .046, 95% CI [–4.85, –0.03; Figure 4]. Table 2 shows the means for each condition.

Discussion

Participants' mean ratings for expression, cohesion, and likability were all less variable across each condition compared to the same ratings of the audio–visual stimuli. This is consistent with our prediction that our performance condition manipulations would not be easily detected by participants presented with audio-only stimuli. Past research has shown that nonmusicians and musicians have a difficult time perceiving differences between musical performances based on the auditory component alone (Davidson, 1993; Tsay, 2013; Vines et al., 2011). Experiment 2 provides further support for this concept.

We also observed an effect of the availability of auditory feedback between performers for cohesion and likability ratings, indicating that participant ratings were sensitive to whether or not musicians could fully hear each other during the initial recording sessions. That said, several of the ends of the 95% confidence intervals approached zero (FvFa vs. NvPa, NvFa vs. FvPa, and NvFa vs. NvPa for the cohesion ratings; FvFa vs. FvPa for the likability ratings), which suggest small effects. We conclude that our performance manipulations had minimal effect on our performers' acoustic output, and that listeners were not sensitive to them. This suggests movements helpful for coordination did not lead to differences in their resultant sound, at least among this population. It remains unclear, however, whether greater musical training would help draw attention to these differences—a topic that could provide an interesting avenue for future research.

EXPERIMENT 3: VISUAL ONLY

In Experiment 3, participants rated point-light display videos without sound. This clarified participants' ability to differentiate expression, cohesion, and likability between performer conditions on the basis of movement differences that are by definition ancillary (i.e., they led to no differences in ratings in Experiment 2).

Method

Participants

A new group of 63 undergraduates from McMaster University participated in the experiment for course credit. Eight participants were excluded due to technical problems with PsychoPy or incorrectly completing the task.

Stimuli and Evaluation Procedure

Participants followed a similar procedure as in Experiment 1, but saw only the visual component of the audio–visual stimuli described previously. Therefore participants watched 32 point-light display videos without sound. Before the experimental trials began, participants heard an audio recording of the musical excerpt that was to be performed by the musicians in the point-light displays. For Experiment 3, participants were instructed to focus on the movements of performers when rating videos. We defined expression as how well the performers used their body movements to convey emotion. Cohesion was defined as how well the performers used their body movements to work together during the performance. Participants completed two practice trials consisting of silent point-light display videos of a different Brahms excerpt. The experiment was programmed and run on PyschoPy v1.85.

Analyses

Due to technical difficulties, 43 of 1,760 (2.3%) expression ratings and 88 of 1,760 (5%) of cohesion ratings were missing. All 1,760 ratings of likability were successfully captured. Consequently, we ran MI (five imputations, 10 iterations) to estimate the expression and cohesion ratings that were missing. To pool ANOVA-like estimates, we fit the imputed expression ratings and imputed cohesion ratings into separate multilevel models. The repeated-measures predictors for each multilevel model were the same as in the previous two experiments: the availability of visual feedback (full vs. none) and the availability of auditory feedback (full vs. partial) between performers, which were specified in the model as fixed effects. Although the likability ratings did not have missing data and we did not run MI on them, we maintained consistency in the analyses by still fitting them to a multilevel model with the same two predictors. In each multilevel model, we did not specify any random slopes. Similar to Experiments 1 and 2, we collected data from every level of interest for each independent variable and we intended to generalize our findings to the population. Participant ID, however, was included in our multilevel models as a random intercept. The ICCs of the intercept-only models are quite high, so adding participant ID as a random intercept controls for individual differences across participants. Thus, we report the ICCs for each intercept-only model and the findings for the multilevel models with our two predictors of interest.

Results

Expression Ratings

When we fit the imputed expression ratings to an intercept-only model with no predictors, we have an estimated ICC of 0.27. To control for individual differences between participants, we included participant ID as a random intercept in our multilevel models. Fitting the imputed expression data to the model with the availability of visual (full vs. none) and auditory (full vs. partial) feedback between performers as predictors revealed an estimated intercept of $\hat{\gamma}_0 = 65.05$, 95% CI [61.73, 68.37], which represents the mean expression rating in the FvPa performance setting. The estimated effect of no visual feedback between performers was significant, $\hat{\gamma}_1 = -5.23$, $t(6735.93) = -4.41$, $p < .001$, 95% CI [–7.55, –2.90]: When performers were unable to see each other, perceived expression decreased by an average of 4.41 units, controlling for the effect of the availability of auditory feedback between the performers. The estimated effect of partial auditory feedback and the estimated interaction effect between visual and auditory feedback were nonsignificant, $\hat{\gamma}_2 = -2.28$, $t(48770.83) = -4.41$, $p = .053$, 95% CI [–4.59, 0.02], and $\hat{\gamma}_{1\times2} = -2.76$, $t(55674.50) = -1.66$, $p = .097$, 95% CI [–6.02, 0.50], respectively (Figure 5). See Table 3 for mean expression ratings for each condition.

Post hoc comparisons were calculated on expression ratings between the four conditions using Tukey adjustments. Expression ratings were significantly higher for FvFa, $p < .001$, 95% CI [7.94, 12.60], NvFa, $p < .001$), 95% CI [2.72, 7.36], and FvPa, $p < .001$, 95% CI [5.67, 10.30], compared to NvPa. Moreover, expression ratings in FvFa and FvPa performance settings both were higher than those of NvFa: $p < .001$, 95% CI [2.90, 7.55] and $p = .013$, 95% CI [0.62, 5.27], respectively.

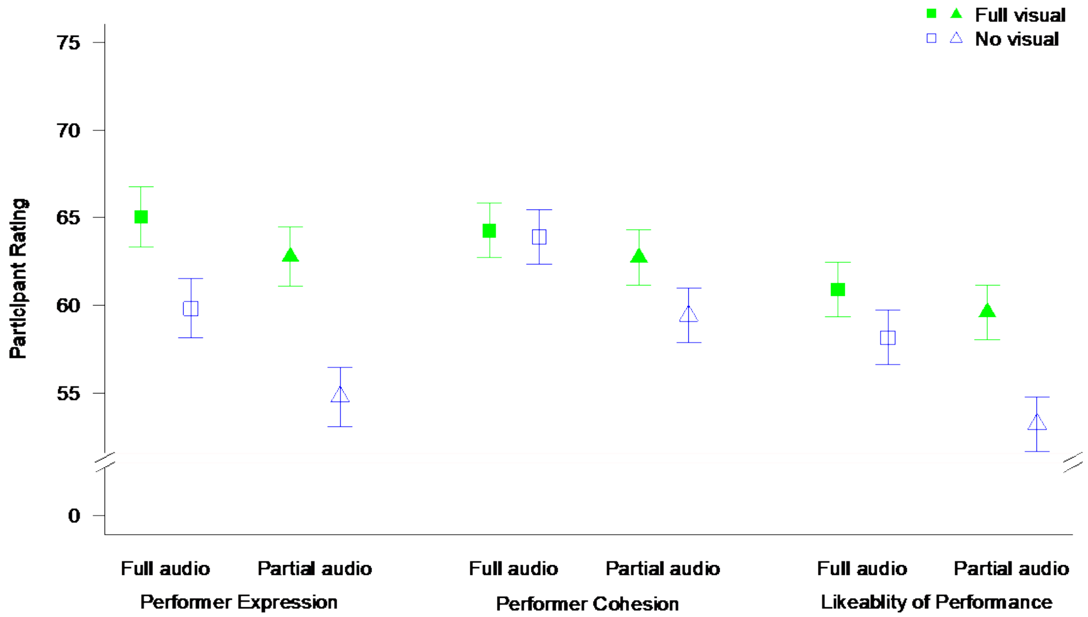


Figure 5. Participant ratings of visual-only stimuli. Error bars represent standard error about the mean for evaluations of performer expression (left), performer cohesion (middle), and participant liking of performance (right).

Table 3. Descriptive Statistics for Visual-Only Stimuli.

Parameter	Condition	Description	M	SD
Expression	1	FvFa	65.05	12.56
	2	NvFa	59.83	12.56
	3	FvPa	62.77	12.56
	4	NvPa	54.79	12.56
Cohesion	1	FvFa	64.27	11.54
	2	NvFa	63.91	11.51
	3	FvPa	62.74	11.61
	4	NvPa	59.42	11.50
Likability	1	FvFa	60.92	11.58
	2	NvFa	58.17	11.58
	3	FvPa	59.61	11.58
	4	NvPa	53.23	11.58

Note. *n* = 55

Cohesion Ratings

The intercept-only model for the imputed cohesion ratings with no predictors estimated an ICC of 0.23. Results from fitting the model to the imputed cohesion ratings with the same two predictors as fixed effect and participant ID as a random intercept revealed an intercept estimate of $\hat{\gamma}_0$ = 64.27, 95% CI [61.22, 67.32], which is the mean cohesion rating for the FvFa performance setting. The model estimated no significant effects of having no visual feedback between performers, $\hat{\gamma}_1$ = –0.36, *t*(742.83) = –0.29, *p* = .772, 95% CI [–2.78, 2.07], or having partial auditory feedback between performers, $\hat{\gamma}_2$ = –1.53, *t*(85345.58) = –1.28, *p* = .199, 95% CI [–3.87, 0.81], and no significant interaction effect between visual and auditory feedback, $\hat{\gamma}_{1\times2}$ = –2.96, *t*(22592.65) = –1.75, *p* = .080, 95% CI [–6.28, 0.35]; Figure 5). See Table 3 for mean cohesion ratings for each condition.

Likability Ratings

The intercept-only model of the likability ratings revealed an estimated ICC of 0.26. Consequently, we added participant ID as a random intercept in our multilevel model to control for considerable individual differences as estimated by the ICC. Fitting the likability ratings to the multilevel model with the same two predictors revealed an estimated intercept (i.e., mean likability rating in FvFa) of $\hat{\gamma}_0$ = 60.92, 95% CI [57.84, 63.99]. Significant estimates were found for the effect of no visual feedback between performers, $\hat{\gamma}_1$ = –2.75, *t*(1702) = –2.46, *p* = .014, 95% CI [–4.93, –0.56]. This means that likability ratings decreased by an average of 2.75 units when performers could not see each other while controlling for the availability of auditory feedback. There was also a significant interaction effect between visual and auditory feedback, $\hat{\gamma}_{1\times2}$ = –3.63, *t*(1702) = –2.30, *p* = .021, 95% CI [–6.72, –0.54]. The estimated effect of partial auditory feedback between performers, however, was nonsignificant, $\hat{\gamma}_2$ = –1.31, *t*(1702) = –1.17, *p* = .241, 95% CI [–3.49, 0.88] (see Figure 5).

Post hoc pairwise comparisons with Tukey adjustments were conducted on likability ratings between each condition to inspect the significant interaction effect further. The simple main effect of the lack of visual feedback between performers was significant when there was only partial auditory feedback between the performers: Likability ratings were significantly higher when performers could see each other than when they could not, holding partial auditory feedback between performers constant (i.e., FvPa vs. NvPa performance settings), *p* < .001, 95% CI [3.51, 9.25]. However, ratings did not differ depending on whether or not the performers could see each other when there was full auditory feedback between them (i.e., FvFa vs. NvFa), *p* = .066, 95% CI [–0.12, 5.61]. The simple main effect of the lack auditory feedback between performers was significant when there was also no visual feedback between the performers: When performers could not see each other, likability ratings were significantly higher when there was full auditory feedback between performers (NvFa) than when there was partial auditory feedback between them (NvPa), *p* < .001, 95% CI [2.07, 7.81]. However, likability ratings for performance settings where performers could fully or partially hear one another did not differ when they could fully see each other (i.e., FvFa vs. FvPa), *p* = .645, 95% CI [-1.562, 4.176]. Additional pairwise comparisons revealed that participants liked FvFa performances significantly more than NvPa ones, *p* < .001, 95% CI [4.82, 10.55]. See Table 3 for mean likability ratings for each condition.

Discussion

Participants in Experiment 1 (audio–visual stimuli) rated the movements of performances where musicians could see and hear each other (Condition 1) as most expressive, cohesive, and likeable. They also rated performances where musicians could not see each other, and the clarinetist could not hear the pianist (Condition 4), as being least expressive, cohesive, and likeable. Ratings from Experiment 3 (visual–only) suggest this difference is driven more by the performances' visual—rather than auditory—component. Together, these experiments are consistent with previous findings on the efficacy of point-light displays as tools for studying ancillary gestures (Dahl & Friberg, 2007; Davidson, 1993). More importantly however, they provide novel evidence for the musical implications of musician's body movements—showing these movements convey information about the conditions in which musicians performed.

Visual feedback had a main effect on participant ratings for expression and likability. This indicates that our manipulations of visual feedback (i.e., whether the musicians could both see one another or not) may have been detected by participants watching the musicians' movements in the absence of hearing performances. Although these effects were small, they reflect that participants were at least sensitive to the manipulations of visual feedback even though they were uninformed of those manipulations. This could mean that having visual information available to musicians during a performance may be more important for expression than having auditory information available. However, it could also be the case that during the no-visual feedback conditions (NvFa, NvPa), musicians were more affected by the visual manipulation since it disrupted sensory information for both performers rather than just one, as occurred during the partial auditory feedback manipulation.

We find it interesting that visual feedback and auditory feedback interacted for likability ratings. This shows that participants are sensitive to the sensory information available to performers and the subsequent effect this had on performance qualities. When we removed the musicians' visual communication channel, their performance was affected differently, depending upon whether or not their auditory communication was impaired. Overall, Experiment 3 shows that visual information, specifically ancillary gestures viewed by audiences, can play an important role in evaluating a musical performance.

GENERAL DISCUSSION

The current studies examined two levels of communication that exist in a musical performance. The first level dealt with intermusician communication. We wanted to see how the presence of auditory and visual information affected musicians' performance as perceived by an audience. The second level examined how musician duos communicate with audience members. To test these two aims, participants rated various presentations of the musical duos with different sensory modalities (i.e., audio–visual, audio-only, or visual-only stimuli). We observed what happened to participant ratings of performances resulting from a change in musicians' ability to communicate with their coperformers.

Intermusician Communication

To examine intermusician communication, we explored whether participants detected changes in musicians' gestures and audio output as a result of our manipulations in recording conditions—the degree to which musicians could see and hear one another. As participants rated audio-only stimuli consistently across all performer conditions (Experiment 2), we found no evidence that removing musicians' ability to see one another affected evaluation of their sound's cohesion. Furthermore, we found no evidence that removing clarinetists' ability to hear the pianists affected ratings of their sound. Intriguingly however, participants ratings of both audio–visual (Experiment 1) as well as visual-only (Experiment 3) stimuli suggest they were sensitive to communication between the performers.

We also found evidence that participants are sensitive to musicians' movements, as their ratings distinguished between conditions where the musicians could see one another—in the audio–visual and visual-only stimuli. When evaluating the audio-alone stimuli, only their ratings of cohesion distinguished between conditions (see Table 4). We interpret these results as reflecting changes in ancillary gestures between performer conditions, as participants noticed differences in performer movement even in cases where they did not detect differences in their auditory output. This indicates musicians' modulated movements may be serving a communicative purpose, complementing and extending past studies suggesting ancillary gestures are used mainly for expressive purposes (Teixeira, Loureiro, Wanderley, & Yehia, 2014; Teixeira, Yehia, & Loureiro, 2015).

Table 4. Summary of Results from Experiments 1, 2, and 3.

Experiment	**Parameter**		
	Expressivity	Cohesiveness	Likability
1 (audio–visual)	A V	A V	A NS
2 (audio-only)	NS NS	A NS	A NS
3 (visual-only)	A V	NS NS	NS V (I)

Note. Summary of main effects and interactions. *Expressivity*: We found significant estimated effects (but no interaction effect) for both the auditory (Fa vs. Pa) and visual (Fv vs. Nv) manipulations in the audio–visual and visual-only stimuli, but neither estimated effects (nor an interaction effect) in the audio-only condition. *Cohesion*: We found significant estimated effects for both the auditory and visual manipulations (and no interaction effect) in the audio–visual stimuli, but only a significant estimated effect of the auditory manipulation (and no interaction effect) in the audio-only stimuli. *Likability*: We found only a significant estimated effect of the auditory manipulation (but not visual manipulation, nor an interaction effect) in the audio–visual and audio-only stimuli, and only a significant estimated effect of the visual manipulation (as well as a significant interaction effect) in the visual-only stimuli.

Musician to Audience Communication

To observe musician-to-audience communication, we examined how musicians' ability to communicate among themselves affected participant ratings. When performers played under normal performance settings (full vision, full audio), participants consistently rated musicians as most expressive, cohesive, and likable, regardless of the sensory information available to participants. When performers could not see each other and the clarinetist could not hear the pianist (no vision, partial audio), participants rated musicians as least expressive, cohesive, and likable, across all experiments. It appears that musicians are affected by the experimental manipulations and this subsequently affects participant ratings.

We found participants to be more sensitive to performer manipulations when presented with audio–visual stimuli compared to audio-only and visual-only stimuli. The visual-only experiment yielded more differentiation between conditions than the audio-only stimuli. This was true for all ratings—expression, cohesion, and likability—indicating that visual information may allow for better discernment of musical differences than auditory information alone. It appears that participants' perception of expression and likability was influenced by whether or not the musicians could see each other when they watched the point-light displays without sound. On the other hand, participants' cohesion and likability ratings were influenced by whether the performers had full or partial auditory information between them when listening to the audio recordings.

Measuring expressivity, Vuoskoski et al. (2014) also found visual kinematic cues contributing more substantially to participant ratings than auditory information. Vuoskoski et al. (2014) created their stimuli using performances of two solo pianists whose natural performance movements varied greatly in style and magnitude. We addressed this limitation by using three clarinetists and three pianists, creating nine balanced pairings. The intent was that performer-dependent movement information would be repeated in different musician pairings so that potentially unique performer movements would be rated multiple times by participants. This design helped control for performer-dependent gestures that might otherwise skew results.

The use of point-light displays in the present study allowed us to conclude that observed differences in the visual-only stimuli are attributed solely to the performers' body movements. Point-light displays have been used in many experiments to study body movements since they isolate ancillary gestures from other visual influencers such as physical appearance, facial expressions, and lighting cues (Davidson, 1993; Sevdalis & Keller, 2011; Vines et al., 2006; Wanderley et al., 2005). Our study complements this field of research and confirms that point-light displays are a valuable tool for separating visual kinematic cues from the entirety of musical performances.

Differences Between Sensory Stimuli

Another interesting outcome of our study is that the removal of audio information lowered participants' ratings. Specifically, mean ratings for expression, cohesion, and likability in the visual-only experiment were lower than those in either the audio-only or audio–visual experiments. This is similar to what Vines et al. (2011) found, who attributed lower ratings to novelty of stimuli. Participants are not familiar with watching point-light display videos without sound but are familiar with listening to music alone. Even though the audio–visual stimuli contained point-light displays, the concept of the figures moving to actual sound is familiar; the novel condition is stick-figures moving in the absence of sound. It is possible that familiarity with stimuli types resulted in more enjoyment in general when sound was present, leading to higher expression, cohesion, and likability ratings. Vines et al. (2006) also found that visual information strengthens overall expressiveness of performances when musician gestures correspond to the emotion of the auditory component. Our results were consistent with that, as we found higher mean ratings for audio–visual stimuli compared to visual-only stimuli. Vuoskoski et al. (2014) attributed higher participant ratings to cross-modal interactions when visual and auditory information could be integrated in a meaningful way. In our audio–visual experiment, participants should have been able to integrate the information, theoretically leading to cross-modal interactions that led to increased ratings.

Future Investigations

These studies have some limitations that should be considered when interpreting the results. We did not fully balance the performer manipulations due to the nature of the instruments. The visual feedback was balanced in that both performers could either see each other or not, but the auditory feedback was not even. In conditions with partial auditory feedback, the clarinetist could not hear the pianist, but the pianist could always hear everything. The condition where the pianist could not hear the clarinetist was not included in the protocol as it is hard to mute an acoustic clarinet. Although an electric clarinet that could be silenced might have been used, we wanted to keep our experiment as ecologically valid as possible.

We chose to use a clarinet and piano piece in this study in order to examine how communication abilities between a soloist (the clarinetist) and collaborator (the pianist) were affected by the manipulations, and how audience perception was changed as a result, as data on this type of musical ensemble dynamic is limited in the joint action literature. Future studies could balance performer roles using piano duets, as electronic pianos are easily muted. Goebl and Palmer (2009) used piano duets to examine the role auditory feedback has on synchronizing musical parts. Pianists heard both parts, the assigned leader heard only themselves while the follower heard both parts, or both pianists only heard themselves. The authors found reduced auditory feedback led to decreased auditory output synchronization, but increased head movement synchronization between piano players. Given these findings, we could gain clarity on the current study results if we had fully balanced audio manipulations.

Another interesting avenue of investigation would be testing trained musicians as audience participants using the same experimental paradigm. In the current study, participants on average had low levels of musical training. Musicians may have a more fine-tuned perception of expression and cohesion, especially clarinet and piano players. Previous research with similar paradigms have found comparable emotion ratings between nonmusicians and musicians (Vines et al., 2011), so our expression ratings may be similar regardless of musical training. However, Vines et al. (2011) did not directly measure cohesion, and it is possible that trained musicians recognize what movements are communicative in purpose and provide different ratings than nontrained participants.

CONCLUSIONS

This study demonstrated that visual information is an important aspect of musical performance in a solo instrument–accompanist setting, both for interperformer communication and communication to the audience. Musicians change ancillary gestures depending on the sensory information available to them but are able to keep audio output consistent regardless. We have attributed changing ancillary gestures to the need for musicians to communicate with their coperformers when sensory feedback is obscured. Ancillary gestures can communicate novel information that increases an audience's sensitivity to performer expression and cohesion. Visual information may be more important than auditory information when audiences are asked to indicate distinctions between performances. Our findings strongly suggest that live music performances, where performers interact with one another and with the audience, may be more enjoyable for an audience than recordings. Live audiences are able to see and hear musicians, which adds to overall enjoyment through increased perception of expression and cohesion. Our findings also inform music pedagogy practices. Music students should be taught how to properly implement ancillary gestures in order to create the most expressive and cohesive performances possible.

IMPLICATIONS FOR THEORY AND APPLICATION

By connecting research on ancillary gestures and interpersonal synchronization, these experiments complement and extend previous work, showing how future studies exploring the complex relationship between physical gesture, interperformer coordination, and audience response could shed new light on interpersonal communication. Refining our understanding of how musicians' gestures simultaneously affect performances on numerous levels provides useful insight for musical training. Previous research on musical movements typically explores either ancillary gestures' effects on audiences or on coperformers. Although the specificity afforded by this bifurcation is helpful from a theoretical perspective, musicians' decisions regarding ancillary movements simultaneously affect both their audiences and their coperformers. By exploring these issues in tandem, our research results offer insight useful in applying such research to musical performances. This topic is timely given the recent explosion of interest in socially distanced performances, where musicians may appear together despite having recorded individual parts in isolation.

ENDNOTES

1. See http://www.qualisys.com/software/qualisys-track-manager/ for the particulars on the movement tracking software used in the study.
2. For more information on the software go to https://www.apple.com/ca/imovie/
3. The software details can be accessed at http://www.psychopy.org/
4. See http://www.reaper.fm/ for more information on the software.
5. A more detailed description of the software can be found at http://www.qualisys.com/
6. The degrees of freedom (*df*) for many of these tests appear to be very large as a result of using the mice package on R to impute missing data. The *df* values vary drastically due to the combination of the number of imputations and maximum iterations as specified in the mice function. Consequently, we ran the MI with five imputations (i.e., the default in the mice package) and 10 maximum iterations. We chose 10 iterations because the imputed data appeared to have good convergence with this value.

REFERENCES

Badino, L., D'Ausilio, A., Glowinski, D., Camurri, A., & Fadiga, L. (2014). Sensorimotor communication in professional quartets. *Neuropsychologia, 55*(1), 98–104.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1–48.

Bishop, L., & Goebl, W. (2018). Performers and an active audience: Movement in music production and perception. *Jahrbuch Musikpsychologie, 28*, 1–17.

Brahms, J. (1951). *Sonata in F minor, for clarinet and piano*, op. 120, no. 1 [Musical score]. New York, NY USA: G. Schirmer. (Original work published 1895)

Broughton, M., & Stevens, C. (2009). Music, movement and marimba: An investigation of the role of movement and gesture in communicating musical expression to an audience. *Psychology of Music, 37*(2), 137–153.

Burger, B., & Toiviainen, P. (2013). MoCap Toolbox: A Matlab toolbox for computational analysis of movement data. In R. Bresin (Ed.), *Proceedings of the Sound and Music Computing Conference 2013* (SMC 2013; pp. 172–178). Stockholm, Sweden: Logos Verlag Berlin.

Chang, A., Livingstone, S. R., Bosnyak, D. J., & Trainor, L. J. (2017). Body sway reflects leadership in joint music performance. *Proceedings of the National Academy of Sciences, 144*(21), E4134–E4141.

Dahl, S., & Friberg, A. (2007). Visual perception of expressiveness in musicians' body movements. *Music Perception, 24*(4), 433–454.

D'Amario, S., Daffern, H., & Bailes, F. (2018). Synchronization in singing duo performances: The roles of visual contact and leadership instruction. *Frontiers in Psychology, 9*, 1–17.

Davidson, J. W. (1993). Visual perception of performance manner in the movements of solo musicians. *Psychology of Music, 21*(2), 103–113.

Fulford, R., Hopkins, C., Seiffert, G., & Ginsborg, J. (2018). Reciprocal auditory attenuation affects looking behaviour and playing level but not ensemble synchrony: A psychoacoustical study of violin duos. *Musicae Scientiae, 24*(2), 168–185.

Goebl, W., & Palmer, C. (2009). Synchronization of timing and motion among performing musicians. *Music Perception, 26*(5), 427–438.

Grund, S. (2018, June 25). *Dealing with missing data in ANOVA models* [Web log post]. Retrieved from https://simongrund1.github.io/posts/anova-with-multiply-imputed-data-sets/

Grund, S., Robitzsch, A., & Luedtke, O. (2019). mitml: Tools for multiple imputation in multilevel modeling. R package version 0.3-7.

Mehr, S. A., Scannell, D. A., & Winner, E. (2018). Sight-over-sound judgments of music performances are replicable effects with limited interpretability. *Plos One, 13*(9), e0202075.

Platz, F., & Kopiez, R. (2012). What the eye listens: A meta-analysis of how audio–visual presentation enhances the appreciation of music performance. *Music Perception: An Interdisciplinary Journal, 30*(1), 71–83.

Schutz, M. (2008). Seeing music? What musicians need to know about vision. *Empirical Musicology Review, 3*(3), 83–108.

Schutz, M., & Kubovy, M. (2009). Deconstructing a musical illusion: Point-light representations capture salient properties of impact motions. *Canadian Acoustics, 37*(1), 23–28.

Sevdalis, V., & Keller, P. E. (2011). Perceiving performer identity and intended expression intensity in point-light displays of dance. *Psychological Research, 75*(5), 423–434.

Sevdalis, V., & Keller, P. E. (2012). Perceiving bodies in motion: Expression intensity, empathy, and experience. *Experimental Brain Research, 222*(4), 447–453.

Teixeira, E. C. F., Loureiro, M. A., Wanderley, M. M., & Yehia, H. C. (2014). Motion analysis of clarinet performers. *Journal of New Music Research, 44*(2), 97–111. https://doi.org/10.1080/09298215.2014.925939

Teixeira, E. C. F., Yehia, H. C., & Loureiro, M. A. (2015). Relating movement recurrence and expressive timing patterns in music performances. *The Journal of the Acoustical Society of America, 138*(3), EL212–EL216.

Thompson, W. F., Graham, P., & Russo, F. A. (2005). Seeing music performance: Visual influences on perception and experience. *Semiotica, 2005*(156), 203–227.

Tsay, C.-J. (2013). Sight over sound in the judgment of music performance. *Proceedings of the National Academy of Sciences of the United States of America, 110*(36), 14580–14585.

Tsay, C.-J. (2014). The vision heuristic: Judging music ensembles by sight alone. *Organizational Behavior and Human Decision Processes*, *124*(1), 24–33.

van Buuren, S., & Groothuis-Oudshoorn, K. (2011). mice: Multivariate imputation by chained equations in R. *Journal of Statistical Software, 45*(3), 1–67.

Vines, B. W., Krumhansl, C. L., Wanderley, M. M., Dalca, I. M., & Levitin, D. J. (2011). Music to my eyes: Cross-modal interactions in the perception of emotions in musical performance. *Cognition*, *118*(2), 157–170.

Vines, B. W., Krumhansl, C. L., Wanderley, M. M., & Levitin, D. J. (2006). Cross-modal interactions in the perception of musical performance. *Cognition, 101*(1), 80–113.

Volpe, G., D'Ausilio, A., Badino, L., Camurri, A., & Fadiga, L. (2016). Measuring social interaction in music ensembles. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *371*(1693), article 20150377. http://doi.org/10.1098/rstb.2015.0377

Vuoskoski, J. K., Thompson, M. R., Clarke, E. F., & Spence, C. (2014). Crossmodal interactions in the perception of expressivity in musical performance. *Attention, Perception, & Psychophysics, 76*(2), 591–604.

Wanderley, M. M. (2002). Quantitative analysis of non-obvious performer gestures. In I. Wachsmuth & T. Sowa (Eds.), *Gestures and sign languages in human computer interaction* (pp. 241–253). Berlin, Germany: Springer Verlag.

Wanderley, M. M., Vines, B. W., Middleton, N., McKay, C., & Hatch, W. (2005). The musical significance of clarinetists' ancillary gestures: An exploration of the field. *Journal of New Music Research, 34*(1), 97–113.

Authors' Note

We thank our research assistants, Maxwell Ng, Shanthika Ravindran, Nicole Divincenzo, Brannon Senger, Rachel Heo, Max Maglal-Lan, and Veronica Panchyshyn, for helping with this project and the LIVE Lab staff, Dave Thompson, Dan Bosnyak, and J. J. Booth, for helping with the technology. This research was supported financially through grants to Dr. Michael Schutz from the Natural Sciences and Engineering Research Council of Canada (NSERC RGPIN/386603-2010), Ontario Early Researcher Award (ER10-07-195), and the Canadian Foundation for Innovation (CFI-LOF-30101). All authors state that they have no conflicts of interest.

All correspondence should be addressed to
Anna Siminoski
McMaster University
424 Togo Salmon Hall
1280 Main Street West
Hamilton, ON, Canada, L8S 4M2
annasiminoski@gmail.com

Human Technology
ISSN 1795-6889
www.humantechnology.jyu.fi

HUMAN TECHNOLOGY

ISSN: 1795-6889

www.humantechnology.jyu.fi Volume 16(3), November 2020, 283–309

THE COMMUNICATION OF MELANCHOLY, GRIEF, AND FEAR IN DANCE WITH AND WITHOUT MUSIC

Lindsay A. Warrenburg
School of Music
The Ohio State University
USA

Lindsey Reymore
School of Music
The Ohio State University
USA

Daniel Shanahan
School of Music
The Ohio State University
USA

Abstract: *Professional dancers were video recorded dancing with the intention of expressing melancholy, grief, or fear. We used these recordings as stimuli in two studies designed to investigate the perception and sociality of melancholy, grief, and fear expressions during unimodal (dancing in silence) and multimodal (dancing to music) conditions. In Study 1, viewers rated their perceptions of social connection among the dancers in these videos. In Study 2, the same videos were coded for the amount of time that dancers spent in physical contact. Results revealed that dancers expressing grief and fear exhibited more social interactions than dancers expressing melancholy. Combined with the findings of Warrenburg (2020b, 2020c), results support the idea that—in an artistic context—grief and fear are expressed with overt emotional displays, whereas melancholy is expressed with covert emotional displays.*

Keywords: *dance, multimodality, emotional expressions, perception, sociality.*

DOI: https://doi.org/10.17011/ht/urn.202011256766

INTRODUCTION

An attunement to emotions experienced by others is essential for human well-being and survival. In general, emotion perception facilitates understanding of the emotions, actions, and sensations of others. Emotions are thought to have an evolutionary past: They function to change our behavior, and such behaviors can provide valuable information to others about the environment as well as possibly influence how others behave (Adolphs, 2017; Ekman, 1992; Keltner & Haidt, 1999; LeDoux & Hofmann, 2018). When people experience an emotion, related changes take place within their physiology and expressions. Some of these changes are common across cultures and some differ as a function of social learning, interoceptive cues, and environmental context (Barrett, 2017; Gross & Feldman Barrett, 2011; Russell, 2003; Tracy & Randles, 2011).

Historically, the study of emotion perception has emphasized one's ability to decipher others' facial expressions or vocal cues (e.g., Ekman & Friesen, 1975). Ekman and colleagues asked actors to make facial expressions of six emotions: surprise, happiness, sadness, disgust, anger, and fear. Researchers have asked people around the globe to match these facial expressions to specific emotional terms, such as sadness. The fact that people in many cultures have been successful in this task has given rise to the claim that some emotional expressions are universal (although see Barrett, 2017, for a refutation of this argument).

In the 21st century, emotion researchers tend to separate subjective feelings of emotion from emotional displays (see Warrenburg, 2020a, for a review comparing emotion theories in aesthetic and nonaesthetic scenarios). Feelings of sadness, for example, may sometimes be visible to others through behaviors such as crying and at other times may be undetectable by an observer. The act of crying may signal sadness, but also could indicate happiness (such as at weddings), fear, or relief. People also can control consciously some facial and bodily expressions, for instance, choosing to smile even when they are feeling sad. Therefore, in everyday life, there is no direct mapping between emotional behaviors or displays and the underlying affective feelings (Barrett, 2017; Russell, 2003).

In art forms such as dance and music, however, artists may rely on certain behaviors and expressions in order to convey emotions to their audiences. Some researchers believe that the emotions expressed and evoked when engaging in art, known as "aesthetic emotions" or "refined emotions," differ from "everyday" emotions (Frijda & Sundararajan, 2007; Juslin, 2013; Scherer, 2004). Characteristics of naturalistic movements—such as force, velocity, timing, and spatial orientation—can be used to express (and perceive) emotions in dance (Camurri, Lagerlöf, & Volpe, 2003; Camurri, Mazzarino, Ricchetti, Timmers, & Volpe, 2003; Van Dyck, Burger, & Orlandatou, 2019). Dances depicting grief, for example, tend to be performed at a relatively stable tempo, compared to dances expressing anger and fear, which contain more tempo changes (Camurri, Lagerlöf et al., 2003). People are able to discriminate emotions in dance performances even when the dancers use identical choreography (Camurri, Lagerlöf et al., 2003). In music, factors such as tempo, mode, dynamics, and articulation have been shown to vary among expressions of emotions such as fear, grief, and melancholy (Juslin & Sloboda, 2010; Warrenburg, 2020b).

Overt and Covert Emotional Displays

A subjective feeling state, like happiness or anger, emerges from a combination of processes, including underlying physiology and cognition, innate behaviors, learned associations and previous experiences, and the physical and social environment (Barrett, 2017). Sometimes these affective feelings are accompanied by certain facial expressions (such as smiling), vocal characteristics (such as a breaking voice), bodily movements (such as jumping), or interpersonal behaviors (such as self-isolation). The physical manifestations that accompany a subjective feeling state are called "emotional displays." Even though specific emotional displays do not correlate one-to-one with subjective feeling states, some affective states tend to be associated with easily observable displays and some with more subtle displays.

Two general types of emotional displays can accompany subjective feeling states: overt and covert (e.g., Huron, 2015, 2016). Overt displays refer to clearly communicated emotional displays, meaning that an observer usually can detect how the emotion experiencer is feeling. Emotions such as excitement or anguish, for example, often—but not always—are accompanied by facial expressions (smiling, crying), vocalizations (laughing, bawling), and bodily expressions (jumping up and down, curling into a fetal position). Covert displays, on the other hand, are difficult to detect by observers because these displays tend to be subtle or even invisible. For example, it is not easy to determine whether another person is experiencing love or jealousy unless the feelings or experiences are articulated directly.

Put succinctly, overt emotional displays can be used to elicit a response in an observer (Huron, 2016; Tomkins, 1980). After having suffered a death in the family, a grieving individual may need help or support from another person. Emotional displays such as crying can facilitate communication between the grieving person and an observer. Similarly, if a person is excited, it is often to that person's advantage to let others know that their facial and bodily expressions represent feelings of happiness and friendliness, as opposed to displays of aggression (Ohala, 1994). Overt auditory and visual displays, then, make it easier for observers to interpret the experiencer's emotions. When others respond, emotion experiencers can, in turn, more effectively regulate their affect.

Alternately, covert emotional displays typically do not result in social interactions. For example, when a person is feeling calm or jealous, obvious emotional expressions or behaviors that let observers know how the experiencer is feeling are missing. The lack of obvious emotional displays can be advantageous for these affective states. A person may wish to hide feelings of jealousy, for example, as jealousy could possibly hurt one's social standing. Thus, because minimal visual or auditory displays are associated with affective states such as melancholy, jealousy, or love, it can be difficult for an observer to perceive accurately these affective states in another person.

In line with the idea that subjective feeling states can be separated from emotional displays, some affective feelings utilize both overt and covert displays. In expressing fear, for example, a person might project widened eyes, choose to run away, or simply freeze in place. Furthermore, the relative overtness of an emotional display is not categorical but continuous: In one instance, feelings of happiness might be accompanied by a smile, but at another time could be accompanied by both a smile and gesticulations. To summarize, then, emotional feelings in everyday life that utilize overt displays are inherently social, as these expressions tend to facilitate interactions with others.

We draw on the role of overt and covert emotional behaviors for our research. The goal of the current study was to study emotional displays in an aesthetic context: We studied how dancers use overt and covert displays to express emotions to an audience, as well as how observers interpret these displays.

Melancholy, Grief, and Fear

We selected melancholy, grief, and fear as the three target emotions because they are all negatively valenced emotions. These three emotions can also be accompanied by various emotional displays and expressions.

Researchers who study sadness often distinguish a low-energy sadness (melancholy) from a high-energy sadness (grief). Among the researchers who make this distinction, grief usually is defined as a negatively valenced emotion associated with high physiological arousal, whereas melancholy usually is defined as a negatively valenced emotion associated with low physiological arousal (Andrews & Thomson, 2009; Darwin, 1872; Urban, 1988; Vingerhoets & Cornelius, 2012). We use these definitions in the current study.

The functions of melancholy and grief, as well as their corresponding subjective feeling states and physiological displays, differ significantly. People experience grief after a significant loss, due to death, loss of safety, loss of autonomy, or loss of identity (Archer, 1999; Epstein, 2019). However, researchers disagree about the evolutionary purpose of grief. On one hand, according to Archer (1999), grief is a maladaptation. Archer's argument is that it is biologically important to form close personal relationships and to experience emotions like love and trust. When one of these relationships is lost, whether due to death or another circumstance, grief is experienced as a maladaptive side effect of the lost relationship; there is no useful purpose of grief.

On the other hand, some researchers argue that the function of grief is to solicit help, compassion, comfort, and prosocial responses from others in times of need (Huron, 2015, 2016; Urban, 1988; Vingerhoets & Cornelius, 2012). This theory of grief is driven by observations that when people are in a grieving state, they often exhibit conspicuous (overt), multimodal displays of emotion, including crying (visual), wailing (auditory), and pheromone release (olfactory; Frick, 1985; Gelstein et al., 2011; Mazo, 1994; Rosenblatt, Walsh, & Jackson, 1976; Urban, 1988; Vingerhoets & Cornelius, 2012). Observers easily understand that a person is grieving and therefore are able to respond to the grieving person with compassionate or prosocial behaviors.

In contrast, melancholy is a negatively valenced emotion associated with few—or no—overt emotional displays. When a person needs to self-reflect about a failed goal, melancholy may be the emotion experienced (Ekman, 1992). This self-reflection is a solitary activity that usually does not require the assistance of other people (Huron, 2015). Melancholic individuals therefore tend not to exhibit any conspicuous displays of emotion. Rather, behaviors and expressions associated with melancholy are simply effects of low physiological arousal: A melancholic person tends to be mute and display relaxed facial expressions (Andrews & Thomson, 2009; Nesse, 1991). Accordingly, it is difficult for an observer to differentiate among people in melancholic, bored, relaxed, or sleepy states, despite the differences in valence in these states (Andrews & Thomson, 2009; Nesse, 1991).

It is important to note that in the case of a significant loss, such as the death of a marital partner, a person is likely to experience both melancholy and grief. Psychic pain tends to involve alternating periods of quiescent melancholy and louder grief. This alternating pattern of melancholy and grief constitutes the mourning cycle (Huron, 2016).

Given the theoretical and behavioral differences between melancholy and grief, we propose that comparatively overt expressions of grief will result in higher levels of perceived sociality than the comparatively covert expressions of melancholy. Dancers expressing grief may be more likely to interact with each other, increasing perceptions of prosociality. However, because melancholy tends to be a private, self-directed emotion, we expect that dancers expressing melancholy may exhibit comparatively fewer prosocial behaviors.

The third emotion examined in the current study is fear. Unlike grief, which typically is accompanied with overt emotional displays, fear can be accompanied by either overt or covert emotional displays, as indicated by the classic “fight, flight, or freeze” response. Several bodily responses to fear-induced adrenaline are involuntary and overt, such as trembling, which can affect the quality of the voice (Huron, 2015). On the other hand, some involuntary responses to fear are quite subtle and may not be noticeable to others, depending on their proximity and attention, such as pupil dilation (Leuchs, Schneider, Czisch, & Spoormaker, 2017) and shortness of breath (Milosevic & McCabe, 2015), while still others are primarily internal, including increased heart rate (Lang, Levin, Miller, & Kozak, 1983).

In some cases, overt displays of fear can provide benefits: As with grief, overt displays may encourage help from others. However, in the face of a threat, one often is advantaged by hiding expressions of fear, particularly when the threat comes from a human or animal that is able to interpret those signs. Therefore, the level of interpretability of emotional displays of fear is unclear. We included fear in the current study in order to investigate whether dancers use comparatively more covert or overt displays of fear and whether observers could discern fear relatively more easily (like grief) or less easily (like melancholy).

Multimodality

In the current study, participants were asked to rate both unimodal (dance–only) and multimodal (dance–music) videos. In multimodal situations involving music and dance, dance movements have been shown to express similar emotions as the accompanying music (Krumhansl & Schenk, 1997). For example, in prior studies, researchers asked participants to dance naturally to a series of musical samples while wearing motion capture sensors (Burger, Saarikallio, Luck, Thompson, & Toiviainen, 2013). Sad music correlated with fairly simple movements, low speeds, and little tension, whereas dancers used irregular and nonfluid movements when dancing to angry music. Multimodal presentations in aesthetic conditions, such as dance with accompanying music, are thought to facilitate emotional learning and perception, potentially leading to a higher accuracy in emotional perception than in dance–only or music–only conditions (Huron, 2016; Moreno & Mayer, 2007). With respect to this previous research, analysis of the current study will address whether multimodality enhances identification of the intended emotions in an aesthetic context.

HYPOTHESES

The current study was based on two primary goals. The first was to explore people’s accuracies in identifying expressions of three negative emotions in dance. Then we sought to determine whether the emotional expressions of the three emotions affected perceptions of sociality among the dancers.

Our first goal was to explore the relative accuracy in identifying expressions of melancholy, grief, and fear in dance. In Hypothesis 1, we proposed that emotions with overt displays, like grief, would be detectable by observers, but emotions with covert displays, like melancholy, might not be identified as easily. However, for an emotion such as fear, which can be associated with either covert or overt displays, the level of interpretability might be unclear, and consequently, we did not make a directional prediction.

> H1: Observers will be more accurate in their perception of grief than in their perception of melancholy in unimodal (dancing in silence) and multimodal (dancing to music) videos.

The second goal was to examine the relative degrees of sociality present in the performances of melancholy, grief, and fear in unimodal (dance–only) and multimodal (dance–music) conditions. As described in the Introduction, people expressing grief may exhibit more overt behaviors than people expressing melancholy, due to the theorized evolutionary function of grief to solicit compassionate or prosocial responses from other people (Huron, 2015, 2016). We quantified social behaviors expressed in uni- and multimodal dance videos using two methods. First, we used a rating task to examine the extent of sociality perceived by observers in videos expressing melancholy, grief, and fear (H2). Second, we quantified the amount of physical connection in videos of dancers expressing melancholy, grief, and fear (H3) by calculating the proportion of time that included physical touch among the dancers for each video. We did not make any a priori hypotheses about fear because of fear's mixture of both overt expressions (e.g., running away, eyes widening) and covert expressions (e.g., standing still).

> H2: Observers will perceive more sociality in videos expressing grief than in videos expressing melancholy.
>
> H3: Dancers will objectively exhibit more social behaviors when dancing to express grief than when dancing to express melancholy.

STUDY 1: PERCEPTIONS OF EMOTION AND SOCIALITY

Study 1 tested Hypotheses 1 and 2, which investigated the ability of observers to identify correctly an emotional expression and the amount of sociality perceived in its expression. We sought to determine if the outcomes are dependent on whether the target emotion utilizes overt or covert displays. Specifically, we predicted that overt displays (H1) facilitate the ability to interpret correctly an emotion in expressive dance and (H2) increase the perceived sociality among dancers.

Method

Participants

We drew on two separate sources for our participant pool. The 101 participants were either visitors to the public science museum, the Center for Science and Industry (COSI, *n* = 61) in Columbus, Ohio, or second-year music majors recruited from The Ohio State University's School of Music (*n* = 40), who received course credit for participation. At COSI, research assistants stationed outside an exhibit asked entering visitors if they would be interested in participating in a short study. Our decision to recruit from COSI in addition to the OSU participant pool was motivated by the desire to increase the diversity of the participants.

Materials

A recording session was conducted with four female members of the SYREN Modern Dance company, a professional group of dancers based in New York City. The dancers wore dark clothing and were video recorded against a neutral background. At the beginning of the session, we instructed the dancers to improvise choreography to express one of three specific emotions. The dancers were unaware of which emotion they would be expressing until approximately a minute before each of the recordings began, when the experimenters held up a sign with the name of the emotion. Dancers performed as a group using modern dance style and were instructed to maintain neutral facial expressions.

We recorded the dances for each of the three emotions six times, with each recording lasting approximately 60 s. First, dancers improvised three times without music. We recorded each improvisation without music first in order to avoid the possibility that experience with the music might affect the silent improvisation for the same emotion. For example, we first recorded three 60 s videos of the expression of melancholy without music. The dancers then listened to the musical excerpt one time. Then, we recorded three videos of the expression of melancholy with the accompanying melancholic music (Fauré, 1878/2015, track 8, 00:00-00:52). We repeated this process for the emotion of grief (Arnold & Price, 2012, track 18, 00:42-01:32) and fear (Elfman, 1992, track 5, 00:00-00:46). After all the stimuli had been recorded, one experimenter conducted an informal interview with the four SYREN dancers in a group setting in order to understand how the dancers interacted with each other and how they communicated the three emotions.

We edited the dance videos to be approximately 15 s in duration (Fauré, 1878/2015, track 8, 00:00-00:13; Arnold & Price, 2012, track 18, 01:11-01:26; Elfman, 1992, track 5, 00:09-00:25). The shorter videos contained the section of music that corresponded to previously validated excerpts that expressed the intended emotions (Eerola & Vuoskoski, 2011; Warrenburg, 2020c). The silent dance videos contained the first 15 s of dance from each video.

Procedure

Each participant viewed nine videos from the 18 recorded videos (nine silent dance videos, nine music and dance videos), selected by the randomizer function in Qualtrics. Participants therefore saw different collections of videos, some that were unimodal (dance–only) and some that were multimodal (dance–music). The participants answered an online series of six questions assessing the perception of emotional expression and sociality in the dances after each video. We also collected basic demographic information, namely age, gender, years of musical training, years of dance training, race/ethnicity, first language, and zip code.

To test the hypothesis that expressions of grief would be more accurately identified than expressions of melancholy, as well as to explore the relative interpretability of fearful expressions, viewers were asked, “What emotion do you think these dancers are expressing?” Participants selected one response from the following three options: fear, grief, or melancholy. This three-alternative forced choice (3-AFC) task mirrored the methodology in Warrenburg (2019, 2020c) for distinguishing perceived emotions in music samples. After answering the 3-AFC question, participants were asked to answer, “How intense was this emotion?” on a sliding scale from 0 (*not intense at all*) to 100 (*extremely intense*). We elected to use the word *intense* instead of the typical research term *physiological arousal* to aid in interpretability for our participants.

The aim of Hypothesis 2 was to determine whether viewers would perceive and experience more sociality in dances expressing grief than in dances expressing melancholy. We defined sociality in two ways. First, participants were asked, “How connected do you think the dancers feel to each other?” and “How connected do you feel with the dancers?” Responses were provided on a continuous sliding scale from 0 (*not at all connected*) to 100 (*extremely connected*). Second, we asked participants to complete the Inclusion of Other in the Self (IOS) scale (Aron, Aron, & Smollan, 1992; see Figure 1). The IOS scale, used primarily in the social science literature, measures social connectedness (e.g., Weinstein, Launay, Pearce, Dunbar, & Stewart 2016).[1] Here, participants responded to the prompt, “Please select the picture that best describes the relationship among the four dancers.” Finally, we measured experienced sociality through the prompt, “Please select the picture that best describes your current relationship with the video.”

Results

Overall Accuracy of Emotion Perception

The first hypothesis addressed whether observers were more accurate in their perception of grieving expressions than in their perception of melancholic expressions. Of a possible 909 responses, 102 responses contained blank values for the questions regarding which emotion the participants believed the dance was expressing and that emotion’s intensity. Across the remaining 807 responses to the dance videos (both modalities), the accuracy of identifying dances expressing melancholy was 65.4%, the accuracy of identifying dances expressing grief was 66.8%, and the accuracy of identifying dances expressing fear was 69.2% (Table 1). For each of the three emotion conditions, a binomial test was conducted to determine whether observers selected the correct emotion more often than chance (33.3%).

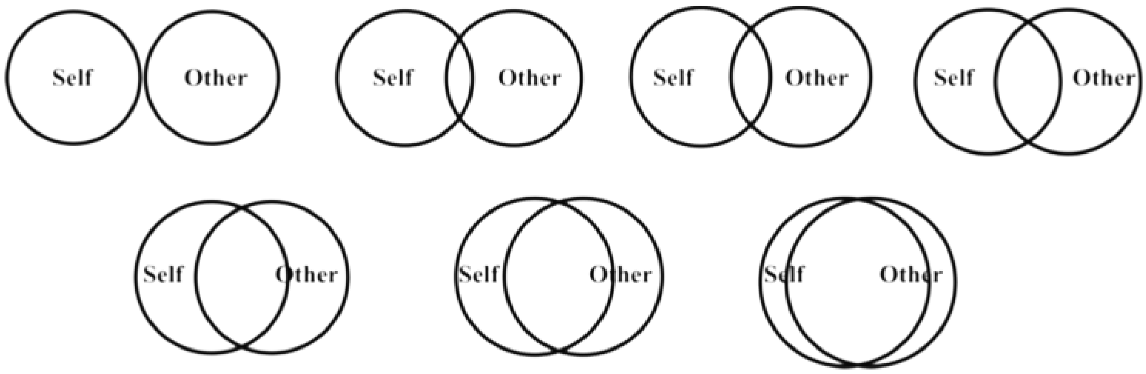


Figure 1. The Inclusion of Other in the Self (IOS) scale, a single-item pictorial measure of social connectedness. Reproduced from Aron, A., Aron, E. N., & Smollan, D. (1992). Inclusion of Other in the Self Scale and the structure of interpersonal closeness. *Journal of Personality and Social Psychology, 63*(4), 596–612. Reprinted with permission of the American Psychology Association as publisher.

Table 1. Results of Overall Emotion Perception Accuracy in Study 1 for Dances Expressing Melancholy, Grief, and Fear in Both Unimodal and Multimodal Conditions Over 807 Trials.

	Overall Accuracy	
Melancholy	65.4%	(174 of 266 trials)
Grief	66.8%	(179 of 268 trials)
Fear	69.2%	(189 of 273 trials)

Observers correctly identified dances expressing melancholy more often than chance, successful trials = 174, total trials = 266, $p < .01$, 95% CI [.597, .711]. In the case of dances of melancholy, Cohen's *g*—the difference between the two percentages—is .321 (Cohen, 1988; Rosnow & Rosenthal, 2003). Any Cohen's *g* larger than .250 is considered a large effect size (Cohen, 1988). Similarly, observers correctly identified dances expressing grief more often than chance, successful trials = 179, total trials = 268, $p < .01$, 95% CI [.612, .724], with a similarly large effect size (Cohen's $g = .335$). Finally, observers also correctly identified dances expressing fear more often than chance, successful trials = 189, total trials = 273, $p < .01$, 95% CI [.638, .747], again with a large effect size (Cohen's $g = .359$).

We conducted a χ^2 test to compare the distributions of correct and incorrect emotion identification responses across the three emotion conditions (melancholy, grief, and fear). People did not differ in accuracy in the three emotion conditions, $\chi^2 = .915$, $df = 2$, $p = .633$. The results therefore do not support Hypothesis 1, which was that people would identify emotions accompanied by overt displays with more accuracy than emotions accompanied by covert displays.

Accuracy of Emotion Perception in Unimodal and Multimodal Settings

Overall, participants correctly identified the expressed emotion with 69.6% accuracy in the multimodal (dancing to music) condition and 64.8% accuracy in the unimodal (dancing in silence) condition (Table 2). The results of a Fisher's exact test were not consistent with the idea that people differed in accuracy between uni- and multimodal conditions, odds ratio = .805, $p = .155$. Follow up tests for each emotional expression revealed that, in the melancholy condition, people did not differ in accuracy in the multimodal condition (63.0%) or the unimodal condition (67.9%), odds ratio = 1.246, $p = .440$. People were more accurate, however, in identifying dances expressing grief in the unimodal condition (73.3%) than in the multimodal condition (60.2%), odds ratio = 1.822, $p = .027$. Finally, people were more accurate in identifying dances expressing fear in the multimodal condition (85.3%) than in the unimodal condition (53.3%), odds ratio = .197, $p < .01$.

Perception of Sociality in Dance Videos

Hypothesis 2 indicated that observers would perceive more sociality in dance videos expressing grief than in videos expressing melancholy. We used multiple linear regression to determine whether certain a priori features (including emotion type and modality) predicted the perceived connection among dancers. The variables used to predict perceived connection among dancers were (a) modality (unimodal vs. multimodal), (b) emotion type (melancholy, grief, fear), (c) intensity

Table 2. Results of Emotion Perception Accuracy for Unimodal (Dancing in Silence) Conditions and Multimodal (Dancing to Music) Conditions Over 807 Trials.

	Unimodal Accuracy		Multimodal Accuracy	
Overall	64.8%	(261 of 403 trials)	69.6%	(281 of 404 trials)
Melancholy	67.9%	(89 of 131 trials)	63.0%	(85 of 135 trials)
Grief	73.3% *	(99 of 135 trials)	60.2% *	(80 of 133 trials)
Fear	53.3% *	(73 of 137 trials)	85.3% *	(116 of 136 trials)

Note. The asterisk (*) indicates when the differences in accuracy between the unimodal and multimodal conditions were statistically significant.

(0—*not at all intense* to 100—*extremely intense*), (d) years of music training, and (e) years of dance training. We investigated modality and emotion types because they are the variables central to the study hypotheses; emotion type was one-hot encoded. Intensity and years of music and dance training were included in the regression analysis to act as statistical controls: Their inclusion in the model allows us to examine the effects of the main variables (modality and emotion type) independent of these features. Such a process helped minimize the possibility of spurious or epiphenomenal effects in the model.

Trials where the participants left some answers blank were not included in the analysis. Furthermore, only trials where the participants correctly identified the emotion expressed by the dancers were included in the analysis (515 trials), a decision that increased power and helped the main effects remain clear.[2] Some assumptions of linear regression were met, while others were not. The data did not suffer from multicollinearity (all VIF, other than the one-hot encoded emotion types, were < 5) and met the assumptions of homoscedasticity, Breusch-Pagan test: Lagrange multiplier statistic, $LM = 6.005$, $p = .539$. A Jarque-Bera test indicated that the data were not normally distributed, but a Q-Q plot suggested that the deviations from normality were relatively small, $JB = 29.526$, $p < .01$, skew = -.523, kurtosis = 3.532. However, the errors of the model were not independent, Durbin-Watson test: $d = 1.803$, $p = .02$.

The regression analysis, summarized in Table 3, suggested that, when controlling for perceived emotional intensity and years of music and dance training, both fear and grief conditions resulted in higher ratings of perceived sociality among the dancers than melancholy conditions, as predicted by Hypothesis 2. In this regression analysis, perceived sociality was operationalized with the question, “How connected do you think the dancers feel to each other?”

The regression analysis results are consistent with the idea that videos with higher ratings of emotional intensity result in ratings of more sociality among the dancers. Participants with more years of music training perceived more sociality among the dancers, though the number of years of dance training did not affect the perceived sociality among dancers. The complete regression model resulted in an adjusted R^2 of .42, meaning that the variables (emotion type, emotional intensity, years of music training) explained 42% of the variance in the sociality scores.

A second regression model was performed using an alternative operationalization of sociality. In this second model, the dependent variable was the IOS scale (Aron et al., 1992), which we discerned by asking participants to "Please select the picture that best describes the relationship among the four dancers" (see Figure 1). Once again, only data where the participants correctly identified the emotional expression were included in the model. The data did not suffer from multicollinearity (all VIF other than the one-hot encoded emotion types were < 5) and the errors of the model were independent, $d = 1.830$, $p = .05$. Although the Jarque-Bera test indicated that the data were not normally distributed, a Q-Q plot suggested that the deviations from normality were relatively small, $JB = 11.753$, $p < .01$, skew = -.368, kurtosis = 2.928. The data did not meet the assumptions of homoscedasticity, Breusch-Pagan test: $LM = 16.841$, $p = .018$.

The five predictors were the same as in the first model: modality, emotion type, intensity of perceived emotion, years of music training, and years of dance training. For the most part, the results replicated those of the first regression: (a) videos where dancers were expressing grief or fear resulted in more perceived sociality among the dancers than did videos expressing melancholy, (b) higher scores of perceived emotional intensity led to more perceived sociality, (c) no effect of modality (dancing in silence versus dancing to music) was evident, and (d) dance training had no effect. The only finding that differed from the first model was that, in this second regression model, the effect of music training was not significant. This second regression model resulted in an adjusted R^2 of .35. In relation to Hypothesis 2, the results of the first and second regression models converge.

Table 3. Results of Linear Regressions for Perceived Sociality Among Dancers.

(a) Predicting Scores of Connection Among Dancers

	Estimate	*Std. Error*	*t*	*p*
(Intercept)	19.637	2.110	9.307	< .01
Fear	12.311	1.644	7.488	< .01
Grief	13.460	1.691	7.958	< .01
Melancholy	-6.135	1.441	-4.256	< .01
Intensity	.494	.041	11.898	< .01
Years of Music Training	.530	.176	3.011	< .01

(b) Predicting Scores of the Inclusion of Other in the Self Among Dancers

	Estimate	*Std. Error*	*t*	*p*
(Intercept)	2.001	.157	12.730	< .01
Fear	1.069	.122	8.723	< .01
Grief	1.152	.126	9.139	< .01
Melancholy	-.219	.107	-2.041	.04
Intensity	.030	.003	9.645	< .01

Note. Predictors were modality (unimodal vs. multimodal), emotion type (one-hot encoded), emotional intensity (0–100 scale), years of music training, and years of dance training. (a) Predicting scores of perceived connection among the dancers (0–100 scale), (b) Predicting IOS scores among the dancers (1–7 scale).

In order to test the difference in sociality perceptions between the dances expressing melancholy and grief, a post hoc logistic regression was conducted without the data related to fear expressions. The model predicted whether the dance videos were expressing melancholy or grief from the following predictors: modality (unimodal or multimodal), emotional intensity (0–100 scale), scores of perceived connection among the dancers (1–100 scale), scores of experienced connection between the self and the dancers (1–100 scale), self–other IOS (1–7 scale), dancer–dancer IOS (1–7 scale), years of music training, years of dance training, and emotion identification (correct vs. incorrect). The assumptions of logistic regression were met, as all continuous predictors and the logit of the outcome variable were linearly related. No multicollinearity was detected (all predictors had a VIF < 5).

The post hoc logistic regression analysis, summarized in Table 4, resulted in an accuracy of 70.4% and an AUC of .774. A confusion matrix of the predicted and observed values contained 174 correct melancholy classifications, 180 correct grief classifications, 78 false grief classifications (the dance expression was melancholy, but the model predicted grief), and 71 false melancholy classifications (the dance expression was grief, but the model predicted melancholy). The logistic regression model provides supporting evidence for the claim that slightly higher ratings of emotional intensity and higher scores of perceived sociality among dancers are more likely to be associated with videos where the dancers were expressing grief than with videos where the dancers were expressing melancholy. This sociality-related finding is consistent across two operationalizations of sociality: ratings of connection among the dancers and ratings of the dancer–dancer IOS. Therefore, the results of the logistic regression are consistent with Hypothesis 2. There was no effect of experienced sociality between the participants and dancers, operationalized in two ways (ratings of connection of self to dancers, ratings of self–video IOS). Modality, years of music training, years of dance training, and correct or incorrect emotion identification also were nonsignificant in the logistic regression model.

Discussion, Study 1

Emotional Accuracy

The first aim of Study 1 was to investigate whether observers would be more accurate in identifying dances expressing grief—which were conjectured to contain overt emotional displays

Table 4. Results of Post Hoc Logistic Regression Predicting Dances Expressing Melancholy (Coded as 0) versus Dances Expressing Grief (Coded as 1).

	Estimate	Std. Error	*z*	*p*
(Intercept)	-2.807	.384	-7.313	< .01
Intensity	.014	.006	2.494	.01
Perceived connection among dancers	.022	.006	3.520	< .01
IOS between dancers	.243	.084	2.901	< .01

—than dances expressing melancholy—which were conjectured to contain covert emotional displays. The results of Study 1 were not consistent with this hypothesis: There was no effect of grief versus melancholy in the accuracy of emotion perception.

The fact that observers did not differ in their accuracy between identifying dances expressing grief than dances expressing melancholy differs from Huron's (2015, 2016) theory that grief may function as an ethological signal and melancholy may function as an ethological cue. Huron proposed that because grief oftentimes is accompanied by overt displays and melancholy is often accompanied by covert displays, observers should be able to discern expressions of grief with greater accuracy than expressions of melancholy. An important aspect of the current study, however, is that we only examined melancholic and grieving expressions in aesthetic conditions. It is possible that, in aesthetic conditions such as dance, covert displays of an emotion like melancholy are exaggerated in order to communicate more directly with the observers. During performances, dancers may wish to alert the audience to emotions of their characters and therefore use overt expressions and behaviors to express all emotional states, regardless of how these states are expressed in everyday life. Additional research is needed to determine whether the similar accuracy in identifying melancholy and grief is replicated in nonaesthetic contexts.

Moreover, when designing the study, we expected that dances expressing fear would sometimes be accompanied by overt cues (related to high physiological arousal) and other times would be accompanied by covert cues (related to low physiological arousal). The intensity ratings of fear conditions (M = 57.271, SD = 24.811) were skewed toward higher ratings. Mean fear intensity fell between mean melancholy intensity (M = 45.414, SD = 27.431) and mean grief intensity (M = 60.366, SD = 22.979). The fact that fear was expressed in dances with relatively higher intensity could suggest that dancers use more overt displays to represent emotional states in order to aid in communication with audiences, as discussed above.

The similar accuracies also could be related to the 3-AFC paradigm used in Study 1. That is, the ability to differentiate grieving expressions from melancholic expressions might depend on the corresponding physiological arousal (intensity) of these two subjective feeling states: Participants could observe cues of low physiological arousal—which typically result in relatively covert displays—in the videos expressing melancholy and observe cues of high physiological arousal—which typically result in comparatively overt displays—in the videos expressing grief. Thus, the results of Study 1 could be interpreted in two ways: They could indicate that participants are sensitive to (a) cues of high and low physiological arousal or (b) highly overt and highly covert emotional displays. Future studies should utilize numerous negative emotions typically associated with high physiological arousal equally with those typically associated with low physiological arousal. Additionally, future research should directly measure the correlation between continuous scales of intensity and the relative amount of overt emotional displays. Finally, new research should replicate the current study and include more than three negative emotion choices (melancholy, grief, fear), include positive emotion choices, or use a free-response format.

The test comparing accuracy of emotion identification between unimodal (dancing in silence) and multimodal (dancing to music) conditions did not reveal a difference in accuracy, a result that differs from previous findings showing that dance–music conditions typically result in more accurate emotion identification than dance–only conditions (Burger et al., 2013). In that study by Burger and colleagues, the researchers found that dance–only conditions expressing negative emotions were sometimes mistaken as expressing positive emotions. The researchers theorized that the erroneous detection of positive emotions could be due to the fact that dance is usually an enjoyable experience. In our study design, we did not examine any positively valenced emotions and observers were not given options to select a positive emotion. Moreover, the lack of accuracy differences between uni- and multimodal conditions in the present study could suggest that, when presented with a list of only negatively valenced emotions, observers find it easier to identify the expressed emotion because they need to discern differences only in comparatively overt or covert displays.

It is possible that the lack in difference of accuracy between the uni- and multimodal conditions in the current study is due to cue redundancy in dance and music. Namely, a similar number (or type) of emotional expressions may be present in both dance and music so that an observer gains no additional information from bimodal expressions (dance–music), as compared to a unimodal expression (dance–only, music–only). Recent research examined 18 music structural features in melancholic and grieving musical samples, including the melancholic and grieving excerpts used in the present study (Warrenburg, 2020b). The results of that music study suggest that melancholic music tends to be quiet, low in register, and contain narrow pitch intervals, while grieving music tends to contain sustained tones, gliding pitches, and harsh timbres. It is possible, then, that dance-related expressions of melancholy mirror these music compositional cues; for example, they may contain small movements and only use a confined amount of space. Expressions of grief in dance may also be similar to their musical parallels and employ large, sweeping movements and utilize a wider area of space.

It is enlightening to compare the emotion identification accuracy of dance–only expressions, music–only expressions, and dance–music expressions of grief and melancholy. Warrenburg (2020c) previously conducted a study of emotion perception in musical passages that express melancholy and grief, including the musical excerpts used in the present study. The music study used a different methodology than the current dance study: The music study included both positively and negatively valenced musical excerpts and a 5-AFC question format (grief, melancholy, happiness, tenderness, none). Despite these methodological differences, we analyzed the emotion identification responses in the music study for the melancholic excerpt by Fauré (1878/2015, track 8) and the grieving excerpt by Arnold & Price (2012, track 18). In this analysis, the positively valenced emotion choices were ignored—only the responses of melancholy and grief were compared. The music–only condition of melancholic aesthetic expressions resulted in an accuracy of 70%, as opposed to the dance–only condition (68% accuracy) and the dance–music condition (63% accuracy). Similarly, the music–only condition of grieving aesthetic expressions resulted in an accuracy of 84%, compared to the dance–only condition (73% accuracy) and the dance–music condition (60% accuracy). Future work needs to directly test whether participants observe similar or different cues in dance expressions and musical expressions. As one example, researchers could mismatch music and dance conditions—in one condition, the music would express grief while the dance would express melancholy. Results from this type of study could assess the relative importance of music-related and dance-related cues.

Sociality

The second goal of Study 1 was to test whether observers would perceive more sociality in videos expressing emotions with overt displays (grief) than in videos expressing emotions with covert displays (melancholy). We based this hypothesis on the theory that although melancholy may have evolved (in part) to enable self-reflection, grief may have evolved (in part) to signal that the experiencer needs assistance from other people (Huron, 2015, 2016). The results of Study 1 support the idea that when dancers express emotions that typically are accompanied by overt displays, such as grief, observers perceive the dancers as being more socially connected than when dancers express an emotion like melancholy, which usually is accompanied by covert displays.

Thus, we might expect that, in aesthetic conditions such as dance and music, the combination of grieving sonic and movement cues in expressions of grief (and fear) may result in “repercussive” feelings of compassion or prosociality on the part of the observer or listener (Huron & Vuoskoski, 2020). These feelings of compassion or prosociality theoretically should be diminished in aesthetic expressions of melancholy. Warrenburg (in press) found in a study of free classifications of experienced emotions to grieving and melancholic music that respondents reported more feelings of crying and death/loss in music expressing grief, whereas instances of reflection were more frequent with music expressing melancholy. She theorized that reflection, a nonsocial behavior, may correspond to the theorized function of melancholy, which is to ruminate or self-reflect on a failed goal or poor experience. On the other hand, crying and death/loss are social behaviors that align with the theorized function of grief, which is to solicit assistance or comfort from another person (Huron, 2015, 2016; Urban, 1988; Vingerhoets & Cornelius, 2012). The findings of Warrenburg (2020c) and the results of the current study are consistent with the idea that in aesthetic conditions—music–only, dance–only, and dance–music—participants perceive more sociality in emotions often accompanied by overt displays, such as grief, than in emotions often accompanied by covert displays, such as melancholy. The post hoc logistic regression presented in Table 4 was consistent with the idea that people perceive more sociality among the dancers in videos that expressed grief compared to videos where the dancers expressed melancholy.

Earlier, we noted the possibility that emotional expressions may be heightened in aesthetic conditions (dance, music) as compared to nonaesthetic conditions (everyday life). If this speculation is correct, it could help explain our finding that fear expressions, in addition to grief expressions, also resulted in higher perceived sociality among the dancers than did melancholic expressions. Although fear can be accompanied by covert or overt displays in nonaesthetic conditions, it could be that, in aesthetic contexts, it is more desirable for performers to utilize overt expressions, including specific dance movements and social interactions with other dancers. These overt characteristics might better enable audience members to discern accurately the emotional state of their characters. We tested this conjecture directly in Study 2, where social interactions were operationalized as physical touch among dancers.

STUDY 2: SOCIALITY AMONG DANCERS

Study 1 addressed audience perception of sociality by testing whether participants would perceive more sociality among dancers expressing grief than among dancers expressing melancholy. Study 2 characterized sociality among the dancers in a different way, that is, with the prediction that dancers would exhibit more social behavior in emotions typically associated with overt displays, such as grief, compared to emotions typically associated with covert displays, such as melancholy (Hypothesis 3). In this second study, social behaviors were operationalized as the proportion of time in each improvisation that the dancers spent in physical contact with one another.

Method

The iPhone stopwatch application was used to record the length of time in seconds in each of the 60 s videos during which two or more dancers were in physical contact. Two independent coders who were not aware of the hypotheses performed this analysis. The coders watched each video silently, regardless of whether there was musical accompaniment, and were unaware of which videos had been recorded with music and which had not. Coders watched each video three times, and each coder watched the videos in a unique random order. On the first viewing, coders were asked to simply watch the video and observe patterns of touch without recording anything. On the second and third viewings, coders recorded the amount of time the dancers spent in physical contact by starting the stopwatch each time any two dancers came into contact and stopping it when all dancers ceased contact; these two recorded times were then averaged. This value was divided by the total time of each video, resulting in a proportion of time spent in physical contact.

Results

Table 5 lists the proportion of time spent in physical contact among two or more dancers in each of the 18 videos for each coder, as well as the average proportion of the two coders. Each proportion represents the amount of time of physical contact (touching) among at least two dancers, compared to the amount of time with no physical contact among any of the four dancers. Although the two coders' times did not always agree precisely, their recorded times were strongly correlated at $r = .99$.

Discussion, Study 2

The second goal of the paper was to examine whether differences could be identified in the sociality of dancers in performances expressing melancholy, grief, and fear. In Study 1, we tested observers' perception of sociality among dancers (Hypothesis 2); in Study 2, we tested the physical connection of the dancers (Hypothesis 3). The results of Study 1 were consistent with the idea that observers perceive relatively more sociality among performers in dances expressing grief and fear—which often are accompanied by overt emotional displays—than in dances expressing melancholy and its usually covert emotional displays. The results of Study 2 converge with the results of Study 1: When operationalizing social behavior as the amount of time that dancers spent in physical contact during the performance, we found that, on average, dancers spent a higher proportion of time in physical contact when they expressed grief and fear as compared to when they expressed melancholy. Furthermore, in both the grief and fear conditions, dancers spent more than 50% of the time in physical contact with each other (grief percent of touching M = 65%; fear percent of touching M = 59%). In the melancholic condition, dancers spent only 30% of the time in physical contact. As demonstrated in Figure 2, this order of grief, fear, melancholy is preserved not only in the mean proportions of time in physical contact, but also in the range of times spent in physical contact for each set of six videos per emotion condition. Due to the small sample size of videos, we did not run statistical analyses comparing these conditions. To establish the possible significance of a difference in physical contact between dances expressing grief, fear, and melancholy, we recommend replicating this study with a larger sample of videos, ideally with more than one group of dancers.

Table 5. Proportion of Time Spent in Physical Contact, Viewed on Video.

Condition	Video	Proportion of time spent in contact		
		Coder 1	*Coder 2*	*Average*
Melancholy, dance-only	1	.02	.08	.05
	2	.45	.45	.45
	3	.32	.31	.32
Melancholy, dance-music	1	.00	.00	.00
	2	.33	.29	.31
	3	.66	.70	.68
Grief, dance-only	1	.43	.38	.41
	2	.67	.70	.69
	3	1.00	1.00	1.00
Grief, dance-music	1	.45	.45	.45
	2	.48	.41	.45
	3	.90	.88	.89
Fear, dance-only	1	.21	.22	.22
	2	.65	.65	.65
	3	.90	.92	.91
Fear, dance-music	1	.48	.48	.48
	2	.35	.34	.35
	3	.91	.91	.91

Note. Each proportion represents the amount of time that some level of physical touch took place between at least two dancers as compared to the time in which no physical contact occurred among any of the four dancers.

The results do not indicate a consistent trend of time spent in physical contact in dance–only versus dance–music conditions across emotion types. Furthermore, because all dance–music conditions were recorded after dance–only conditions, we cannot eliminate the possibility of order effects. Future research could expand on the current results through further application of the touch timing method presented here. The study could also be replicated using dancers without formal training, as opposed to the professional SYREN dancers in the current study. However, as groups of people unfamiliar with each other may be uncomfortable making physical contact in a dance improvisation, a study could explore the level of contact among amateur dancers, that is, with pairs or groups of people who are in relationships where physical touch is common, such as couples or families.

The differences in the amount of physical contact among dancers may have acted as information to help participants distinguish between dances expressing grief/fear versus dances expressing melancholy. That is, in response to a dance video with a high amount of physical touch, participants may have used this characteristic (likely unconsciously) to help make their decision to select grief (or fear) as opposed to melancholy. Alternately, in a dance video with a low amount of physical contact, participants may have used this display to aid in their decision to select melancholy as opposed to grief/fear. Because there was little difference in the amount

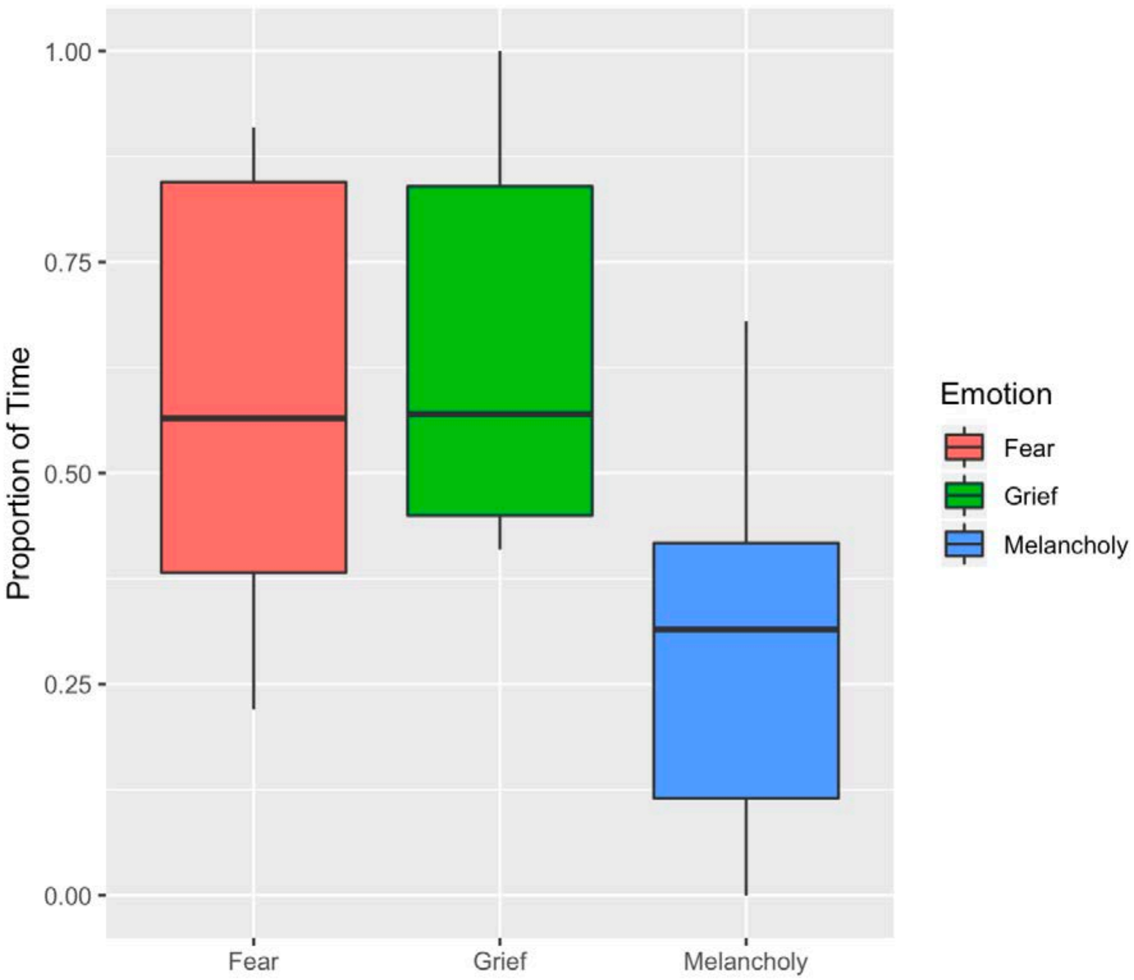


Figure 2. Box and whiskers plot of the proportions of time spent in physical contact among the dancers in each emotion condition (dance–only and dance–music combined). Times used in this graph are the average of both coders (see "Average" column of Table 5).

of physical contact in grief and fear conditions, the amount of physical touch among dancers probably was not a consequential factor in differentiating these two expressions. Instead, other types of bodily expressions likely were used to help participants distinguish fear and grief.

The idea that physical touch can help differentiate expressions of melancholy and grief, but not help differentiate expressions of grief and fear, is in line with the emotion literature discussed in the Introduction. Recall that both melancholy and grief are considered to be subtypes of sadness and can be experienced in response to the same event, such as the death of a marital partner. Fear, on the other hand, is elicited in different types of circumstances than melancholy and grief. Although the function of melancholy may be to self-reflect on a failure or minor loss, such as being rejected from a job application, and the function of grief may be to seek compassionate behaviors after a loss, Ekman (1992) hypothesized that fear may surface when a person expects to fail at a goal or task or anticipates suffering a loss of consequence. In the way that sadness may consist of multiple subtypes (e.g., grief, melancholy), fear itself may be considered to be a broad emotional category that can be broken into subtypes (e.g., anxiety, panic; Ji & Maren, 2007; Mobbs et al., 2007). Given this research, it may be more likely for a person's emotional response to the death of their marital partner to alternate regularly between periods of melancholy and grief, and less likely to regularly alternate between periods of melancholy and fear (or periods of grief and fear). Of course, personal situations, contexts and environments, personalities, and physical responses vary greatly among people and this trend will not explain many peoples' responses to a tragic event.

In response to the 3-AFC task of selecting the terms melancholy, grief, and fear to best describe dance expressions, one approach taken by participants may have been first to decide between general sadness (melancholy or grief) and fear. Subsequently, if they perceived sadness in the videos, they could then have used additional factors, including physical touch, to determine whether the subtype of sadness was melancholy or grief (see Warrenburg, 2020b, for more details about distinguishing characteristics of melancholy and grief). Another approach taken by participants could have been to first use cues of physical touch to differentiate emotions typically associated with low physiological arousal (such as melancholy) from emotions typically associated with high physiological arousal (such as grief and sometimes fear). After this distinction, the participant then may have examined emotional displays to distinguish expressions of fear and grief.

GENERAL DISCUSSION

Two experiments tested the theory that the difference between overt and covert emotional displays is pertinent to emotional expressions in dance. One goal of the present study was to test whether people perceive dances expressing emotions such as grief—which tend to be accompanied by overt displays—with higher accuracy than dances expressing emotions like melancholy, which tend to be accompanied by covert displays. The results of the first study were not consistent with this idea: Participants did not differ significantly in accuracy among conditions of melancholy, grief, and fear. It is possible that the overt/covert distinction, usually discussed in nonaesthetic conditions (everyday life), does not carry over to aesthetic conditions (dance, music). Because one of the major goals of art, including dance and music, is to communicate with audiences, emotional displays may be heightened in expressions of melancholy, grief, and fear.

The second goal of the study was to examine whether dances expressing grief contained more social interactions than dances expressing melancholy. The results of the second study are in line with this conjecture. First, we found that observer participants perceived more connection among the dancers in dances expressing grief and fear than in dances expressing melancholy. Second, we found that dancers were in physical contact more often, as determined by two independent timing coders, when expressing grief and fear than when expressing melancholy. The findings of both studies have implications for the role of sociality in aesthetic conditions.

Sociality in Aesthetic Conditions

Dancing, music making, and music listening are all inherently social activities that can unify groups of people, contribute to emotional contagion, and give people a sense of identity (Huron, 2005). For example, dance and music are essential components in many kinds of religious ceremonies and war dances. Behaviors such as synchronization to music and coordination through dance may promote social bonding and prosocial actions, possibly leading to feelings of pleasure (Semin & Caccioppo, 2008). Huron (2005) theorized that the emotions arising from listening to, performing, and dancing to music are contagious. Because emotional contagion can lead to social congruence and cooperation, Huron suggested that the effectiveness of a group's actions may improve when dancing to music. Expressing emotions in aesthetic conditions, then, may be beneficial for the affective state of each group member.

In addition to occurring within the self, emotions may occur on different social levels, including on an interpersonal level, an intergroup level, and a sociocultural level (Haidt, 2012; Keltner & Haidt, 1999). Although the natural human state may be to focus on the needs and goals of the individual (for survival purposes), at times individual competition is not as important as group cooperation (Darwin, 1872). Certain emotions, such as awe or transcendence, can be associated with group levels of focus, which in turn can help group members define group-related roles and help them assume cultural identities (Haidt, 2012). Research has shown that people who are better able to understand others' emotions tend to be more imaginative, exhibit more forgiveness, and act more prosocially (Strayer & Roberts, 1989; Wieseke, Geigenmüller, & Kraus, 2012). Additionally, some negative emotions, such as grief, can also be experienced interpersonally, such as when partners experience the loss of their child. These shared emotions may lead to heightened displays of sociality among the group experiencers.

In order to learn about the dancers' experiences of sociality in the melancholic, grieving, and fearful conditions, we conducted a group interview with the SYREN dancers once they had finished improvising to all of the emotional prompts. In this interview, we aimed to capture the dancers' mental schemas of the target emotions and to gain insight into the dancers' experiences in expressing these three emotions. The comments from these interviews further support the idea that group dynamics were magnified during improvisations expressing grief and fear, as compared to improvisations expressing melancholy.

The dancers were asked to talk about their own definitions of melancholy, grief, and fear; which actions they used to express these three emotions; and if (and how) their experience differed when dancing with and without music. We did not inform the dancers of the hypotheses or the purpose of the experiment; however, the theme of sociality arose as the dancers addressed the types of actions they used to express each of the emotions.

> Researcher (R): How did you choose the specific actions that communicated each of the emotions?
>
> Dancer 1 (D1): *...sometimes I feel like, as a group, we definitely collectively decided we're going to do* [the dance] *this time together. In grief, one of the times, we were very much just like—*
>
> Dancer 2 (D2): *—comforting.*
>
> *D1: —getting through it together. So I think there were different lenses that each one took little bit that way.*
>
> R: I remember what you're talking about, that time where you were all very integrated in the grief. Did you feel like you had that experience when doing the melancholy?
>
> Dancer 3 (D3): *At times.*
>
> D2: *I feel like grief and fear felt more natural to find that interaction.*
>
> D3: *Right. I thought melancholy was a little—I don't want to say playful, but there were more moments of "Here I'm going to do something with this person and, real quick, I'm going to leave them and go into something with another person."*

D1: *I don't feel like we had that kind of group sense, where it was like we're all going to tackle this together—maybe that's because melancholy is something we kind of each go through on our own, and fear and grief we—*

Dancer 4 (D4)*: —we should reach out more.*

D2: *There's definitely more support in fear and grief, I would—like, I would say, in day to day. So, I feel like our improv reflected that.*

In light of these statements, it is possible that the dancers interpreted emotions like grief and fear as emotional states that give rise to a collective experience among all dancers. For example, they seem to have interpreted fear as a shared emotion, so that all of the dancers portrayed fear: They all responded as a single group to an external aggressor, rather than having some dancers portraying aggressors and other dancers portraying fearful victims. Similarly, in dances expressing grief, the dancers seem to have interpreted the emotion as a shared grieving experience, where all dancers expressed grief in an interpersonal manner.

Emotional Displays Cannot Be Interpreted as Affective Feelings

As use of recording technology has expanded in the last century, and particularly with the advent of multiple video streaming sites, it has become possible to listen to music or to watch dances in nonsocial settings. The advent of headphones and portable electronic devices, such as laptops, tablet computers, CDs, and mobile phones, allow people to watch videos and listen to music outside of the home, as well. Today, much of the music listening in the West is done privately (Greasley & Lamont, 2011; Krause, North, & Hewitt, 2016). Accordingly, research into music and dance often examines emotional responses to these art forms within an individual setting. The studies presented here, as well as the interview with the SYREN dancers, suggest that displays and experiences of sociality are central to artistic emotional expressions.

Technologies such as motion capture, video recordings, and psychophysiology methods often have been used in academic research to learn how emotions are expressed and perceived in nonaesthetic conditions (Adolphs, 2017; Cordaro et al., 2020; Cowen & Keltner, 2017; Keltner, Sauter, Tracy, & Cowen, 2019). The results of the current study suggest that technology can also capture and communicate certain aspects of how artists express emotions in aesthetic conditions (also see Alaerts, Nackaerts, Meyns, Swinnen, & Wenderoth, 2011; Cowen, Fang, Sauter, & Keltner, 2020; Trevor, 2018; Vuoskoski, Thompson, Clarke, & Spence, 2020, for this type of work).

Recall, however, that a 1-to-1 correlation does not exist between emotional expressions or behaviors and underlying subjective feeling states (Barrett, 2017; Russell, 2003). Although technologies and physiological measures provide insight into a person's emotional expressions and behaviors, these displays cannot be interpreted as affective feeling states. As mentioned in the Introduction, without directly asking a person to label their subjective emotional state, and without understanding the situational context, the person's bodily interoceptive responses, and native language, it is impossible to conclusively determine the emotional state of another person. Given these considerations, future research on classifying emotional expressions could benefit from new approaches. Instead of classifying an emotional expression into a single category (using classification metrics), technologies could provide likelihoods that an expression fits under multiple dimensions (e.g., Bayesian approaches, convoluted neural networks).

Technology that analyzes expressions and behaviors should also be integrated with additional sources of information, including situational and contextual cues, as well as multimodal and continuous responses (e.g., bodily expressions, facial expressions, vocal signals). Drawing information from across multiple sources should aid in increased interpretability and ecological validity to the study of emotional expression in aesthetic and nonaesthetic contexts.

The distinction between overt and covert displays, as well as the relative sociality of these displays, is central to interpreting a person's emotional expressions or behaviors. Overt emotional displays will be conspicuous and possibly multimodal. Technologies will have a relatively easier time recognizing overt emotional expressions in people because of potential differences among facial, vocal, and/or bodily expressions. The baseline performance in classifying covert emotional expressions should be expected to be lower. For example, the physiological characteristics and expressions of people in melancholic and sleepy states overlap extensively, with some researchers even suggesting that sleepiness and melancholy can only be differentiated by invisible cognitive factors (Andrews & Thomson, 2009; Nesse, 1991). Emotions accompanied by covert displays may not contain any observable or distinctive facial, vocal, or behavioral expressions. Thus, a technological device may not be able to differentiate among covert emotional displays without additional information from the user. Baseline understandings of a person's emotional granularity, first language (and that language's emotional lexicon), verbal self-reported emotional feelings, personality, past history, situational location (alone or with others), time of day, and relative health are all important differentiators in the detection of covert (and overt) emotional displays.

CONCLUSIONS

The current study investigated how people perceived expressions of negatively valenced emotion in silent dances and dances with musical accompaniment. Four members of a professional dance company, SYREN Modern Dance, were recorded dancing with the intention of expressing melancholy, grief, or fear. In the first study, observer participants were asked to identify which emotion the dancers were expressing in each video using a 3-AFC paradigm. In addition, the participants were asked to rate how socially connected they believed the dancers to be throughout the dance. In the second study, two individuals who did not participate in the first study and were not informed of the intent of our study were asked to code the percentage of time the dancers were in physical contact in each video.

Hypothesis 1 predicted that observers would be more accurate when identifying dance expressions of grief than when identifying dance expressions of melancholy because feelings of grief are often accompanied by overt emotional displays and feelings of melancholy are often accompanied by covert emotional displays. Although participants were able to identify expressions of melancholy, grief, and fear more than chance, there was no difference in accuracy among these three emotion conditions.

Hypothesis 2 predicted that observers would perceive more sociality among dancers expressing grief than among dancers expressing melancholy. The results were consistent with this hypothesis: Controlling for perceived emotion intensity and years of music/dance training, dances expressing grief (and fear) resulted in higher observer ratings of social connection among the SYREN dancers than dances expressing melancholy.

Hypothesis 3 predicted that dancers would objectively exhibit more social behaviors while dancing to express grief than while dancing to express melancholy. Social behavior was defined as the amount of time the dancers spent in physical contact with each other. Once again, the results were consistent with this idea. Dancers behaved more socially while expressing grief (and fear) than while expressing melancholy. The results of the two studies are consistent with the idea that aesthetic expressions of grief and fear facilitate more social behaviors than aesthetic expressions of melancholy.

IMPLICATIONS FOR RESEARCH AND APPLICATION

The presented research provides insight into how negative emotions are expressed in dances with and without music, as well as how dance and music work together to facilitate perceptions of the performers' social connection. Future research focusing on the nature of emotional expressions in aesthetic conditions can build on our findings by examining not only the movements of dancers, but also on the movements of musicians and audience members.

The implications of using technology to understand expressions of emotion in aesthetic and nonaesthetic conditions extends beyond academia and into industry, as many companies specialize in detecting and evaluating consumers' emotions through technology present in cell phones, computers, and videos. Even though technologies and physiological measures certainly can provide users, researchers, and businesses with information about a person's emotional displays, the expressions captured by technology should not be interpreted as affective feeling states. However, by including information about overt and covert emotional displays, as well as social versus nonsocial emotional expressions, technology companies could expect improvements in model accuracy, sensitivity, and specificity, as well as their overall classification of emotional expressions. The ability of technology to aid in interpreting emotional displays in aesthetic (and nonaesthetic) contexts affects businesses of all sizes, as well as governmental systems like SPOT (Screening Passengers by Observation Techniques), which was designed to detect signs of deception or fear in airport passengers.

ENDNOTES

1. For further detail, see Cialdini, Brown, Lewis, Luce, & Neuberg (1997) and Hauser, Preston, & Stansfield (2014). Although an empathic person is able to distinguish between the self and the other during an empathic state, most people feel a "self–other overlap" (Aron et al., 1992) with others. This self–other overlap is thought to have evolved so that when an individual helps another person in need, it contributes to positive feelings in oneself.
2. That is, if noncorrect responses were included in the model, it is possible that each main effect could be split into separate trends: one trend for correct responses and a second trend for incorrect responses. If the direction of these two trends differed, the result could be a horizontal line with no direction. Separating trends due to correct and incorrect responses could have been handled through a mediation, moderation, or conditional process analysis had we had a larger number of participants and responses. Future research should replicate the findings of the present study with a larger number of participants.

REFERENCES

Adolphs, R. (2017). How should neuroscience study emotions? By distinguishing emotion states, concepts, and experiences. *Social Cognitive and Affective Neuroscience*, *12*(1), 24–31.

Alaerts, K., Nackaerts, E., Meyns, P., Swinnen, S. P., & Wenderoth, N. (2011). Action and emotion recognition from point light displays: An investigation of gender differences. *PloS ONE*, *6*(6), e20989.

Andrews, P. W., & Thomson, J. A., Jr. (2009). The bright side of being blue: Depression as an adaptation for analyzing complex problems. *Psychological Review*, *116*(3), 620–654.

Archer, J. (1999). *The nature of grief: The evolution and psychology of reactions to loss.* New York, NY, USA: Taylor & Francis Group.

Arnold, D., & Price, M. (2012). Blood on the pavement. On *Sherlock: Music from Series 2 (Original Television Soundtrack)* [mp3]. London, England: Hartswood Films Ltd.

Aron, A., Aron, E. N., & Smollan, D. (1992). Inclusion of Other in the Self Scale and the structure of interpersonal closeness. *Journal of Personality and Social Psychology, 63*(4), 596–612.

Barrett, L. F. (2017). *How emotions are made: The secret life of the brain*. New York, NY, USA: Houghton Mifflin Harcourt.

Burger, B., Saarikallio, S., Luck, G., Thompson, M. R., & Toiviainen, P. (2013). Relationships between perceived emotions in music and music-induced movement. *Music Perception: An Interdisciplinary Journal, 30*(5), 517–533.

Camurri, A., Lagerlöf, I., & Volpe, G. (2003). Recognizing emotion from dance movement: Comparison of spectator recognition and automated techniques. *International Journal of Human–Computer Studies*, *59*(1), 213–225.

Camurri, A., Mazzarino, B., Ricchetti, M., Timmers, R., & Volpe, G. (2003, April). Multimodal analysis of expressive gesture in music and dance performances. In A. Camurri & G. Volpe (Eds.), *International Gesture Workshop* (pp. 20–39). Berlin, Germany: Springer.

Cialdini, R. B., Brown, S. L., Lewis, B. P., Luce, C., & Neuberg, S. L. (1997). Reinterpreting the empathy–altruism relationship: When one into one equals oneness. *Journal of Personality and Social Psychology*, *73*(3), 481–494.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ, USA: Erlbaum.

Cordaro, D. T., Sun, R., Kamble, S., Hodder, N., Monroy, M., Cowen, A., Bai, Y., & Keltner, D. (2020). The recognition of 18 facial-bodily expressions across nine cultures. *Emotion*, *20*(7), 1292–1300.

Cowen, A. S., Fang, X., Sauter, D., & Keltner, D. (2020). What music makes us feel: At least 13 dimensions organize subjective experiences associated with music across different cultures. *Proceedings of the National Academy of Sciences, 117*(4), 1924–1934.

Cowen, A. S., & Keltner, D. (2017). Self-report captures 27 distinct categories of emotion bridged by continuous gradients. *Proceedings of the National Academy of Sciences, 114*(38), E7900–E7909.

Darwin, C. (1872). *The expression of emotions in man and animals*. New York, NY, USA: Oxford University Press.

Eerola, T., & Vuoskoski, J. K. (2011). A comparison of the discrete and dimensional models of emotion in music. *Psychology of Music, 39*(1), 18–49.

Ekman, P. (1992). An argument for basic emotions. *Cognition & Emotion, 6*(3-4), 169–200.

Ekman, P., & Friesen, W. V. (1975). *Unmasking the face: A guide to recognizing the emotions from facial cues.* Englewood Cliffs, NJ, USA: Prentice Hall.

Elfman, D. (1992). Selina transforms, Pt. 1. On *Batman Returns* [mp3]. Los Angeles, CA, USA: Warner Records Inc.

Epstein, S. (2019). *Four types of grief nobody told you about*. Retrieved from https://www.psychologytoday.com/us/blog/between-the-generations/201904/four-types-grief-nobody-told-you-about

Fauré, G. (2015). Aprés un rêve, Op. 7 No. 1 [Recorded by Y. Ma & K. Stott]. On *Songs from the Arc of Life* [mp3]. Arlington, MA, USA: Sound Postings LLC. (Original music published in 1878)

Frick, R. W. (1985). Communicating emotion: The role of prosodic features. *Psychological Bulletin, 97*(3), 412–429.

Frijda, N. H., & Sundararajan, L. (2007). Emotion refinement: A theory inspired by Chinese poetics. *Perspectives on Psychological Science*, *2*(3), 227–241.

Gelstein, S., Yeshurun, Y., Rozenkrantz, L., Shushan, S., Frumin, I., Roth, Y., & Sobel, N. (2011). Human tears contain a chemosignal. *Science, 331*(6014), 226–230.

Greasley, A. E., & Lamont, A. (2011). Exploring engagement with music in everyday life using experience sampling methodology. *Musicae Scientiae*, *15*(1), 45–71.

Gross, J. J., & Feldman Barrett, L. (2011). Emotion generation and emotion regulation: One or two depends on your point of view. *Emotion Review, 3*(1), 8–16.

Haidt, J. (2012). *The righteous mind: Why good people are divided by politics and religion*. New York, NY, USA: Pantheon Books.

Hauser, D. J., Preston, S. D., & Stansfield, R. B. (2014). Altruism in the wild: When affiliative motives to help positive people overtake empathic motives to help the distressed. *Journal of Experimental Psychology: General*, *143*(3), 1295–1305.

Huron, D. (2005). The plural pleasures of music. In J. Sundberg & W. Brunson (Eds.), *Proceedings of the 2004 Music and Music Science Conference* (pp. 1–13). Stockholm, Sweden: Kungliga Musikhögskolan & KTH.

Huron, D. (2015). Affect induction through musical sounds: An ethological perspective. *Philosophical Transactions of the Royal Society B: Biological Sciences, 370*(1664), 20140098. Retrieved from https://royalsocietypublishing.org/doi/pdf/10.1098/rstb.2014.0098

Huron, D. (2016). Cues and signals: An ethological approach to music-related emotion. *Signata: Annales des sémiotiques/Annals of Semiotics*, *6*, 331–351.

Huron, D., & Vuoskoski, J. K. (2020). On the enjoyment of sad music: Pleasurable compassion theory and the role of trait empathy. *Frontiers in Psychology*, *11*, 1–16. https://doi.org/10.3389/fpsyg.2020.01060

Ji, J., & Maren, S. (2007). Hippocampal involvement in contextual modulation of fear extinction. *Hippocampus, 17*(9), 749–758.

Juslin, P. N. (2013). What does music express? Basic emotions and beyond. *Frontiers in Psychology*, 4, 1–14. https://doi.org/10.3389/fpsyg.2013.00596

Juslin, P. N., & Sloboda, J. (Eds.). (2010). *Handbook of music and emotion: Theory, research, applications.* New York, NY, USA: Oxford University Press.

Keltner, D., & Haidt, J. (1999). Social functions of emotions at four levels of analysis. *Cognition & Emotion, 13*(5), 505–521.

Keltner, D., Sauter, D., Tracy, J., & Cowen, A. (2019). Emotional expression: Advances in basic emotion theory. *Journal of Nonverbal Behavior, 43*(2), 133–160.

Krause, A. E., North, A. C., & Hewitt, L. Y. (2016). The role of location in everyday experiences of music. *Psychology of Popular Media Culture*, *5*(3), 232–257.

Krumhansl, C. L., & Schenk, D. L. (1997). Can dance reflect the structural and expressive qualities of music? A perceptual experiment on Balanchines's choreography of Mozart's Divertimento No. 15. *Musicae Scientiae, 1,* 63–83.

Lang, P. J., Levin, D. N., Miller, G. A., & Kozak, M. J. (1983). Fear behavior, fear imagery, and the psychophysiology of emotion: The problem of affective response integration. *Journal of Abnormal Psychology*, *92*(3), 276–306.

LeDoux, J. E., & Hofmann, S. G. (2018). The subjective experience of emotion: A fearful view. *Current Opinion in Behavioral Sciences*, *19*, 67–72.

Leuchs, L., Schneider, M., Czisch, M., & Spoormaker, V. I. (2017). Neural correlates of pupil dilation during human fear learning. *NeuroImage*, *147*, 186–197.

Mazo M. (1994). Lament made visible: A study of paramusical elements in Russian lament. In B. Yung & J. Lam (Eds.), *Themes and variations: Writings on music in honor of Rulan Chao Pian* (pp. 164–211). Boston, MA, USA: Harvard University Press.

Milosevic, I., & McCabe, R. E. (Eds.). (2015). *Phobias: The psychology of irrational fear*. Santa Barbara, CA, USA: ABC-CLIO, LLC.

Mobbs, D., Petrovic, P., Marchant, J. L., Hassabis, D., Weiskopf, N., Seymour, B., Dolan, R. J., & Frith, C. D. (2007). When fear is near: Threat imminence elicits prefrontal-periaqueductal gray shifts in humans. *Science, 317*(5841), 1079–1083.

Moreno, R., & Mayer, R. (2007). Interactive multimodal learning environments. *Educational Psychology Review, 19*(3), 309–326.

Nesse, R. M. (1991). What good is feeling bad? The evolutionary benefits of psychic pain. *The Sciences*, *31*(6), 30–37.

Ohala, J. (1994). The frequency code underlies the sound-symbolic use of voice pitch. In L. Hinton, J. Nichols, & J. Ohala (Eds.), *Sound symbolism* (pp. 325–347). Cambridge, UK: Cambridge University Press.

Rosenblatt, P. C., Walsh, R. P., & Jackson, D. A. (1976). *Grief and mourning in cross-cultural perspective*. New Haven, CT, USA: Yale University, Human Relations Area Files.

Rosnow, R. L., & Rosenthal, R. (2003). Effect sizes for experimenting psychologists. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale, 57*(3), 221–237.

Russell, J. A. (2003). Core affect and the psychological construction of emotion. *Psychological Review*, *110*(1), 145–172.

Scherer, K. R. (2004). Which emotions can be induced by music? What are the underlying mechanisms? And how can we measure them? *Journal of New Music Research, 33*(3), 239–251.

Semin, G. S., & Caccioppo, J. T. (2008). Grounding social cognition. In G. S. Semin & E. R. Smith (Eds.), *Embodied grounding: Social, cognitive, affective, and neuroscientific approaches* (pp. 119–147). Cambridge, UK: Cambridge University Press.

Strayer, J., & Roberts, W. (1989). Children's empathy and role taking: Child and parental factors, and relations to prosocial behavior. *Journal of Applied Developmental Psychology, 10*(2), 227–239.

Tomkins, S. S. (1980). Affect as amplification: Some modifications in theory. In R. Plutchik & H. Kellerman (Eds.), *Emotion: Theory, research, and experience, Volume 1: Theories of emotion* (pp. 141–164). New York, NY, USA: Academic Press.

Tracy, J. L., & Randles, D. (2011). Four models of basic emotions: A review of Ekman and Cordaro, Izard, Levenson, and Panksepp and Watt. *Emotion Review, 3*(4), 397–405.

Trevor, C. (2018). *Cognitive and theoretical analyses of expressive performance choices.* (Unpublished doctoral dissertation, The Ohio State University, Columbus, OH, USA).

Urban, G. (1988). Ritual wailing in Amerindian Brazil. *American Anthropologist, 90*(2), 385–400.

Van Dyck, E., Burger, B., & Orlandatou, K. (2019). The communication of emotions in dance. In M. Lesaffre, P. J. Maes, & M. Leman (Eds.), *The Routledge companion to embodied music interaction* (pp. 122–130). New York, NY, USA: Taylor & Francis.

Vingerhoets, A. J., & Cornelius, R. R. (Eds.). (2012). *Adult crying: A biopsychosocial approach*. New York, NY, USA: Taylor & Francis.

Vuoskoski, J., Thompson, M., Clarke, E., & Spence, C. (2020). *The contribution of visual and auditory cues to the perception of emotion in musical performance*. Manuscript submitted for publication.

Warrenburg, L. A. (2019). *Subtle semblances of sorrow: Exploring music, emotional theory, and methodology.* (Unpublished doctoral dissertation, The Ohio State University, Columbus, OH, USA).

Warrenburg, L. A. (2020a). Comparing musical and psychological emotion theories. *Psychomusicology: Music, Mind, and Brain, 30*(1), 1–19.

Warrenburg, L. A. (2020b). Redefining sad music: Music's structure suggests at least two sad states. *Journal of New Music Research, 49*(4), 373–386.

Warrenburg, L. A. (2020c). *Melancholy and grief music express different affective states*. Manuscript submitted for publication.

Warrenburg, L. A. (in press). People experience different emotions from melancholic and grieving music. *Music & Science*. https://doi.org/10.1177/2059204320977384

Weinstein, D., Launay, J., Pearce, E., Dunbar, R. I., & Stewart, L. (2016). Group music performance causes elevated pain thresholds and social bonding in small and large groups of singers. *Evolution and Human Behavior: Official Journal of the Human Behavior and Evolution Society*, *37*(2), 152–158.

Wieseke, J., Geigenmüller, A., & Kraus, F. (2012). On the role of empathy in customer–employee interactions. *Journal of Service Research*, *15*(3), 316–331.

Authors' Note

We thank the SYREN Dance Troupe (www.syrendance.org) and Laura Wagner, without whom this research would not have been possible. Research was assisted by Nate Centa, Xintong Li, Hansae Park, Jessica Sarff, Diana Sari, and Feiyu Xie.

All correspondence should be addressed to
Lindsay Warrenburg
School of Music, The Ohio State University
Weigel Hall, 1866 College Rd N #110
Columbus, OH 43210
lindsay.a.warrenburg@gmail.com

Human Technology
ISSN 1795-6889
www.humantechnology.jyu.fi

HUMAN TECHNOLOGY

ISSN: 1795-6889

www.humantechnology.jyu.fi Volume 16(3), November 2020, 310–347

EXPLORING RELATIONSHIPS BETWEEN EFFORT, MOTION, AND SOUND IN NEW MUSICAL INSTRUMENTS

Çağrı Erdem
RITMO Centre for Interdisciplinary Studies in Rhythm, Time and Motion
University of Oslo
Norway

Qichao Lan
RITMO Centre for Interdisciplinary Studies in Rhythm, Time and Motion
University of Oslo
Norway

Alexander Refsum Jensenius
RITMO Centre for Interdisciplinary Studies in Rhythm, Time and Motion
University of Oslo
Norway

Abstract: *We investigated how the action–sound relationships found in electric guitar performance can be used in the design of new instruments. Thirty-one trained guitarists performed a set of basic sound-producing actions (impulsive, sustained, and iterative) and free improvisations on an electric guitar. We performed a statistical analysis of the muscle activation data (EMG) and audio recordings from the experiment. Then we trained a long short-term memory network with nine different configurations to map EMG signal to sound. We found that the preliminary models were able to predict audio energy features of free improvisations on the guitar, based on the dataset of raw EMG from the basic sound-producing actions. The results provide evidence of similarities between body motion and sound in music performance, compatible with embodied music cognition theories. They also show the potential of using machine learning on recorded performance data in the design of new musical instruments.*

Keywords: *EMG, music, machine learning, musical instrument, motion, effort, guitar, embodied.*

DOI: https://doi.org/10.17011/ht/urn.202011256767

VERTAISARVIOITU KOLLEGIALT GRANSKAD PEER-REVIEWED www.tsv.fi/tunnus

INTRODUCTION

What are the relationships between action and sound in instrumental performance, and how can such relationships be used to create new instrumental paradigms? These two questions inspired the experiments presented in this paper. Our research is based upon two basic premises: It is possible to find relationships between the continuous, temporal shape of an action and its resultant sound and that embodied knowledge of an existing instrument can be translated into a new performative context with different instrument. Thus, we are interested in exploring whether it is possible to create mappings in new instruments based on measured actions on and sounds from an existing instrument. It is common to create such action–sound mappings based on overt motion features. However, in our study, we were interested primarily in exploring whether covert muscle signals can be used for new musical instruments.

Embodied Knowledge

The body's role in the experience of sound and music is central to the embodied music cognition paradigm (Leman, 2008). Several studies have explored the embodiment of musical experiences by investigating how musicians and nonmusicians transduce what they perceive as musical features into body motion. Sound-tracing is one such experimental paradigm that has been used to study how people spontaneously follow salient features in music (Kelkar, 2019; Kozak, Nymoen, & Godøy, 2012; Nymoen, Caramiaux, Kozak, & Torresen, 2011). Sound mimicry is a similar approach, based on examining how sound-producing actions can be imitated "in the air," that is, without a physical interface (Godøy, 2006; Godøy, Haga, & Jensenius, 2005; Valles, Martínez, Ordás, & Pissinis, 2018). Several other studies have aimed at identifying musical mapping strategies, drawing on concepts of embodied music cognition as a starting point (e.g., Caramiaux, Bevilacqua, Zamborlin, & Schnell, 2009; Françoise, 2015; Maes, Leman, Lesaffre, Demey, & Moelants, 2010; Tanaka, Donato, Zbyszynski, & Roks, 2019; Visi, Coorevits, Schramm, & Miranda, 2017).

In this study, we took bodily imitation as the starting point for the creation of action–sound mappings. The idea was to transfer the acquired skills of playing traditional instruments to a new context. Here the term traditional refers to the recognizability of performance skills, what Smalley (1997) explained as an intuitive knowledge of action–sound causalities in traditional sound-making. The idea was to exploit such proprioceptive relationships between musician and instrument (Paine, 2009). The premise is that skill can be understood as embodied knowledge (Ingold, 2000) that leads to lower information processing at a cognitive level (Dreyfus, 2001). It also builds upon the idea that spectators can perceive and recognize skill as an embodied phenomenon (Fyans & Gurevich, 2011).

One outcome of this research was aimed at developing solutions for creating musical instruments that can be performed in the air. However, it should be clear from the start that we are not interested in making "air" versions of the guitar or any other physical instrument. Rather, our attention is devoted to reusing the embodied knowledge of one type of instrumental performance in new ways (Magnusson, 2019). The lack of a haptic and tactile experience creates a significantly different experience when playing a physical instrument as compared to a touchless air instrument. According to the "gestural agency" concept of Mendoza Garay & Thompson (2017), the instrument is as much an agent in the musical transaction as the performer:

They influence each other within a musical ecosystem. In this system, the agents' communication is multimodal. Therefore, the act of instrument playing accommodates not only the auditory, tactile, and haptic channels but also the visual, kinetic, proprioceptive, or any other kind of interactions that have a musical influence. The human agent becomes the participant that is expected to adapt; thus, any change in the environment can be seen as a creative challenge.

From Body Motion to Musical Actions

Gesture is employed frequently in the literature on music-related body motion (Cadoz & Wanderley, 2000; Gritten & King, 2011; Hatten, 2006). We understand gesture as related to the meaning-bearing aspects of performance actions. In this project, we focus not on such meaning-bearing aspects and thus will not use that term in the following discussion. Instead, we will use *motion* to describe the continuous displacement of objects in space and time, and *force* to explain what sets these objects into motion. Both motion and force are physical phenomena that can be captured and studied using various devices (see Jensenius, 2018a, for an overview of various methods for sensing music-related body motion). Hitting a guitar string is an example of what we call motion, which can be studied through motion capture data of the arm's continuous position. Muscle tension is an example of the force involved in the sound production and can be studied through electromyography (EMG).

Motion and force describe the kinematic and kinetic aspects of performance, respectively. These relate to—but are not the same as—the experienced action within a performance (Jensenius, Wanderley, Godøy, & Leman, 2010). Thus, in our research, we use *action* to describe a cognitive phenomenon that can be understood as goal-directed units of motion and/or force (Godøy, 2017). Many actions are based on visible motion, but an action also can be based solely on force. For example, some electroacoustic musical instruments are built with force-sensitive resistors that can be pressed by the performer, even without any visible motion. Hence the player's action can change drastically over time even with no or only little observable body motion.

Music-related body motion comes in various types (see Jensenius et al., 2010, for an overview). Here we primarily focus on the *sound-producing actions*. These can be subdivided into *excitation* actions, such as the right hand that excites the strings on a guitar, and *modification* actions, such as the left hand modifying the pitch. The excitation action can be divided further into the three main categories proposed by Schaeffer (2017), as sketched in Figure 1: *impulsive*, *sustained*, and *iterative*. An impulsive excitation is characterized by a fast attack and discontinuous energy transfer, while a sustained excitation has a gradual onset and continuous energy transfer. An iterative excitation is based on a series of discontinuous energy transfers.

Action–Sound Coupling and Mappings

Sound production on a traditional instrument is bound by the physical constraints of the instrument and the capabilities of human body. For example, although both are plucked instruments, a banjo, and an oud have different damping characters due to the resonant features of the instruments' bodies. The physical properties of the instruments also define their unique timbre and how they are played. Additionally, the human body has its expressive limitations. These limitations can be in

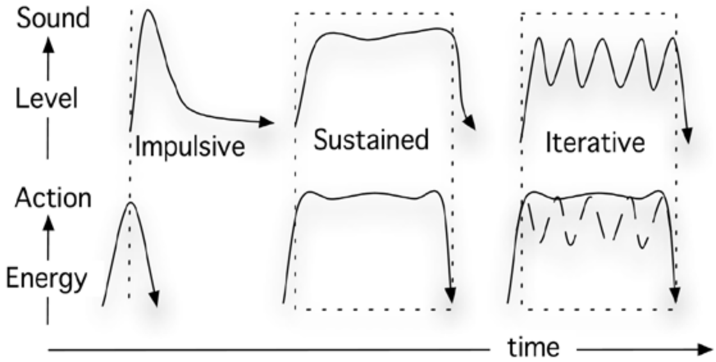


Figure 1. Illustration of the three, basic action–sound types: impulsive, sustained, and iterative (Jensenius, 2007; Used with permission).

the form of what Godøy (2018) suggested as "effort constraints," meaning "limits to endurance," which necessitate an optimization of muscle contractions (i.e., to prevent injuries). He described these limitations as also leading to "coarticulation," which results from multiple individual actions merging into larger units. All these levels of constraints are part of the transformation of biomechanical energy to sound features. We think that during the transformations in *action–sound couplings* (Jensenius, 2007), the relationships between actions and sounds are dictated by the laws of physics.

When playing a traditional instrument, one must exercise muscular exertion to abide by the instrument's physical boundaries. In the case of the guitar, this prevents the player from breaking a string due to excessive effort or not producing sound due to the lack of energy input (Tanaka, 2015a). After centuries of design, the construction of traditional instruments is no longer open to much interpretation, except for using some extended playing techniques or additional equipment. To the contrary, electroacoustic musical instruments are based on the creation of *action–sound mappings*. Here the constraints of hardware and/or software elements often are open to interpretation. In other words, the relationships between biomechanical input and the resultant sound are designed and may not correspond to each other. However, the creation of meaningful action–sound mappings is critical for how an instrument's playing and its sound are perceived (Hunt & Wanderley, 2002; Van Nort, Wanderley, & Depalle, 2014). This is often discussed as the "mapping problem" (Maes et al., 2010), which has been a central research topic in the field of new interfaces for musical expression over the last decades (Jensenius & Lyons, 2017).

New Musical Interactions

The number of artists and researchers interested in using the human body as part of their musical instrument has been growing over the last decades. Such interests often lead to the use of gestural controllers, which are types of wearable sensors or camera-based devices that allow for touchless performance, that is, a type of performance not based on touch of physical objects. As such, these instruments allow for sonic interaction in the air (Jensenius, 2017). Examples of such instruments are the Virtual Air Guitar (Karjalainen, Mäki-Patola, Kanerva, & Huovilainen, 2006), the Virtual Slide Guitar (Pakarinen, Puputti, & Välimäki, 2008), and Google's Teachable Machine, which lets users mimic guitar-playing in front of a web camera (Google, 2020).

The above-mentioned examples focus mainly on creating an air guitar. However, this is not the focus of our current research; rather, we seek to explore new ways of performing in the air. Although motion-based tracking often is employed for air instruments, we are interested specifically in measuring muscle tension through electromyography (EMG). When worn on the forearm, EMG sensors can provide muscle activation information related to the motion of hand and fingers (Kamen, 2013). EMG goes beyond measuring limb positions and provides information of the muscle articulation throughout the preparation for and execution of an action (Tanaka, 2019). The use of muscle activation data in musical performance was pioneered by Knapp & Lusted (1990) and has been practiced extensively by Tanaka (1993, 2015b). Mechanomyograms (MMGs), as a signal for muscle-based performance (Donnarumma, 2015), also have been studied.

Performing in the air introduces several conceptual and practical challenges. For example, when does a sound-producing action begin and end when no physical instrument defines the performance space? How can one handle the use of physical effort as part of that action without being restricted to a physical instrument? To address such problems, we drew on what Tanaka (2015a) suggested as an embodied interaction strategy: He replaced constraints, such as those experienced while playing a traditional instrument, with "restraints," that is, the "internalization of effort" (p. 299). Such restraints can help define a set of affordances that can replace the physical constraints found in a traditional instrument.

Even though we are interested in creating new instrument concepts, this may not necessarily require developing an entirely new action–sound repertoire. Michel Waisvisz, the creator of *The Hands* (Waisvisz, 1985), focused on maintaining the action–sound mappings of his instrument. This helped him develop and maintain a skill set over time. We propose a design strategy based on what Magnusson (2019) referred to as an "ergomimetic" structure. Here *ergon* stands for work memory and *mimesis* for imitation. Such an ergomimetic structure may help in reusing well-known interactions of a performer in a new performative context. Of course, such an approach raises some questions. For example, what types of errors and surprises emerge when a physical pipeline is replaced by software? We aim through our research to contribute to better understanding how a musician's physical skills could transfer to new air instruments.

Machine Learning

Machine learning is a set of artificial intelligence techniques for tackling tasks that are too difficult to solve through explicit programming; it is based on finding patterns in a given set of examples (Fiebrink & Caramiaux, 2016). Deep learning is a subset of machine learning, where artificial neural networks allow computers to understand complex phenomena by building a hierarchy of concepts out of simpler ones (Goodfellow, Bengio, & Courville, 2016). Machine learning has been an important component in the design of and performance with new interfaces for musical expression since the early 1990s (Lee, Freed, & Wessel, 1991). Several easy-to-use tools have been developed over the years for artists and musicians (see, e.g., Caramiaux, Montecchio, Tanaka, & Bevilacqua, 2015; Fiebrink, 2011; Martin & Torresen, 2019), and many new instruments have explored the creative potential of artificial intelligence in music and performance (Caramiaux & Donnarumma, 2020; Kiefer, 2014; Næss, 2019; Schacher, Miyama & Bisig, 2015; Tahiroğlu, Kastemaa & Koli, 2020). However, unlike the applications for generating music in the form of musical instrument digital interface (MIDI) data (Briot, Hadjeres, & Pachet, 2020) or generating music in the wave-form domain (Purwins et al., 2019), the use of deep learning techniques for interactive music is rather rare. We see that deep learning can be particularly useful when dealing with complex muscle signals.

Research Questions

The brief theoretical discussion above has shown that a number of questions remain open regarding how musical sound is performed and perceived and how it is possible to create new empirically based sound-making strategies. Thus, in the current two-experiment study, we were interested particularly in

1. What types of muscle signals are found in electric guitar performance and how do these signals relate to the resultant sound?
2. How can we use deep learning to predict sound based on raw electromyograms?

We begin by explaining the methodological framework that has been developed for the first empirical study, followed by a presentation and discussion of the results. We then reuse some of the data from the first experiment to pursue a preliminary predictive model for action–sound mappings. We conclude with a general discussion of the findings of these two experiments.

EXPERIMENT 1: MUSCLE–SOUND RELATIONSHIPS

Methods

Research Design

This aspect of our research is based on the outcomes of an experiment with electric guitar players. Each of the guitarists performed, while wearing various sensors, a set of basic sound-producing actions as well as free improvisations. To collect the data these actions produced, we built a multimodal dataset of EMG and motion capture data; additionally, video and sound recordings of each performer were made. For this paper, we focus only on a statistical analysis of the EMG data and sound recordings from this first experiment, with a particular emphasis on similarity measures. Prior to conducting the research, we obtained ethical approval from the Norwegian Center for Research Data (NSD), Project Number 872789.

Participants

Thirty-six music students and semiprofessional musicians took part in the study. Five of the datasets turned out to be incomplete and these were excluded from further analysis. Thus, the final dataset consisted of 31 participants (30 male, 1 female, M_{age} = 27 years, SD = 7), all right-handed. All the participants had some formal training in playing the electric guitar, ranging from private lessons to university level education. The recruitment was conducted through an online invitation published on a specified web site of the University of Oslo, Norway, and announced in various communication channels targeting music students. Participation was rewarded with a gift card (valued at approximately €30).

Data Collection

The participants' muscle activity was recorded as surface EMG with two systems: consumer-grade Myo armbands and a medical-grade Delsys Trigno system. The former has a sample rate of 200 Hz, while the latter has a sample rate of 2000 Hz. Overt body motion was captured with a 12-camera Qualisys Oqus infrared optical motion capture system at a frame rate of 200 Hz. This system tracked the three-dimensional positions of reflective markers attached to each participant's upper body and the instrument. A trigger unit was used to synchronize the Qualisys and Delsys Trigno systems. Additionally, we developed a custom-built software solution to capture data from the Myo armbands in synchrony with the audio. Regular video was recorded with a Canon XF105 camera, which was synchronized with the Qualisys motion capture system. Figure 2 demonstrates the two major means for gathering data: the motion-capture configuration and the EMG system.

Procedure

Each participant was recorded individually. One recording session took 90-105 minutes. First, the participants received a brief explanation about the experiment, before they signed the consent form. Following the recording session, they completed a short survey regarding their musical background, their use of musical equipment, and their thoughts on new instruments and interactive music systems.

The participants were instructed to stand at the same marked spot in the laboratory. We asked them to perform tasks based on well-known electric guitar techniques. The hammer-on and pull-off are similar techniques that allow the performer to play multiple notes connected in a legato manner (tied together). In both techniques, the left-hand fingers hit multiple notes with a single excitation action. Hammer-on refers to bringing down another finger with sufficient force to hit a

(a) (b)

Figure 2. **(a)** A participant during the recording session. Motion capture cameras are visible hanging in the ceiling rig behind and on stands in front of the performer. The monitor with instructions for the performer can be seen below the front left motion capture camera. **(b)** The protocol used for placement of the EMG electrodes: Two Delsys EMG sensors were placed on each side of the arm corresponding to the extensor carpi radialis longus and flexor carpi radialis muscles, just below the Myo armbands.

neighboring note on the fretboard. Pull-off refers to moving the finger from one fret to another to modify the pitch. Bending is achieved by a finger pulling or pushing the string across the fretboard to smoothly increase the pitch. The given tasks were as follows:

- A warm-up improvisation with metronome at 70 bpm
- Task 1
  - Softly played impulsive notes B and C in 3rd and 4th octaves, respectively
  - The same task, played strongly
- Task 2
  - Softly played iterative notes
    - Single pitch (B3)
    - Double pitches (B3–C4)
  - The same task, played strongly
- Task 3
  - Softly played legato
  - The same task, played strongly
- Task 4
  - Softly played bending (semi-tone)
  - The same task, played strongly
- A free improvisation (the tone features and the use of metronome are at the participant's discretion)

We based the tasks on performing guitar-like versions of each of the three action–sound types. Tasks 1 and 4, for instance, lie somewhere in between classes considering that the right hand excites the string in an impulsive manner while the left hand keeps sustaining the tone as much as the construction of the instrument allows. In Task 2, participants were asked to alternate between single and double pitches in different takes. Finally, Task 3 presents a hybrid of the impulsive and sustained types. All given tasks focused on the notes B3 and C4 on the D string, played by index and middle fingers.

Each task was recorded as a fixed-form track, 2 min 16 s in duration, along with a metronome click at 70 BPM. The participants were instructed to play for 4 bars, rest for 2 bars, play the variation for 4 bars, rest another 2 bars and repeat this same 12-bar pattern two more times. See Table 1 for a detailed list of finger and style variations. To help the participants perform the tasks correctly, they were standing in front of a custom-built prompter screen. On the screen, they could follow animated circles, which signified the beat and the bar they were supposed to be at with respect to the predefined form of the given task. This allowed for a more comfortable and efficient experiment process. For the pilot study, we used a text-based prompting. However, this increased the cognitive load of the participants. Thus, for the full experiment we implemented a simple geometry-based design.

Table 1. Detailed Fingerings and Playing Styles Instructed to Participants for Particular Tasks.

	Takes 1-3-5	**Takes 2-4-6**
Impulsive	Index	Middle
Iterative	Index	Index–middle
Bending	Middle, as fast as possible	Middle, as slow as possible
Legato	Index–middle, hammer-on	Middle–index, pull-off

Note. Fingering and playing styles were organized based on the odd- and even-numbered takes to have a systematic approach to labeling different action features recorded within a single track. This approach facilitated the groupings of segmented individual takes during the preprocessing step.

Data Acquisition

Figure 3 shows the recording setup, which was based on two separate personal computers running the data collection software. In the first one, we used an external trigger to send the start pulse to the Qualisys motion capture system, which allowed an in-sync recording of the motion capture cameras, the Delsys Trigno EMG sensors, and the Canon video camera. The second computer recorded signals from the Myo armbands and the audio as line input from the guitar amplifier. This was accomplished using a custom-built Python program to record synchronized sensor data and audio. The Myo armbands were interfaced through improving the myo-to-osc framework for the Bluetooth API (Martin, Jensenius, & Torresen, 2018). To overcome possible bandwidth limitations, we implemented low-latency support for the multiple Myo armbands connected to the computer via individual Bluetooth Low Energy adapters. PyAudio was used for the audio recording (Pham, 2006). The Python interface ran as four simultaneous processes: data acquisition from each armband, the metronome, and the audio recording.

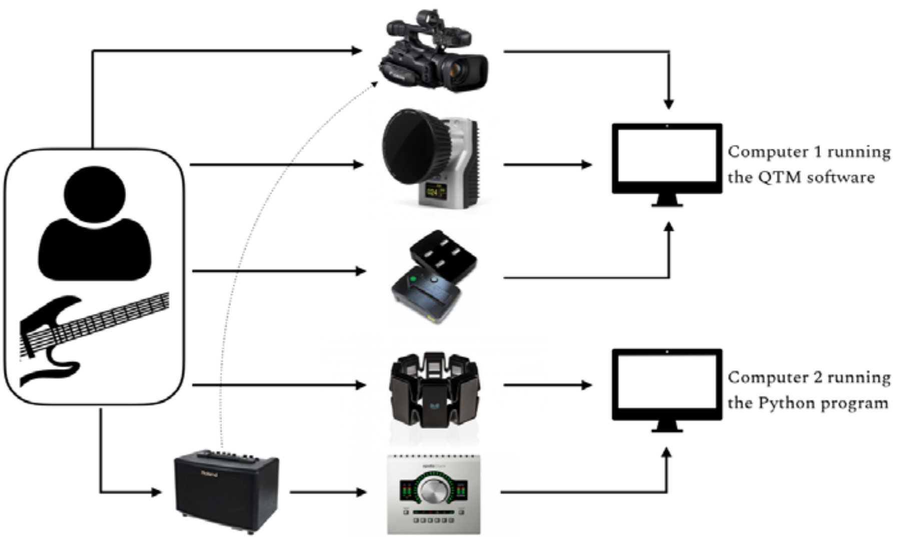


Figure 3. A simplified signal flow diagram of the experimental setup. Representative pictures of the equipment used, from top to bottom: Canon video camera, Qualisys Oqus infrared camera, Delsys Trigno electrodes, Myo armband, and Roland guitar amplifier, and Universal Audio Apollo Twin sound card.

Preprocessing

Preprocessing of our data for further analysis and modeling purposes was handled separately for the data from the Delsys and Myo systems. The medical-grade Delsys system provided high-quality data suitable for analytical purposes, while the Myo is a consumer-grade product that works well for interactive applications (see Pizzolato et al., 2017, for a comparison of various EMG acquisition setups). For the Delsys data, preprocessing included filtering, segmentation, and feature extraction methods. For the Myo data, we worked on interpolation and alignment of the raw data instead.

Synchronization

We synchronized the recorded data and audio through a custom-built metronome script within our Python program. This script recorded the timestamps of the metronome clicks together with the start point of the audio recording in a CSV file. This strategy helped in two ways. First, we could calculate lags at less than 0.1s among the various recording channels. As a result, we could align all the data types, based on their start points, to the metronome timeline. The synchronization strategy also helped in conforming the Qualysis data captured on Computer 1 with the line-audio recordings on Computer 2. Computer 1 ran the Qualisys software, which also recorded a standard video file synchronized with embedded audio.

We first extracted the audio stream from the video recording, and then decomposed the signal into its percussive and harmonic components. Applying an onset detection algorithm on the percussive component made it possible to obtain a timeline of metronome clicks from the ambient audio recording. This allowed us to measure the clicks and compare them to the logged timestamps of the original metronome clicks from Computer 2. Because the Delsys data shared the same timestamps with those of the metronome onsets, and the line audio recording shared the same timestamps with those of the metronome logs, we were able to align all the recorded data and media.

EMG Signal

Drawing on the method proposed by De Luca, Gilmore, Kuznetsov, & Roy (2010), we recorded the raw EMG data at 2000 Hz using the Delsys Trigno system, which were first run through a high-pass filter with a cutoff frequency of 20 Hz, and a low-pass filter with a cut-off of 200 Hz. Both filters were fourth-order Butterworth type (Selesnick & Burrus, 1998). Next, we segmented the synchronized and normalized EMG data into 5-beat sequences (1 bar created from the last beat of the previous bar in the timeline). This was to capture also muscle activation preceding the sound-producing action. The muscle activation necessarily precedes the motion of the hand and the audio onset.

Each task was recorded as a single track that contained six takes (see Table 1). Then, we selected one segment from each of them following this protocol:

1. Takes that featured the index finger on B3 were chosen from the impulsive and iterative tasks. In addition to an effort for narrowing the scope by focusing on the index finger for the impulsive task, we were interested in exploring how two motion types combine in the iterative task.

2. Takes that were played "as slow as possible" were chosen from the bending task. Slow bending (over a period of approximately a bar) is fairly similar to the sustained motion type. The guitar does not actually afford sustained performance in the same way as, for example, a violin does. However, the more the bending is prolonged, the more the damping is shortened. This results in two almost opposing input and output amplitude envelopes. The sustaining muscle amplitude envelope has an increased tension. The sound energy, on the contrary, decays quicker than that of an impulsive attack.
3. Takes that featured the hammer-on technique were chosen from the legato task. We observed that a majority of the participants was more comfortable with the hammer-on technique than a pull-off. This was also something we observed in the recorded data. In addition, hammer-on can be seen as a variation of the impulsive tasks played with both fingers.

Finally, each segment was divided into four EMG channels (i.e., the extensor and flexor muscles of each forearm). This resulted in 992 segments (31 participants, 8 tasks, 4 channels) of EMG data. Each segment had a duration of 4.29 s.

For the feature extraction, we were interested primarily in the amplitude envelopes. This was extracted as the root mean square (RMS) of the continuous signal. The moving RMS of a discrete signal is defined by St-Amant, Rancourt, & Clancy (1996) as

$$\hat{x}_1\,(t) = \left[\frac{1}{N}\sum_{i=t-N+1}^{t} m^2\,(i)\right]^{1/2}$$

where $\hat{x}$ is the EMG amplitude estimate at sample t, using a smoothing window length of N. The recommended window length for calculating the RMS of an EMG signal is 120–300 ms (Burden, Lewis, & Willcox, 2014). After several trials, we noticed that shorter window lengths better covered the peaks of fast attacks. Thus, we used a 50 ms sliding window with 12.5 ms (25%) overlaps.

Muscle onsets were calculated using the Teager-Kaiser Energy (TKE) operation to improve the accuracy of the detection (Li, Zhou, & Aruin, 2007). The TKE operation is defined in the time domain as

$$y(n) = x^2(n) - x(n-1)x(n+1)$$

Audio Signal

The sound analysis was based primarily on the RMS envelopes. Additionally, we computed the spectral centroid (SC) of the sound, as it has been shown to correlate with the perception of brightness in sound (Schubert, Wolfe, & Tarnopolsky, 2004), that is, how the spectral content is distributed between high and low frequencies. The RMS signal is particularly relevant in that our primary interest in this study is in the amplitude envelope of the sound. RMS correlates with perceptual loudness; people can judge whether a signal is loud, soft, or in between but cannot infer where a periodic signal is peaking or is at a zero-crossing (Beranek & Mellow, 2012; Ward, 1971). Thus, for our purposes, RMS served as an appropriate feature, providing more information than simply identifying the peak value within a given time interval.

Analysis

Our analysis focused on exploring similarities between the amplitude envelopes of the EMG signals and the sound. We achieved this by comparing the beginning and the end of the body–sound interactions identified when playing the electric guitar. Muscle activation was observable at the beginning, followed by motion, and then the resulting sound. We conducted the entire analysis through in a custom-built toolbox programmed in Python.

EMG Analysis

The initial component of the EMG analysis focused on exploring the similarities between the RMS of each of the four channels (two per arm) and the sound RMS for each of the participants. We used a Pearson's product–moment correlation, Spearman's rank correlation, and analysis of variance.

Also known as linear correlation coefficient (LCC), Pearson's product–moment correlation is a parametric correlation of the degree to which the change in one variable is linearly associated with a change in another continuous variable. In its equation form, LCC is commonly abbreviated as r while, in our case, x and y represent EMG and audio signals, respectively,

$$r = \frac{\sum(x - \overline{x})(y - \overline{y})}{\sqrt{\sum(x - \overline{x})^2 \sum(y - \overline{y})^2}}$$

where $LCC > 0$ denotes a positive correlation while the opposite ($LCC < 0$) refers to an inverse correlation. The LCC approaches 0 when the correlation weakens. To our knowledge, this measure has not been used to compare audio and EMG signals.

A common assumption of the Pearson's correlation is that the continuous variables follow a bivariate normal distribution. In other cases, where the data is not normally distributed and the relationship of two variables rather seems nonlinear, the Spearman's rank correlation (SCC) is suggested to measure the monotonic relationship (Schober, Boer, & Schwarte, 2018). SCC is fairly similar to LCC, but it calculates the ranks of the pair of values. It is abbreviated as r_s (or ρ) in its mathematical representation where D is the difference between ranks and n denotes the number of data pairs:

$$r_s = 1 - \frac{6\sum D^2}{n(n^2 - 1)}$$

A positive r_s denotes a covariance toward the same direction, whereas a negative r_s refers to fully opposite directions. It is a correlation measure that is commonly used in validating EMG data (Fuentes del Toro et al., 2019; Nojima, Watanabe, Saito, Tanabe, & Kanazawa, 2018).

A third approach was to calculate the pairwise *t* tests and one-way analysis of variance (ANOVA) to explore the variances of correlation values across participants and different dynamics. Here, we tested the assumptions of normality and homogeneity of variances of the independent samples in the dataset using the Shapiro-Wilk and Levene tests (Virtanen et al., 2020), respectively.

In addition to the above-mentioned analysis strategies, we explored other representations of the EMG signals. Inspired by Santello, Flanders, & Soechting (2002) and González Sánchez, Dahl, Hatfield, & Godøy (2019), we applied the time-varying Principal Component Analysis

(PCA) to merge all four channels and investigate prominent features across all participants. The input matrix for the PCA is defined as $A \in \mathbb{R}^{m \times n}$ where m is the number of participants and n denotes the number of EMG channels. For each of the 8 tasks, in which half employed soft dynamics and the other half strong dynamics, we obtained two principal components (PCs), which represented a combination of both excitation and modulation actions on the guitar, as shown by the following equation,

$$EMG_{\mathrm{m}} = \mathrm{meanEMG_m} + PC1 \times EMG1_{\mathrm{m}} + \cdots + PCn \times EMGn_{\mathrm{m}}$$

Additionally, we applied Singular Spectrum Analysis (SSA) to principal components of EMG for further signal–noise separation. SSA is a technique of time series analysis used for decomposing the original series by means of a sliding window into a sum of small number of interpretable components, such as slowly varying trend, oscillatory (periodic) components, and structureless noise (Golyandina & Zhigljavsky, 2013). The algorithm for SSA is similar to that of PCA in multivariate data. In contrast to the PCA, which is applied to a matrix, SSA provides a representation of the given time series in terms of a matrix made of the time series (Alexandrov, 2009). In this way, we applied SSA on the EMG principal components and extracted the trend, which is a smooth additive component that contains information about the time series' global change (Alexandrov, Bianconcini, Dagum, Maass, & McElroy, 2012). This procedure allowed us to obtain better visualizations of the nonlinearity of relationships between EMG and audio waveforms.

It should be noted that researchers in the literature have suggested a variety of specialized methods for choosing the SSA window length (L). Knowing that it is highly difficult to define a universal method to find an optimal L value for an arbitrary time series and that the practitioners should therefore investigate this issue with care, Khan & Poskitt (2011) suggested a rule as $L = (\log N)^c$ with $c \in (1.5, 3.0)$ for assigning a window length that will yield near optimal performance. Starting from there, as the RMS segments of our interest were at a fixed length of $N = 344$, we empirically chose $c = 2.5$, which yielded $L = 10$.

Video Analysis

We used the Musical Gestures Toolbox (Jensenius, 2018b) to extract the sparse optical flow from the video recordings, with the goal of identifying to what extent participants moved unintentionally. This information allowed us to make comparisons with other data at hand and open a better understanding of unexpected muscle activations.

Sound Analysis

Our aim in the sound analysis was to quantify how the different dynamics influenced the overall brightness of the sound. To this end, we averaged the SC across all participants. Note that the sound data in this study is presented in approximately 4.29 s chunks. However, we also investigated chunks of a shorter duration in order to explore whether dynamic fluctuations of particularly the iterative task had an effect on the mean brightness. Moreover, considering the damping character of the guitar, which is relatively short in duration, we explored how decay times influenced the overall brightness value.

Results

The 36 participants completed 360 tasks in total. However, we excluded five datasets due to incomplete data. After also excluding the improvisations—which were intended to be used in the modeling experiment detailed below—we analyzed 248 tasks from 31 participants. An overview of how muscle activation patterns transform to sound features in each task is illustrated in Figure 4.

LCC and SCC

The correlation coefficients among participants were computed using the LCC and SCC measures. Table 2 shows positive correlation, negative correlation, mean, and standard deviation for each factor. Figures 5 and 6 show the distribution of LCC and SCC correlations.

The analysis shows to what extent the muscle activation underlying the sound-producing motion and the resultant sound on the same musical instrument can have similar amplitude envelopes. This is supported by the ANOVA results. The correlation of muscle–sound amplitude envelopes—whether positive, negative, or close to 0—does not exhibit a noteworthy variance between participants. That is, the ANOVAs for EMG–sound similarities across participants (for all EMG channels and tasks) are as follows: LCC, $F(30,961) = 1.6$, $p = 0.02$, and SCC, $F(30,961) = 1.59$, $p = 0.02$.

The comparisons of the correlation values between left and right hands supports the functional distinction between the right and left actions (see Table 3). Another clear distinction was revealed when we compared to what extent the EMG and sound envelopes correlated with respect to soft and strong dynamics (see Table 4). When the participants played strongly, the muscle and resultant sound amplitude envelopes correlated better.

PCA and SSA

Figure 7 shows the waveforms of the two principal components of the combined EMG channels across all participants for impulsive, iterative, bending, and legato tasks, separately for soft and strong dynamics. Each panel shows the activation patterns for the characteristics of these tasks.

The trends of the same principal component waveforms via signal–noise separation were extracted using SSA ($L = 10$) and have been plotted against the averaged sound RMS on the horizontal axis in Figure 8. Here we can observe the varying level of nonlinearities of the muscle–sound relationship for the tasks played at different dynamic levels.

Spectral Centroid

Figure 9 shows the distribution of the SC of the sound across all participants for each soft and strong task, separately. Although stronger dynamics show a clear strength in the upper end of the sound spectrum, the distribution among particular tasks varied depending on the chosen timescale. As such, SC values of all tasks with soft dynamics ($M = 299.03$, $SD = 124.24$), compared to the SC values of tasks with strong dynamics ($M = 585.93$, $SD = 141.22$), demonstrated significantly lower mass of the spectrum, $t(246) = 16.98$, $p < .001$

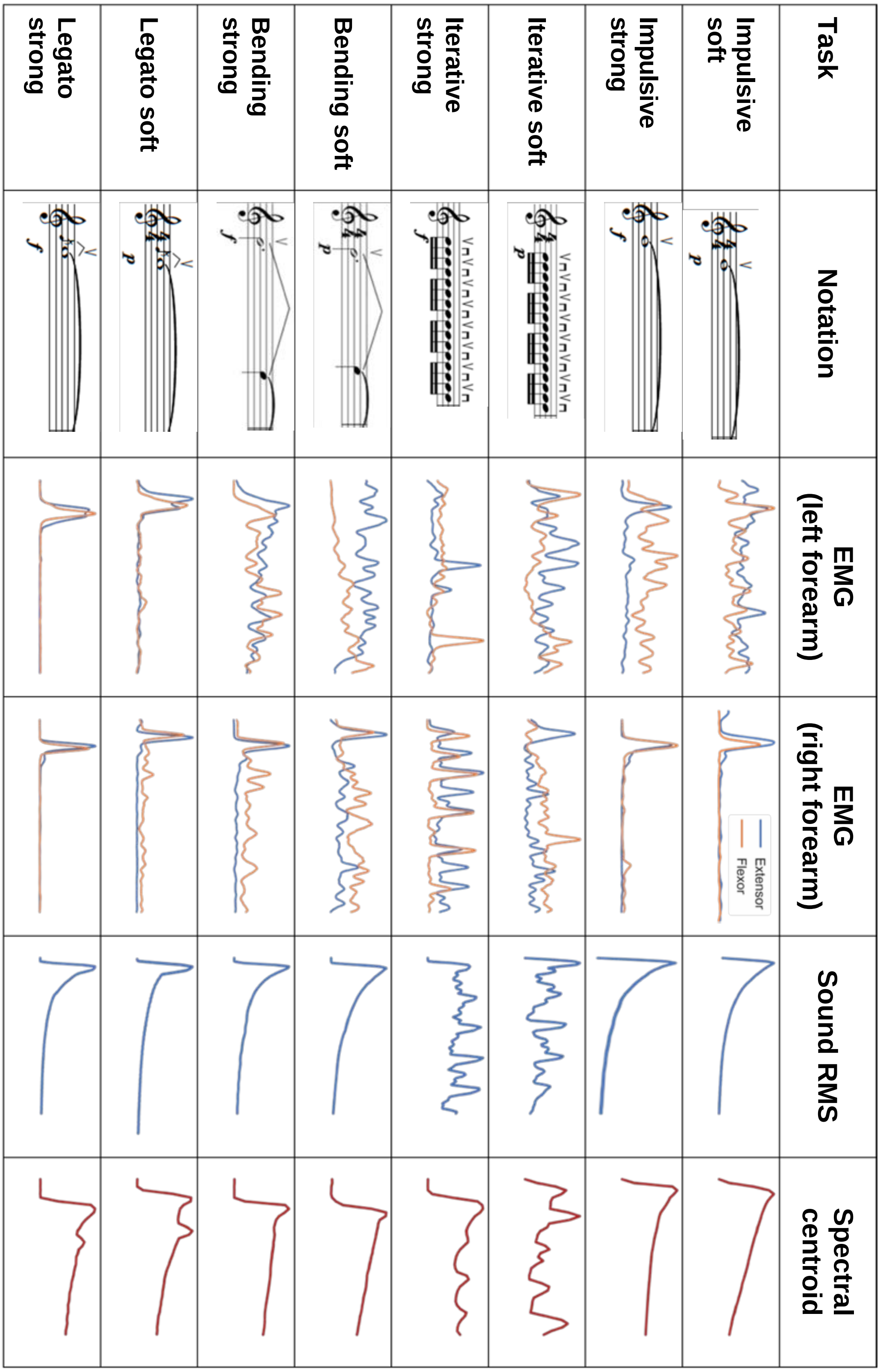


Figure 4. An overview of how notated music transforms into an audio waveform when playing the electric guitar. Trends of signals were extracted using Singular Spectrum Analysis (SSA) with a window length $L = 10$.

Table 2. Correlation Coefficients for Each Factor (LCC and SCC): The Positive, Negative, Mean and Standard Deviation of Correlation Coefficients.

			Impulsive soft	Impulsive strong	Iterative soft	Iterative strong	Bending soft	Bending strong	Legato soft	Legato strong
LCC	r	Extensor (right)	0.66	0.59	0.64	0.68	0.60	0.73	0.46	0.53
		Flexor (right)	0.65	0.54	0.51	0.86	0.65	0.69	0.42	0.55
		Extensor (left)	0.72	0.62	0.74	0.64	0.63	0.76	0.44	0.60
		Flexor (left)	0.55	0.55	0.65	0.65	0.48	0.63	0.51	0.48
	$-r$	Extensor (right)	–0.24	–0.03	–0.24	–0.24	–0.12	–0.10	–0.38	–0.24
		Flexor (right)	–0.34	–0.25	–0.10	–0.07	–0.34	–0.10	–0.33	–0.32
		Extensor (left)	–0.66	–0.61	–0.35	–0.35	–0.51	–0.66	–0.35	–0.33
		Flexor (left)	–0.62	–0.62	–0.53	–0.51	–0.54	–0.46	–0.30	–0.53
	μ	Extensor (right)	0.17	0.24	0.28	0.33	0.26	0.28	0.00	0.09
		Flexor (right)	0.13	0.23	0.22	0.33	0.21	0.27	0.02	0.03
		Extensor (left)	–0.23	–0.08	0.21	0.25	0.18	0.22	–0.02	0.01
		Flexor (left)	–0.34	–0.24	0.20	0.21	0.03	0.15	–0.01	–0.02
	σ	Extensor (right)	0.23	0.14	0.17	0.18	0.18	0.19	0.15	0.20
		Flexor (right)	0.25	0.17	0.17	0.19	0.21	0.17	0.13	0.18
		Extensor (left)	0.35	0.36	0.26	0.23	0.27	0.24	0.16	0.16
		Flexor (left)	0.28	0.25	0.28	0.20	0.14	0.22	0.14	0.12

(continued)

Table 2. Correlation Coefficients for Each Factor (LCC and SCC): The Positive, Negative, Mean and Standard Deviation of Correlation Coefficients. (continued)

			Impulsive soft	Impulsive strong	Iterative soft	Iterative strong	Bending soft	Bending strong	Legato soft	Legato strong
SCC	r_s	Extensor (right)	0.66	0.71	0.68	0.71	0.58	0.78	0.55	0.61
		Flexor (right)	0.49	0.71	0.58	0.74	0.66	0.74	0.27	0.66
		Extensor (left)	0.65	0.84	0.77	0.81	0.81	0.84	0.66	0.42
		Flexor (left)	0.70	0.70	0.69	0.63	0.43	0.70	0.43	0.34
	$-r_s$	Extensor (right)	−0.45	−0.15	−0.25	−0.30	−0.14	−0.17	−0.42	−0.33
		Flexor (right)	−0.41	−0.43	−0.18	−0.04	−0.41	−0.19	−0.19	−0.42
		Extensor (left)	−0.85	−0.89	−0.56	−0.56	−0.61	−0.85	−0.32	−0.61
		Flexor (left)	−0.77	−0.78	−0.50	−0.50	−0.62	−0.78	−0.55	−0.61
	μ	Extensor (right)	0.08	0.27	0.25	0.41	0.27	0.35	−0.01	0.10
		Flexor (right)	0.07	0.26	0.17	0.38	0.18	0.37	0.01	0.02
		Extensor (left)	−0.27	−0.08	0.27	0.35	0.19	0.25	0.00	0.00
		Flexor (left)	−0.38	−0.26	0.21	0.29	0.04	0.17	0.00	0.00
	σ	Extensor (right)	0.22	0.19	0.20	0.23	0.15	0.25	0.14	0.25
		Flexor (right)	0.24	0.21	0.19	0.19	0.18	0.25	0.12	0.20
		Extensor (left)	0.40	0.46	0.31	0.23	0.30	0.24	0.14	0.14
		Flexor (left)	0.31	0.31	0.31	0.23	0.16	0.26	0.13	0.10

Note. The zeros in the table represent rounded values that were smaller than three decimal places, thus a “close-to-zero” correlation.

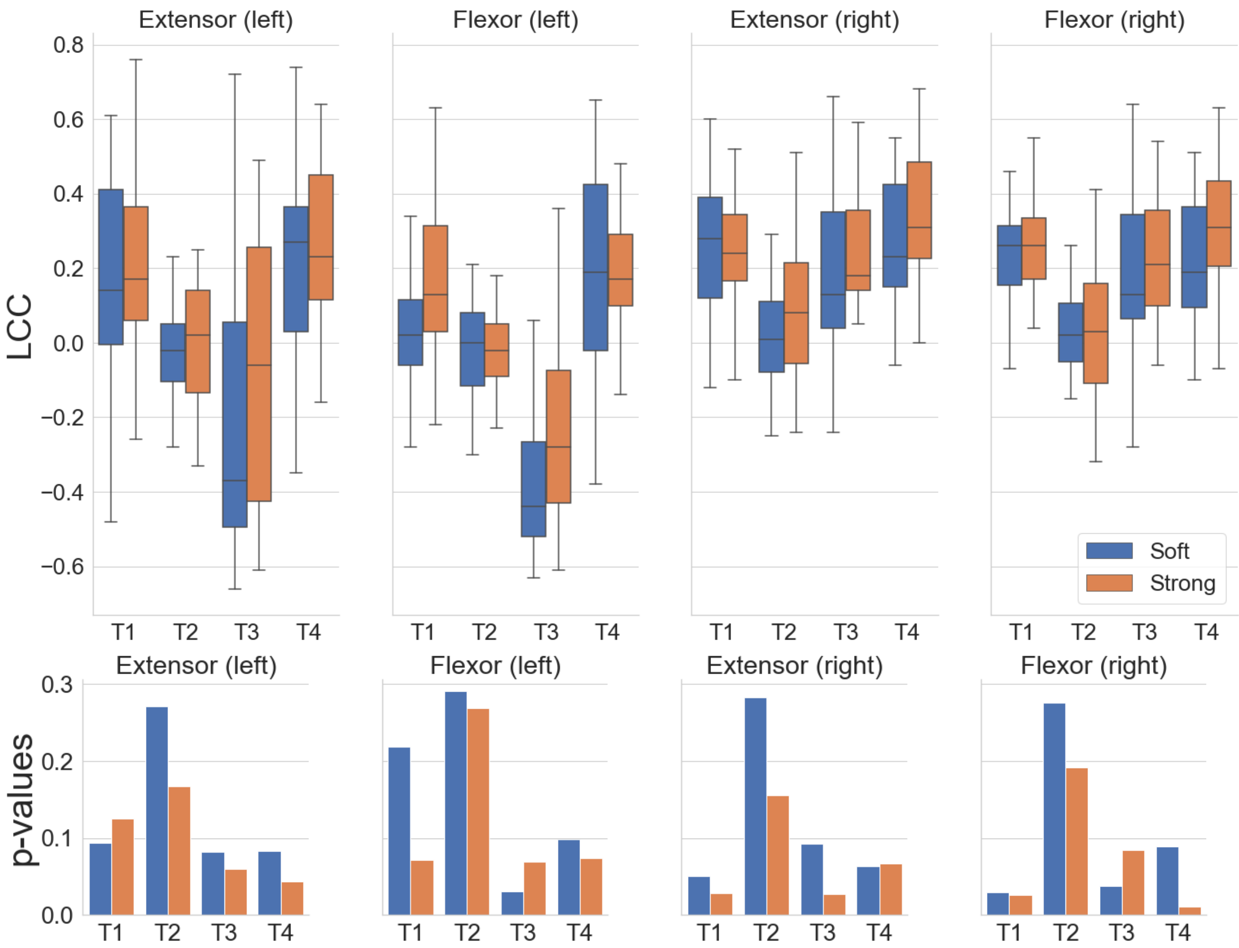


Figure 5. Pearson's product–moment correlations between EMG and Sound RMS envelopes. LCC > 0 denotes a positive correlation while LCC < 0 refers to the negative. The box plots show the interquartile ranges of correlation distribution per task, separately for soft and strong dynamics. The bar plots below show the distribution of *p*-values showing the significance of the correlations. T1, T2, T3 and T4 refer to impulsive, iterative, bending and legato tasks, respectively.

Table 3. Pairwise *t* tests Demonstrating How Modification (Left Forearm) and Excitation (Right Forearm) Actions Have Distinct EMG–Sound Amplitude Envelopes.

	Modification action	Excitation action	Variance
LCC	$M = 0.03$, $SD = 0.30$	$M = 0.19$, $SD = 0.21$	$t(495) = 11.41$, $p < .001$
SCC	$M = 0.05$, $SD = 0.34$	$M = 0.20$, $SD = 0.24$	$t(495) = 9.04$, $p < .001$

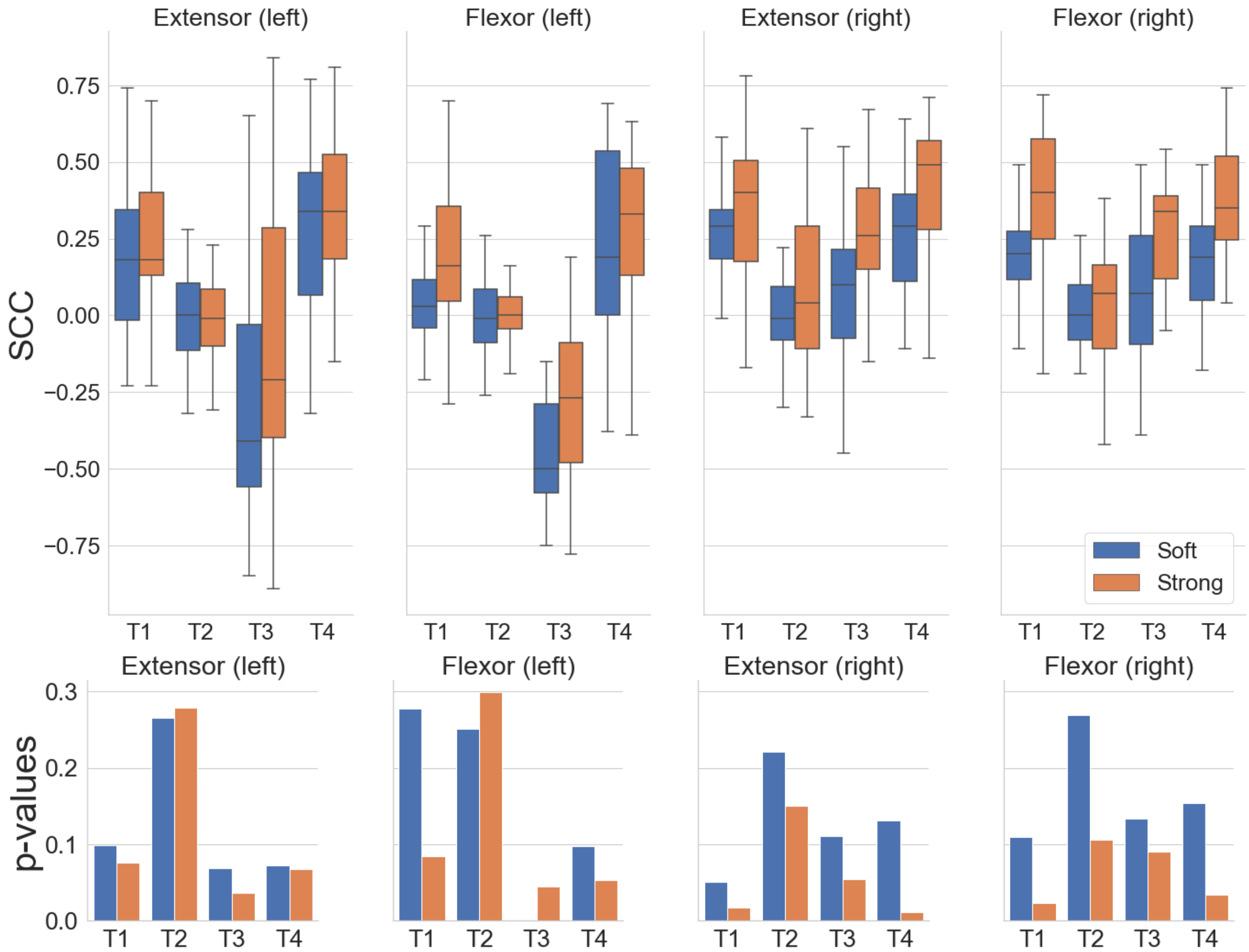


Figure 6. Spearman's rank correlations between EMG and Sound RMS amplitude envelopes. SCC > 0 denotes a covariance in the same direction while SCC < 0 refers to the opposite direction. The box plots show the interquartile ranges of correlation distribution per task, separately for soft and strong dynamics. The bar plots below show the distribution of *p*-values showing the significance of the correlations. T1, T2, T3 and T4 refer to impulsive, iterative, bending and legato tasks, respectively.

Table 4. Means, Standard Deviations and *t*-scores for LCC and SCC Metrics.

	Soft	**Strong**	**Variance**
LCC	$M = 0.08$, $SD = 0.27$	$M = 0.14$, $SD = 0.26$	$t(495) = 5.41$, $p < .001$
SCC	$M = 0.07$, $SD = 0.29$	$M = 0.18$, $SD = 0.31$	$t(495) = 8.33$, $p < .001$

Note. Pairwise t-tests show EMG–sound amplitude envelopes correlations between soft and strong dynamics.

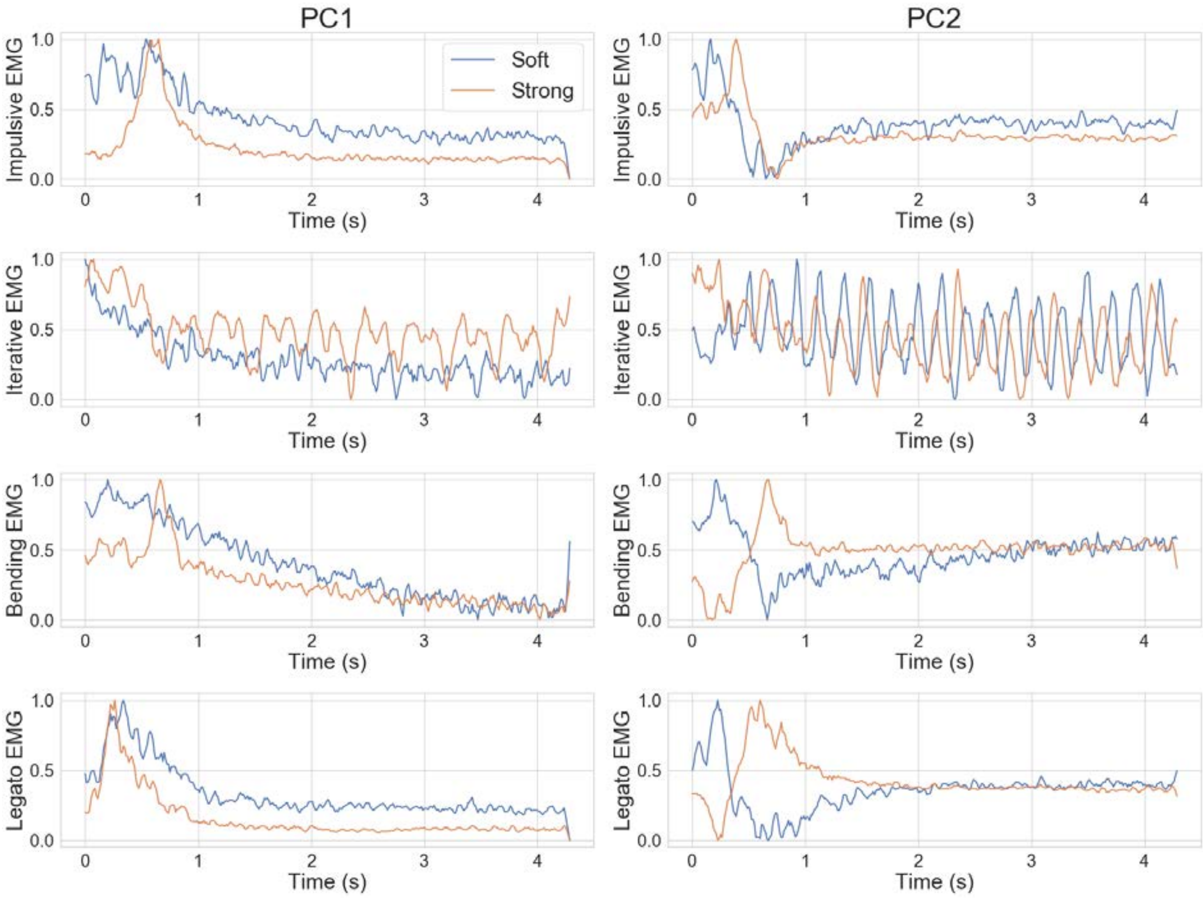


Figure 7. Two principal components (PC1 and PC2) of the combined left and right forearm EMG data of all participants rescaled to (0,…,1) (See the text for more information about the PCA analysis).

Discussion

The analyses showed that sound production on musical instruments is a phenomenon that involves many physical and physiological processes. For example, Figure 10 shows the activation patterns of the extensor and flexor muscles during down- and up-stroking using a plectrum. This figure illustrates only two muscles groups from the right forearm. However, a musical note often is produced as a more complex combination of both arms, as shown in Figure 4.

Similarity Between EMG and Sound Shapes

Our experiment results show that the relations between the muscle energy envelope and the envelope of the resultant sound have similarities between participants. The results show a significant variance when comparing attacks with soft and strong dynamics using pairwise *t*-tests (Table 4). As shown in Figures 5 and 6, the correlation values are higher, and the directionality is more apparent when the same task is played with strong dynamics. This may be due to two factors. First, greater energy input results in larger sound amplitude, which is less biased to base noises, such as the inherent postural instability of the human body.

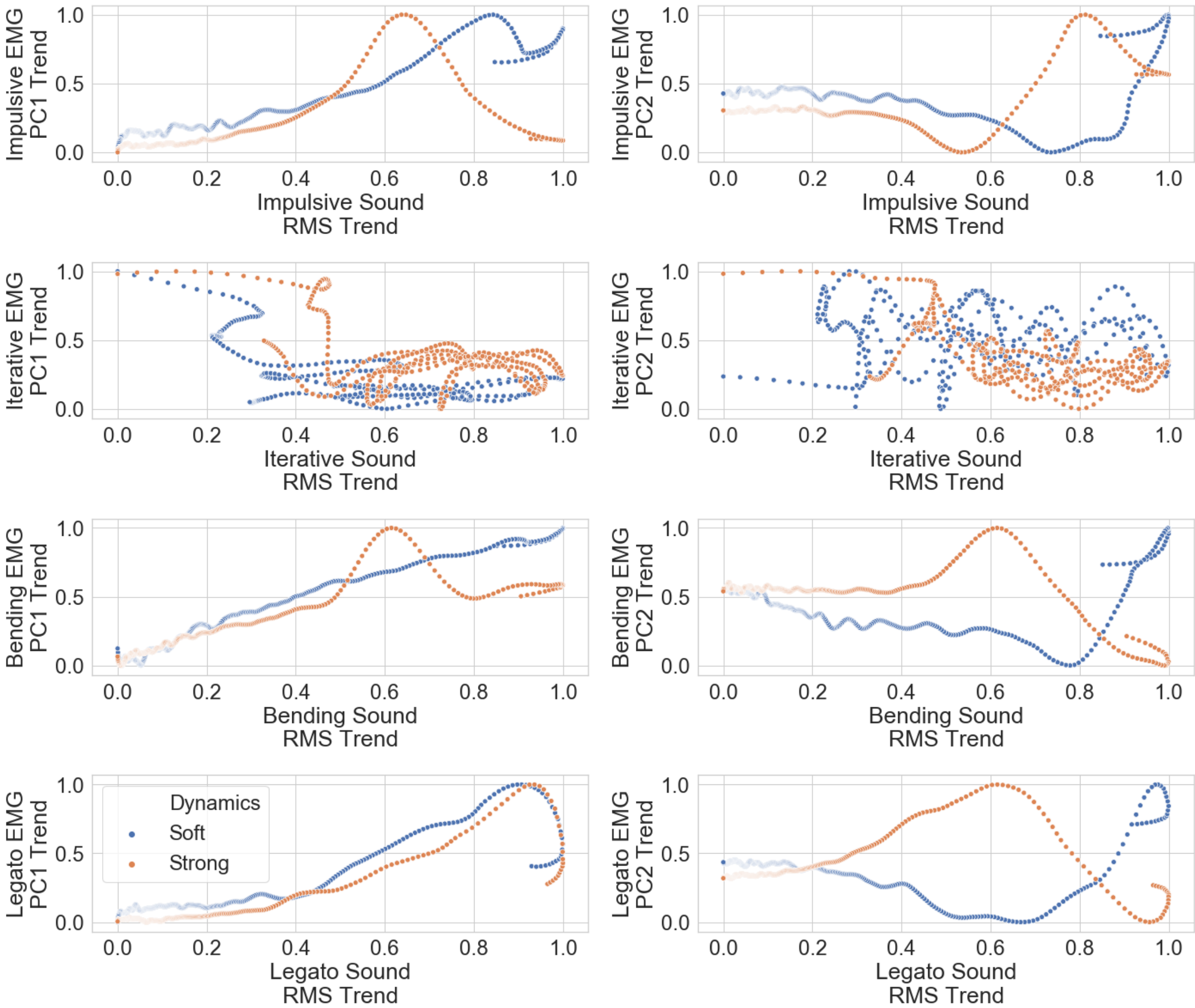


Figure 8. Decomposed principal components (PC1 and PC2) against resultant Sound RMS of all participants (SSA window length $L = 10$). The plots show to what extend the EMG and resultant sound RMS envelopes have a linear relationship at every time step.

Second, we know that expert players tend to use less tension in the forearm muscles (Winges, Furuya, Faber, & Flanders, 2013). Most of our participants can be considered semiprofessionals and thus may have felt less comfortable with stronger dynamics. As a result, they may have employed forearm muscles more explicitly. Unfortunately, we do not have data to check this hypothesis.

The results in Table 3 are in line with the conceptual distinction provided in our Introduction. The excitation action, which typically is performed by the right arm for right-handed players, determines the main characteristics of the resultant sound amplitude envelope. The difference between the activation patterns of both forearms is also observable in Figure 4. The impulsive tasks noted on the top two rows, for example, show the right forearm muscles have envelopes similar to that of the resultant sound while the activation patterns from the left forearm seem to resemble a continuous sound envelope, somewhat between the sustained and iterative types. This is due mainly to a continuous effort exerted by the left forearm over the period of the given task, which is different from the right forearm that excites the string once,

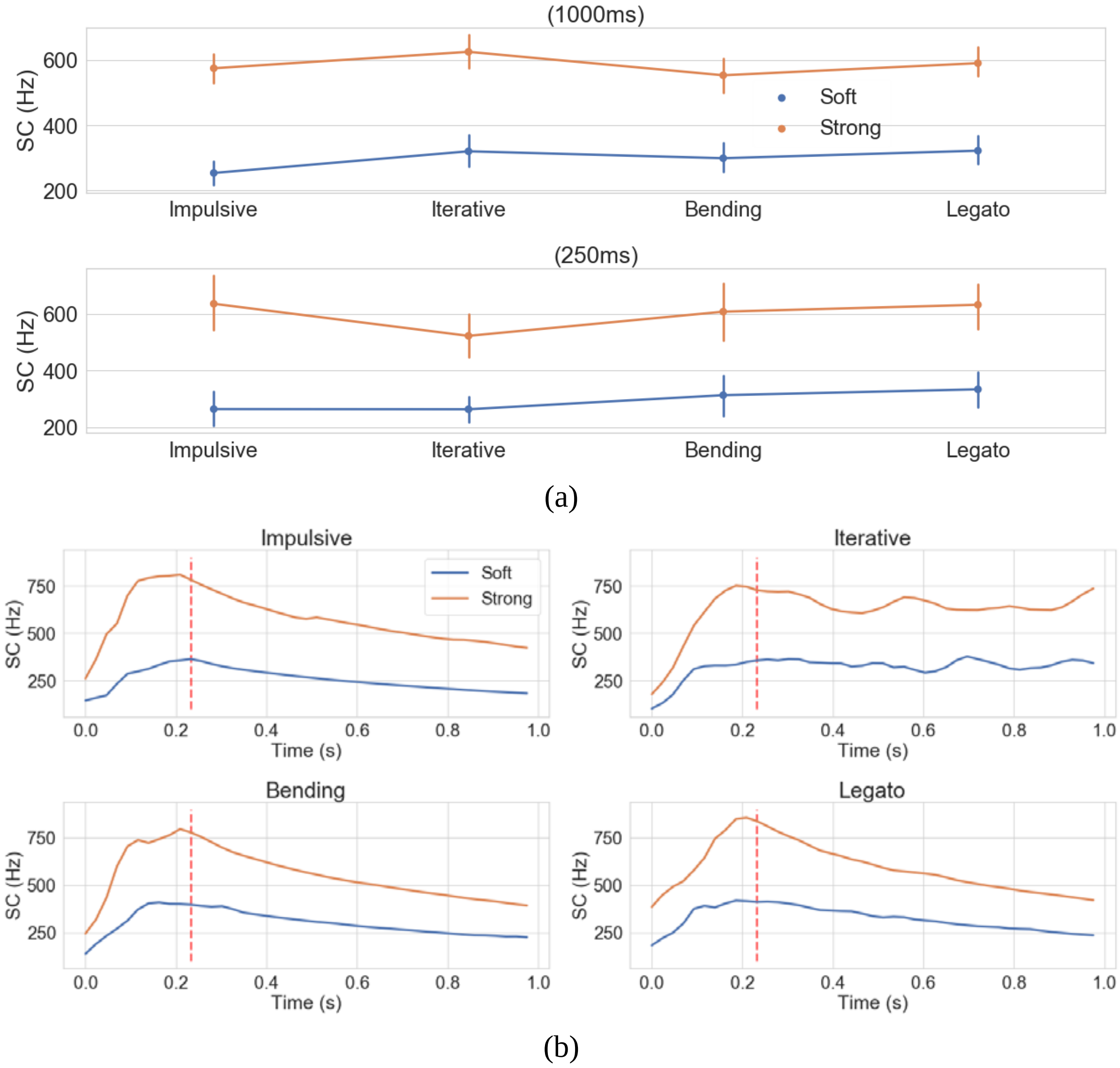


Figure 9. Spectral centroid (SC) of the resultant sound (a) SC distribution between soft and strong dynamics in chunks of 1000 ms and 250 ms duration. (b**)** SC envelopes averaged across all participants. The red vertical lines on the left sides of the plots show the cut point of 250 ms. Note that the segments are 1 s long, which is different than 4 s segments that we initially used. Doing so removed most of the decay that contributes to mean SC.

exerting effort for just a short period. During continuous exertion, we see that bioelectric muscle signals do not exhibit a smooth trend yielding a nearly iterative shape.

Furthermore, any additional ancillary motion, such as moving parts of the body to the beat, or a further modification motion, such as a vibrato to add expression to the sustaining tone, also can be considered as possible artifacts contributing to the envelope of muscular activation. When inspecting the individual participants' video recordings, we noticed that such spontaneous motions are fairly common. Figure 11 provides an example of this. We extracted the sparse optical flow by tracking certain points on a close-up video recording of a participant playing the impulsive task. The participant's ancillary motion is observable in the position of the guitar in relation to the camera and captured possibly by the EMG sensors on the left forearm.

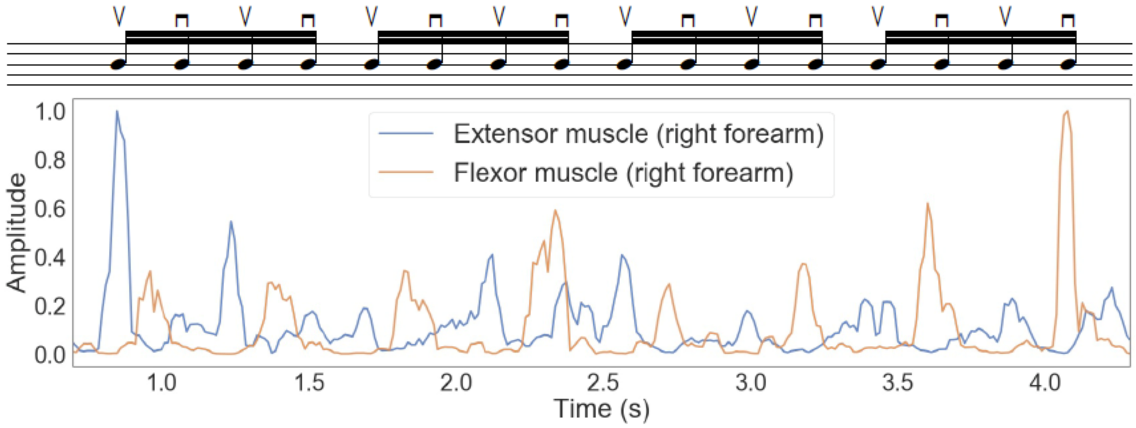


Figure 10. EMG amplitude of the excitation motion during iterative task demonstrating distinct activation of extensor and flexor muscles for down and up strokes, respectively, during a series of 16th notes.

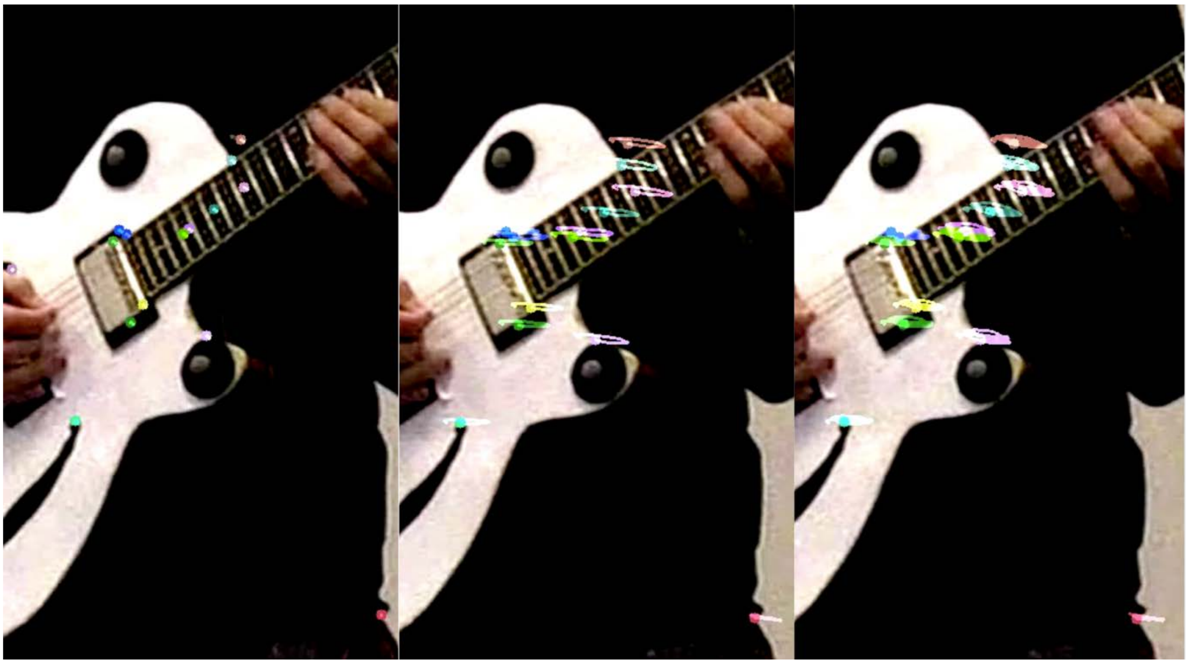


Figure 11. The sparse optical flow shows the trajectory of multiple points on a close-up video segment while a participant is performing an impulsive task. Three subsequent screenshots demonstrate the ancillary motion reflected on the guitar over the period of 1 bar (~3.43 s). The multicolored points on the left picture yield certain patterns in their trajectories reflecting participant movement patterns in the center and right pictures.

We suggest that such ancillary motion influences more directly the ongoing muscle activation as compared to right forearm muscles, which were resting at that moment.

When comparing left and right forearm muscle activation patterns, the negative directionality is noteworthy. This is particularly clear during the bending tasks (see Figures 5 and 6), a playing technique in which the right arm excitation is equivalent to the impulsive task. The left arm modifies the pitch and has a sustained envelope. This is unique to the guitar, as this instrument does not afford sustained sound as do the bowed strings instruments. We should also mention that both the exerted effort and the resultant damping character of the sound would be different if other equipment were used, such as a harder wood and/or pickups with stronger magnets in instrument design, high-gain amplifiers, electronic effects units, or any other room acoustics resulting in greater feedback.

Another interesting observation when comparing data from the left and right forearms is the similarity between positive correlation values of the Impulsive and Legato. This could result from coarticulation. In this task, the left hand executes two consecutive (impulsive) attacks. These are quite different from the impulsive task, however. Because the two consecutive attacks are close temporally, they merge to form one large, coarticulated shape.

Finally, the iterative tasks showed the most idiosyncratic patterns and the least shape similarity. We observed that playing consecutive notes as a series of relatively fast attacks was the most challenging task for many of our participants. Depending on the level of expertise, each participant demonstrated signs of slogging to some extent, which arguably resulted in unique timing characteristics. Effort constraints may be a relevant topic here: Although some players are able to optimize their muscle contractions, others can exert more or less than optimal effort. In addition to the participants' level of expertise, the iterative task may have led to muscle fatigue. None of the participants mentioned this condition, but the possibility deserves further exploration in the context of musical performance.

Exploring Dimensions

The main objective of this investigation was to explore the quantifiable similarities of the amplitude envelopes of sound-producing actions on the electric guitar. In the first part of our analysis, we explored such relationships between two muscle groups against the resultant sound amplitude envelopes from each participant. In the second, we focused on a combination of results from all muscles on both forearms across all participants. We performed PCA on concatenated EMG channels, aiming to render additional observations and visual perspectives. In this part of the analysis, then, we aimed at exploring the signal PCs that can reflect a combination of simultaneous processes. Our interpretation of the PCA is that although PC1 reflected the overall dissipating aspect of the excitation motion, PC2 revealed the variation in the energy input of the modulation motion. This is the case even though we did not specify the decomposition to be separate.

From these observations, we can group all types of EMG patterns under two conceptual categories: (a) impulsive, where a single impulse or a series of impulses is applied, and (b) sustained, denoting a constant muscle energy. The experimental approach of decomposing the PCs using SSA (Figure 8) provided alternative perspectives for exploring the nonlinearities of the relationships. Whereas series of impulses yielded fewer regular patterns, sustaining energy showed clearer similarities. These findings are in line with the results presented in the previous subsection.

The Resultant Sound

Figure 9a demonstrates how SC was distributed across various tasks and dynamics. The main observation here was that stronger dynamics led to a brighter sound. We also should note that plucked strings have what may be called incidental nonlinearities that can have effects, depending on the intensity of excitation (Fletcher, 1999). Moreover, we used 1000 ms and 250 ms segments in these two subplots, respectively. These durations were different from the approximately 4.29 s segments we relied on in our analysis. This shift was intended to remove the tail of the waveform during the decay, which affects the mean brightness value. So, our results support previous work suggesting that timescales shorter than 500 ms reflect most of the timbral features that happen during the attack phase of the excitation (Godøy, 2018).

Figure 9a shows how Iterative had a brighter character than the others when the averaged segments are a longer duration (1000 ms). However, Iterative's mean SC decreased when shorter segments (250 ms) were used for comparison. This indicated a timbral difference between the impulsive and iterative tasks. That is, the impulsive tasks tended to demonstrate a single peak in the exerted energy, reflecting in a brighter sound. The series of attacks of the latter, however, showed more fluctuating energy. This also revealed that during those series, the energy that was transduced into the attacks also made the SC change dynamically. As such, the plots of the averaged SC shaped over time (Figure 9b).

EXPERIMENT 2: A PRELIMINARY PREDICTIVE MODEL

Following the empirical exploration of how biomechanical energy transforms into sound, we used these transformations as part of a machine learning framework based on a long short-term memory recurrent neural network for action–sound mappings. We engaged an interdisciplinary approach that draws on a combination of sound theory and embodied music cognition. Our starting point involved an idea of developing a model that is trained solely on fundamental sound-producing action types. The aim this component of our research was to predict the sound amplitude envelopes of a freely improvised performance. We see this as a preliminary step toward designing an entirely new instrument concept.

Conceptual Design

Our motivating concept was to develop a model that allows for coadaptation, meaning the system not only learns from the user but the user adapts to the behavior of the system (Tanaka & Donnarumma, 2018). Knowing that EMG is a stochastic and nonstationary signal (Phinyomark, Campbell, & Scheme, 2019), even simple trigger actions are quite complex in nature. Although it may seem handy to use well-known machine learning methods, such as classification for triggering sounds or regression to map continuous motion signal (Caramiaux & Tanaka, 2013), we are interested in developing beyond a one-directional control. This vision is conceptually different from, for example, using machine learning for EMG-based control aimed at prosthetic research (Jaramillo-Yánez, Benalcázar, & Mena-Maldonado, 2020).

We also were intrigued with another design concept: predictive modeling. Following various control structures that we had explored in previous work (Erdem, Camci, & Forbes, 2017; Erdem & Jensenius, 2020; Erdem, Schia, & Jensenius, 2019), we were interested more with the ways of how the system can behave differently from interactive music systems that react primarily to the user (Rowe, 1992). Drawing on the work of Martin, Glette, Nygaard, & Torresen (2020), we began exploring the potential of artificial intelligence tools generally, and predictive models in particular, that facilitate not only the input–output mapping of complex signals in new instruments but also enable self-awareness.

Methods

Data Preparation

Our modeling process relied heavily on data from Myo armbands, as they are a cheaper and more portable solution than the Delsys Trigno system. As described in detail in the Methods section of Experiment 1, we synchronized the EMG data and audio arrays based on the recorded metronome timeline. The primary difference in our analysis procedure in this experiment was that we kept all data for modeling. That is, the data were not segmented nor did we eliminate the material collected in-between tasks, when the participants were waiting for the next instruction. This latter set of material made it possible to have the model learn to distinguish between rest and motion states.

We applied linear interpolation to the EMG data and calculated the RMS from the audio signal. The data preparation process resulted in eight segments per participant of EMG and audio data as training examples. The preliminary architecture focused on mapping the raw EMG data to the RMS envelope of the sound as the target.

Predictive Model

We used nine model configurations based on a long short-term memory (LSTM) recurrent neural network (RNN) architecture. Drawing on previous research that suggested 32 or 64 LSTM units in each layer as the most appropriate for integrating the model into an interactive music system (Martin & Torresen, 2019), we wanted to test different configurations. Thus, we used models with one, two, and five hidden layers and each containing 16, 32, and 64 units. Each model was trained on sequences that were 50 data points. This window size refers to 250 ms at Myo armband's 200 Hz sample rate.

Following the LSTM layer(s), a fully connected layer passes a single data point into the activation layer, using a rectified linear activation (ReLU) function. From there, a final layer returns the mean value of the input tensor in order to map an EMG window to one data point of the sound RMS, a many-to-one sequence modeling problem. In short, an array of raw EMG signal with a dimensionality of (50,16) was fed into the network as sliding windows (e.g., sample N_0 to N_{49}, sample N_1 to N_{50}, etc.) to predict a single value of sound RMS at a time step (see Figure 12 for a simplified diagram). The training loss function was defined as

$$\mathcal{L}(x_{\mathrm{RMS}}, \hat{x}_{\mathrm{RMS}}) = \frac{1}{n}\sum_{i=1}^{n}(x_{\mathrm{RMS,i}} - \hat{x}_{\mathrm{RMS,i}})^2 ,$$

where x_{RMS} are the recorded values, $\hat{x}_{\mathrm{RMS}}$ are the values to be predicted, and the sliding window has size n.

Training

The dataset was limited to 160 training examples from 20 participants in which 40 examples were used for validation. We conducted the training using the Adam optimizer (Kingma & Ba, 2014) with a batch size of 100. As we executed multiple trainings to test various configurations, we limited the trainings to 20 epochs. The duration of trainings varied from 4 to 10 hours, depending

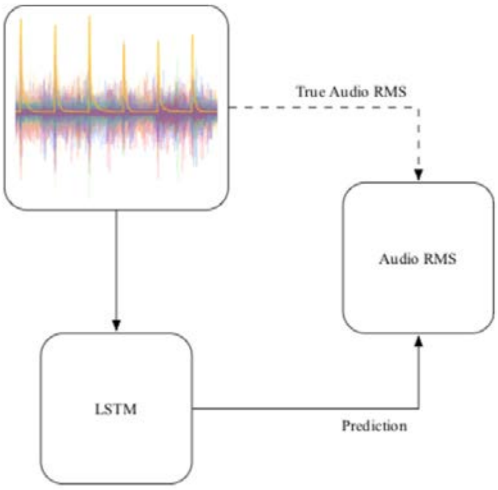


Figure 12. Sketch of the training model: A 16-channel Raw EMG as the source and sound RMS as the target data are passed into an LSTM cell, which then outputs a prediction.

on the quantity of trainable parameters in relation to the number of hidden layers and units. Even though we report here the final results from training locally on a single Nvidia GeForce GTX 1080Ti graphics processing unit (GPU), we also ran the trainings on Google's browser-based coding notebook, *Colaboratory;* we did not observe any remarkable difference in the training duration.

Results

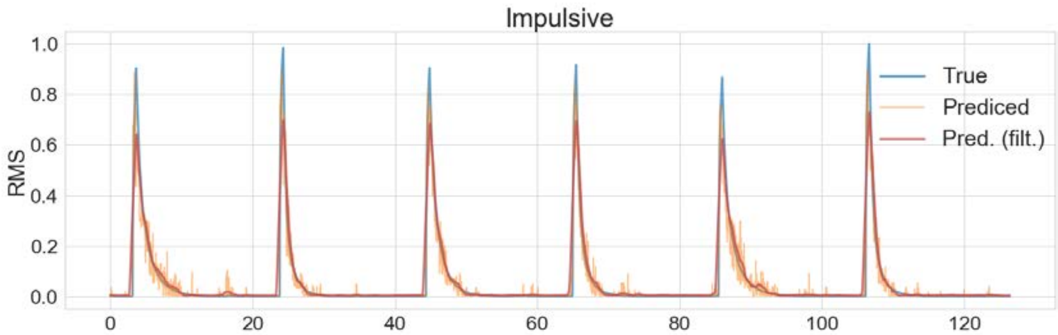
All model configurations were generally capable of predicting the sound RMS (see Figure 13). The model with two hidden layers and 64 units had the best results, which can be seen in the figures of recorded versus predicted RMS of the impulsive (Figure 13a) and iterative tasks (Figure 13b). For the latter, the model could generate similar consecutive envelopes resembling a series of attacks.

One goal in developing this preliminary model was to test the performance of the LSTM based on a limited dataset. In this case, the limitation refers to the type of dataset rather than its size. We were encouraged to see that the model could predict the general trend of the sound energy when tested using the free improvisation dataset (Figure 14).

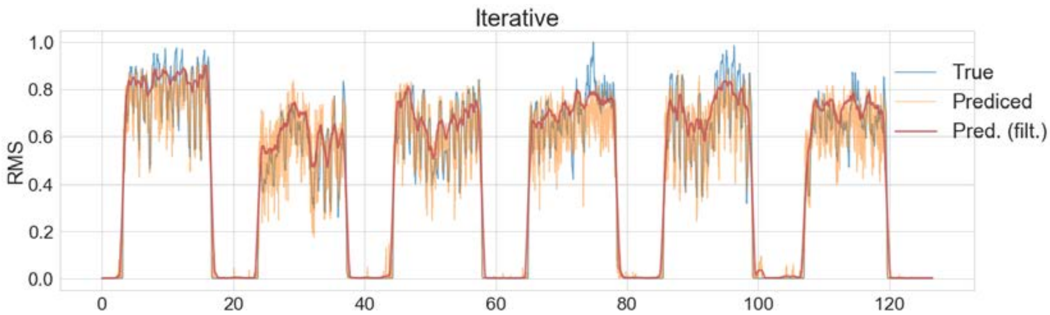
The prediction of the bending task brought an interesting result (Figure 13c). Normal guitar performance does not afford sustained excitation action, although it can be accomplished with a bow on the strings, as Led Zeppelin's guitarist, Jimmy Page, popularized in the late 1960s. However, apart from using extended playing techniques—such as pressing on the strings with the hands or using additional equipment, such as a bow, vibrato arm, or electronic effects processing units—a player can only hit on a string once (impulsive) or as a series of impulses (iterative). Thus, sustained motion is available only for the modification action, such as bending the string with a finger on the left hand.

In the prediction, however, we observed a longer decay as compared to an impulsive, single attack of the right arm. This interesting in-between result suggests a means for augmenting the guitar for creative purposes.

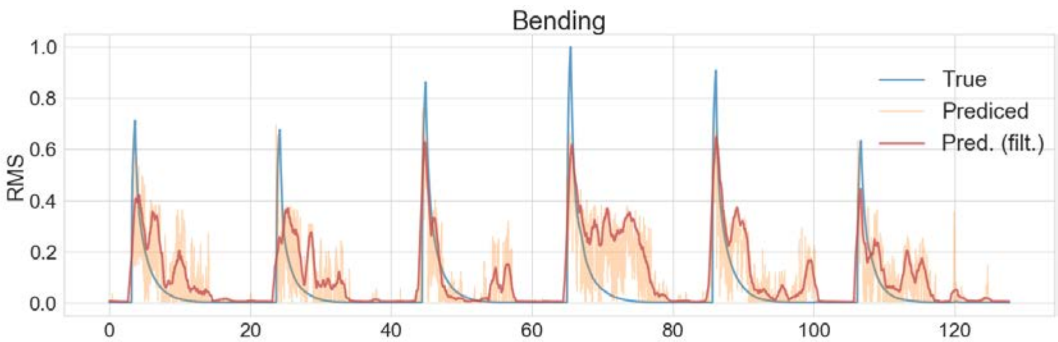
We also tested various model sizes using Euclidean distance measure (EDM), which is a common method for measuring the distance between objects. EDM is calculated as the root of square



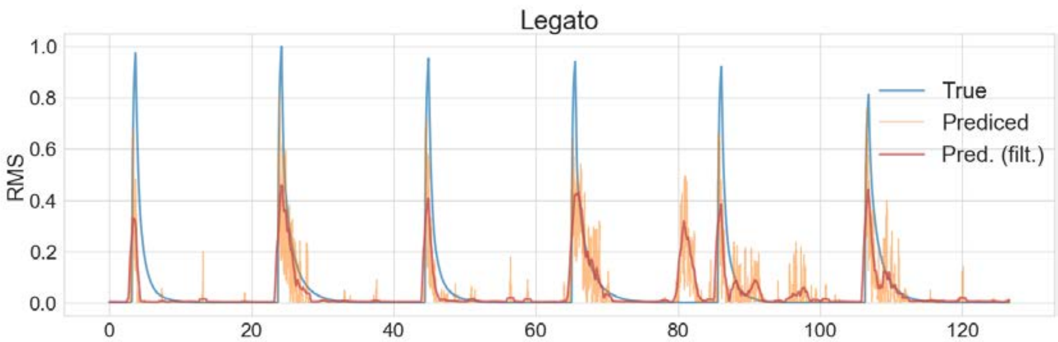
(a) The RMS of the recorded sound and the model prediction for the impulsive task.



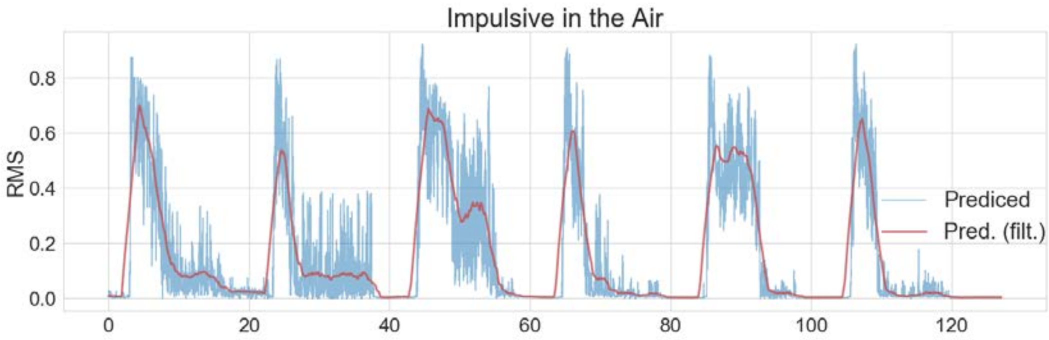
(b) The RMS of the recorded sound and the model prediction for the iterative task.



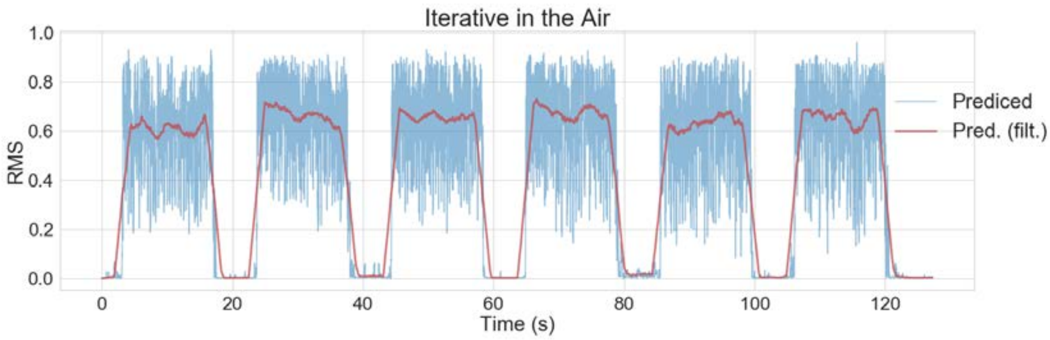
(c) RMS of the recorded sound and the model prediction for the bending task.



(d) RMS of the recorded sound and the model prediction for the legato task.



(e) The predicted sound RMS of impulsive playing in the air.



(f) The predicted sound RMS of iterative playing in the air.

Figure 13. The performance of the model with two hidden layers and 64 units in given tasks. Plots a through d show the true sound RMS and predicted RMS envelopes. Because we recorded impulsive and iterative tasks performed in the air as test data for further exploration, plots e and f show only the predicted sound RMS envelope based on the EMG data of an air performance. The time axis is shared across all plots and predicted curves are processed with a Savitzky-Golay filter (Savitzky & Golay, 1964) to reflect the general shape and facilitate the visual inspection.

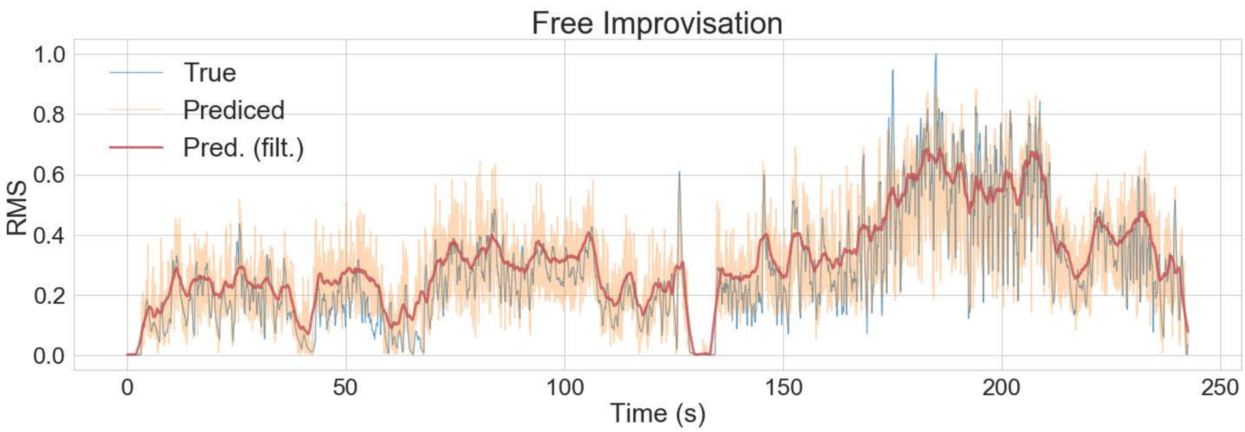


Figure 14. The RMS of the recorded sound and the model prediction of a free improvisation task. Predicted curves are filtered to reflect the general shape and facilitate the visual inspection.

differences between coordinates of two objects (Kang, Cheng, Lai, Shiu, & Kuo, 1996). Given the normalized true and predicted sound RMS vectors $\vec{p}, \vec{s} \in \mathbb{R}^n$, we can find the distances in Euclidean n-space as $\sqrt{(p_1 - s_1)^2 + (p_2 - s_2)^2 \ldots (p_n - s_n)^2}$. The distances between the true RMS and predicted RMS envelopes of the nine models of different configurations were calculated using the free improvisation recordings from 20 participants, of which given tasks were used as training data. This provided us with a statistical measure for evaluating the performance of different model configurations for mapping 16-channel raw EMG data to sound RMS envelope. Figure 15 provides the distribution of distances together with the latency of single-threaded prediction processes on the central processing unit (CPU) of a MacBook Pro 2018. According to results, we observed a trend that the model performance increases along with additional LSTM layers and units; unfortunately, however, the model's performance decreases when the model becomes too large. The prediction time also increases drastically with additional parameters. However, models with a single hidden layer have the least latency even while having a fairly large margin of error. Thus, according to the results, a two-layer stacked LSTM with 32 or 64 units can be seen as a "sweet spot" configuration.

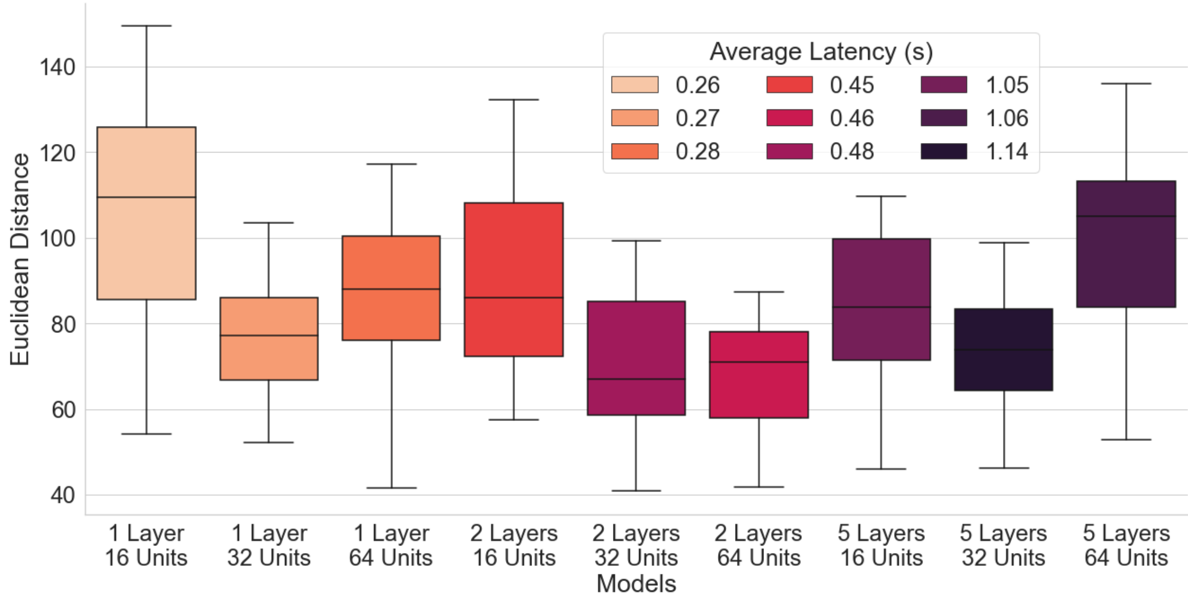


Figure 15. Euclidean distances between true RMS envelope of the free improvisation task and its corresponding prediction of RMS envelope based on nine model configurations. The boxes display the interquartile ranges while the central lines show the median. The whiskers show the minimum and maximum values of the distribution.

Discussion

The implemented model can predict the overall trend of the sound energy of a freely improvised performance based solely on a training dataset of particular action types. As shown in Figure 13, some similarities are evident between the EMG signal and the sawtooth-like patterns of the predicted waveforms. We think this is acceptable, as these fluctuating patterns can be filtered easily and used as an amplitude parameter in the sound synthesis. However, considering that the prediction of a single temporal feature is insufficient for capturing the complexity of musical sound, these patterns might cause problems. These predictions also may lead to unpredictable sound features that could be aesthetically pleasing in an improved model.

Drawing on the results from the tests between different model configurations, we see that, as the model size increases, the distance between the true RMS and predicted RMS generally decreases, but the similarity tends to increase. However, larger model sizes also result in a larger latency, which can cause problems in real-time performance situations. We believe that although a lower similarity can be utilized creatively, higher similarity with a larger latency is much less usable.

Another step in the future development of the system will be to conduct a thorough user study to test the framework. It will be particularly interesting to explore how possible it is to obtain near-optimal latency using the trained model and, moreover, how to use the latency creatively. Also relevant is the exploration of how motion data from an inertial measurement unit can add to the information provided by the EMG data. At its core, the question remains how the spatiotemporality of the performance can be further explored and evaluated.

GENERAL DISCUSSION AND CONCLUSIONS

The main research question that inspired the first experiment of the study regarded the relationships between action and sound in instrumental performance. To answer that, we performed statistical analyses on the data from an experiment in which 31 electric guitarists performed a set of basic sound-producing actions: impulsive, sustained, and iterative. The results showed clear action–sound correspondences, compatible with theories of embodied music cognition. These correspondences' statistical levels varied, depending on the given task. The relatively less-challenging tasks, such as impulsive, yielded higher correlation values. Conversely, we observed how participants' varying level of motor control resulted in unique EMG and audio wave-forms for the iterative tasks, which involved performing a series of impulsive sound-producing actions merged into a single shape. Here, the way participants used rhythms and structured the musical time had a determinant role in the coarticulated muscle activations. Thus, we can argue that complex rhythms yield unique bodily patterns.

An important limitation of Experiment 1 was the gender imbalance. Unfortunately, only one female joined the study. The participants were recruited via local communication channels; thus the range of participants was limited to whoever volunteered. Another limitation was the experimental setup in a controlled laboratory environment, which may have felt unnatural to many participants. The same could be said about the very constrained tasks, which restricted the participants' musical expression. For example, the use of physical effort is most likely quite different than in a live music-making situation. Also, we provided the participants with the instrument, which may have influenced the results. Musicians typically develop bodily habits based on particular instruments—including the string gauge and plectrum. Thus, unfamiliarity with the electric guitar used in this study could have affected the relationships between EMG and audio signals. Furthermore, the analyses clearly showed that these relationships contain nonlinear components, so we could question the reliability of using linear methods. Still, we believe that the use of such methods can provide an example for future work. The results were satisfactory for such an exploratory study, but the choice of statistical methods for correlating bodily signals with sound features remains an open question.

The second research question involved how such relationships between action and sound can be used to create new instrumental paradigms. Relying on the notion of imitating existing interactions in new instruments, we aimed in our second experiment at modeling the action–sound relationships found in playing the guitar. We explored some aspects of this question through a series of analyses in the first experiment. However, we were more focused in Experiment 2, employing our multimodal dataset to train LSTM networks of different configurations. Our results showed that the preliminary models could predict audio energy features of free improvisations on the guitar, relying on an EMG dataset of three distinct motion types. These results satisfied our expectations concerning the size and type of the training dataset. Considering the nonlinear components found in the analysis of the relationships between the EMG and sound RMS envelopes (see Figure 8), the satisfactory outcome of our model corresponded to the known ability of neural networks that, in theory, any continuous function can be approximated by computing the gradient through a neural network. This is achieved by breaking down a complex function into several step-functions computed by the network's hidden neurons. How good the approximation is often depends on the depth or number of layers in the network and the width or number of neurons of each layer (Goodfellow et al., 2016).

A caveat of our research in our second experimental setup is that even the smallest model configuration achieved a much higher latency (see Figure 15 for the results of our analysis on different model configurations) than acceptable ranges (20–30 ms) for real-time audio applications (Lago & Kon, 2004). Although it is possible to reduce the latency using elaborated programming structures, a single predicted feature would still be limited. Moreover, a similar output can be achieved using traditional signal processing methods. Thus, a next step in our research will include expanding the model with spectral, temporal, and spatial features from both motion and audio data. It would also be relevant to explore the potential of what such a deep learning-based framework can afford for musical performance and creativity in a new instrumental concept.

In the future, we will continue to build on this two-fold strategy of combining empirical data collection and machine learning-based modeling. We intend to explore deep learning features for myoelectric control that can be applied to extracting discriminative representations of coarticulated sound-producing actions. We remain interested especially in exploring the creative potential of such models: How can artificial intelligence generally—and deep neural networks particularly—be used to explore the aesthetics of, and embodied interaction with, the transformations of biomechanical waveforms into sound? To answer such a question, we will emphasize exploring the conceptual and practical challenges of space and time, particularly when using the human body as part of the musical instrument. By conducting more user studies, we expect to provide valuable information about conceptual approaches of translating embodied knowledge of actions into the use of new musical instruments.

IMPLICATIONS FOR RESEARCH

The studies presented in this paper are situated within the interdisciplinary research field of music technology (see Serra, 2005). This field involves both practitioners and researchers working with both artistic and scientific methods. Both groups will benefit from the knowledge gained from our empirical studies of basic sound-producing actions and the artificial intelligence methods developed for modeling relationships between muscle energy and audio energy. More broadly, the outcomes of applying multimodal machine learning for creative purposes opens new research activities. These contributions include a new multimodal dataset, the development of custom software tools, statistical analyses between action and sound, and an evaluation of various machine learning configurations. Furthermore, the study provides additional support for previous research on action–sound relationships and embodied music cognition. Our emphasis on EMG irregularities as a control signal suggests an alternative perspective for music technology research on performing arts and human-computer interaction. These irregularities and imperfections open for new creative possibilities.

REFERENCES

Alexandrov, T. (2009). *A method of trend extraction using singular spectrum analysis*. Retrieved from https://arxiv.org/abs/0804.3367

Alexandrov, T., Bianconcini, S., Dagum, E. B., Maass, P., & McElroy, T. S. (2012). A review of some modern approaches to the problem of trend extraction. *Econometric Reviews*, *31*(6), 593–624. https://doi.org/10.1080/07474938.2011.608032

Beranek, L. L., & Mellow, T. J. (2012). Chapter 1: Introduction and terminology. In L. L. Beranek & T. J. Mellow (Eds.), *Acoustics: Sound fields and transducers* (p. 1–19). Cambridge, MA, USA: Academic Press. https://doi.org/10.1016/B978-0-12-391421-7.00001-4

Briot, J.-P., Hadjeres, G., & Pachet, F.-D. (2020). *Deep learning techniques for music generation*. Cham, Switzerland: Springer. https://doi.org/10.1007/978-3-319-70163-9

Burden, A. M., Lewis, S. E., & Willcox, E. (2014). The effect of manipulating root mean square window length and overlap on reliability, inter-individual variability, statistical significance and clinical relevance of electromyograms. *Manual Therapy, 19*(6), 595–601. https://doi.org/10.1016/j.math.2014.06.003

Cadoz, C., & Wanderley, M. M. (2000). Gesture-music. In M. M. Wanderly & M. Battier (Eds.), *Trends in gestural control of music* (Vol. 12, pp. 71–94). Paris, France: IRCAM. Retrieved from https://hal.archives-ouvertes.fr/hal-01105543

Caramiaux, B., Bevilacqua, F., Zamborlin, B., & Schnell, N. (2009). Mimicking sound with gesture as interaction paradigm (Technical Report). Paris, France: IRCAM. Retrieved from http://articles.ircam.fr/textes/Caramiaux10d/index.pdf

Caramiaux, B., & Donnarumma, M. (2020). *Artificial intelligence in music and performance: A subjective art-research inquiry*. Retrieved from https://arxiv.org/abs/2007.15843

Caramiaux, B., Montecchio, N., Tanaka, A., & Bevilacqua, F. (2015). Adaptive gesture recognition with variation estimation for interactive systems. *ACM Transactions on Interactive Intelligent Systems, 4*(4), 18–52. https://doi.org/10.1145/2643204

Caramiaux, B., & Tanaka, A. (2013). Machine learning of musical gestures: Principles and review. In *Proceedings of the International Conference on New Interfaces for Musical Expression* (pp. 513–518). Daejeon, Republic of Korea: Zenodo. https://doi.org/10.5281/zenodo.1178490

De Luca, C. J., Gilmore, L. D., Kuznetsov, M., & Roy, S. H. (2010). Filtering the surface EMG signal: Movement artifact and baseline noise contamination. *Journal of Biomechanics*, *43*(8), 1573–1579. https://doi.org/10.1016/j.jbiomech.2010.01.027

Donnarumma, M. (2015). Ominous: Playfulness and emergence in a performance for biophysical music. *Body, Space & Technology*, *14*, unpaginated. http://doi.org/10.16995/bst.30

Dreyfus, H. L. (2001). Phenomenological description versus rational reconstruction. *Revue Internationale de Philosophie, 216*(2), 181–196. https://doi.org/10.3917/rip.216.0181

Erdem, C., Camci, A., & Forbes, A. (2017). Biostomp: A biocontrol system for embodied performance using mechanomyography. In *Proceedings of the International Conference on New Interfaces for Musical Expression* (pp. 65–70). Copenhagen, Denmark: Zenodo. http://doi.org/10.5281/zenodo.1176175

Erdem, C., & Jensenius, A. R. (2020). RAW: Exploring control structures for muscle-based interaction in collective improvisation. In *Proceedings of the International Conference on New Interfaces for Musical Expression* (pp. 477–482). Birmingham, UK: Birmingham City University.

Erdem, C., Schia, K. H., & Jensenius, A. R. (2019). Vrengt: A shared body–machine instrument for music–dance performance. In *Proceedings of the International Conference on New Interfaces for Musical Expression* (pp. 186–191). Porto Alegre, Brazil: Zenodo. http://doi.org/10.5281/zenodo.3672918

Fiebrink, R. A. (2011). *Real-time human interaction with supervised learning algorithms for music composition and performance* (Doctoral dissertation, Princeton University). Retrieved from https://www.cs.princeton.edu/research/techreps/TR-891-10

Fiebrink, R. A., & Caramiaux, B. (2016). *The machine learning algorithm as creative musical tool*. Retrieved from https://arxiv.org/abs/1611.00379

Fletcher, N. H. (1999). The nonlinear physics of musical instruments. *Reports on Progress in Physics*, *62*(5), 723–764. http://doi.org/10.1088/0034-4885/62/5/202

Françoise, J. (2015). *Motion-sound mapping by demonstration* (Doctoral dissertation, Université Pierre et Marrie Curie). Retrieved from https://www.julesfrancoise.com/documents/JulesFRANCOISE_phdthesis.pdf

Fuentes del Toro, S., Wei, Y., Olmeda, E., Ren, L., Guowu, W., & Díaz, V. (2019). Validation of a low-cost electromyography (EMG) system via a commercial and accurate EMG device: Pilot study. *Sensors*, *19*(23), 1–16. https://doi.org/10.3390/s19235214

Fyans, A. C., & Gurevich, M. (2011). Perceptions of skill in performances with acoustic and electronic instruments. In *Proceedings of the International Conference on New Interfaces for Musical Expression* (pp. 495–498). Oslo, Norway: Zenodo. http://doi.org/10.5281/zenodo.1178019

Godøy, R. I. (2006). Gestural-sonorous objects: Embodied extensions of Schaeffer's conceptual apparatus. *Organised Sound, 11*(2), 149–157. https://doi.org/10.1017/S1355771806001439

Godøy, R. I. (2017). Key-postures, trajectories and sonic shapes. In D. Leech-Wilkinson & H. M. Prior (Eds.), *Music and Shape* (pp. 4–29). New York, NY, USA: Oxford University Press. https://doi.org/10.1093/oso/9780199351411.003.0002

Godøy, R. I. (2018). Sonic object cognition. In R. Bader (Eds.), *Springer handbook of systematic musicology* (pp. 761–777). Berlin, Germany: Springer. https://doi.org/10.1007/978-3-662-55004-5

Godøy, R. I., Haga, E., & Jensenius, A. R. (2005). Playing "air instruments": Mimicry of sound-producing gestures by novices and experts. In S. Gibet, N. Courty, & J. F. Kamp (Eds.), *International gesture workshop: Gesture in human–computer interaction* (pp. 256–267). Berlin, Germany: Springer. http://dx.doi.org/10.1007/11678816_29

Golyandina, N., & Zhigljavsky, A. (2013). *Singular spectrum analysis for time series*. Berlin, Germany: Springer. http://dx.doi.org/10.1007/978-3-642-34913-3

González Sánchez, V. E., Dahl, S., Hatfield, J. L., & Godøy, R. I. (2019). Characterizing movement fluency in musical performance: Toward a generic measure for technology enhanced learning. *Frontiers in Psychology*, *10*, 1–11. https://dx.doi.org/10.3389/fpsyg.2019.00084

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. Cambridge, MA, USA: The MIT Press. Retrieved from https://www.deeplearningbook.org/

Google. (2020). *Teachable machine*. https://teachablemachine.withgoogle.com/

Gritten, A., & King, E. (Eds.). (2011). *New perspectives on music and gesture*. London, UK: Routledge. https://doi.org/10.4324/9781315598048

Hatten, R. S. (2006). A theory of musical gesture and its application to Beethoven and Schubert. In A. Gritten & E. King (Eds.), *Music and gesture* (pp. 1–23). London, UK: Routledge. https://doi.org/10.4324/9781315091006

Hunt, A., & Wanderley, M. M. (2002). Mapping performer parameters to synthesis engines. *Organised Sound, 7*(2), 97–108. https://doi.org/10.1017/S1355771802002030

Ingold, T. (2000). *The perception of the environment*. London, UK: Routledge.¨ https://doi.org/10.4324/9780203466025

Jaramillo-Yánez, A., Benalcázar, M. E., & Mena-Maldonado, E. (2020). Real-time hand gesture recognition using surface electromyography and machine learning: A systematic literature review. *Sensors*, *20*(9), Art. 2467. https://doi.org/10.3390/s20092467

Jensenius, A. R. (2007). *ACTION–sound: Developing methods and tools to study music-related body movement* (Doctoral dissertation, University of Oslo). Retrieved from http://urn.nb.no/URN:NBN:no-18922

Jensenius, A. R. (2017). Sonic microinteraction in "the air." In M. Lesaffre, P.-J. Maes, & M. Leman (Eds.), *The Routledge companion to embodied music interaction* (pp. 431–439). New York, NY, USA: Routledge. https://doi.org/10.4324/9781315621364

Jensenius, A. R. (2018a). Methods for studying music-related body motion. In R. Bader (Ed.), *Springer handbook of systematic musicology* (pp. 805–818). Berlin, Germany: Springer. https://doi.org/10.1007/978-3-662-55004-5

Jensenius, A. R. (2018b). *The musical gestures toolbox for matlab*. In the *Late-Breaking/Demo Session Abstracts for the 2018 International Society for Music Information Retrieval Conference*. Retrieved from http://www.arj.no/wp-content/2018/09/Jensenius-ISMIR2018.pdf

Jensenius, A. R., & Lyons, M. J. (2017). *A nime reader: Fifteen years of new interfaces for musical expression*. Cham, Switzerland: Springer. https://doi.org/10.1007/978-3-319-47214-0

Jensenius, A. R., Wanderley, M. M., Godøy, R. I., & Leman, M. (2010). Musical gestures: Concepts and methods in research. In R. I. Godøy & M. Leman (Eds.), *Musical gestures: Sound, movement, and meaning* (pp. 12–35). New York, NY, USA: Routledge. https://doi.org/10.4324/9780203863411

Kamen, G. (2013). Electromyographic kinesiology. In D. G. E. Robertson, G. E. Caldwell, J. Hamill, G. Kamen, & S. N. Whittlesey (Eds.), *Research methods in biomechanics* (2nd ed., pp. 179–202). North Yorkshire, UK: Human Kinetics, Inc.

Kang, W.-J., Cheng, C.-K., Lai, J.-S., Shiu, J.-R., & Kuo, T.-S. (1996). A comparative analysis of various EMG pattern recognition methods. *Medical Engineering & Physics*, *18*(5), 390–395. https://doi.org/10.1016/1350-4533(95)00065-8

Karjalainen, M., Mäki-Patola, T., Kanerva, A., & Huovilainen, A. (2006). Virtual air guitar. *Journal of the Audio Engineering Society*, *54*, 964–980. Retrieved from http://users.spa.aalto.fi/mak/PUB/AES12860.pdf

Kelkar, T. (2019). *Computational analysis of melodic contour and body movement* (Doctoral dissertation, University of Oslo). Retrieved from http://urn.nb.no/URN:NBN:no-74166

Khan, M. A. R., & Poskitt, D. S. (2011, November). *Window length selection and signal-noise separation and reconstruction in singular spectrum analysis* (Working Paper 23/11). Retrieved from the Monash University Department of Econometrics and Business Statistics website: http://business.monash.edu/econometrics-and-business-statistics/research/publications/ebs/wp23-11.pdf

Kiefer, C. (2014). Musical instrument mapping design with echo state networks. In *Proceedings of the International Conference on New Interfaces for Musical Expression* (pp. 293–298). London, UK: Zenodo. http://doi.org/10.5281/zenodo.1178829

Kingma, D. P., & Ba, J. (2014). *Adam: A method for stochastic optimization*. Retrieved from https://arxiv.org/pdf/1412.6980.pdf

Knapp, R. B., & Lusted, H. S. (1990). A bioelectric controller for computer music applications. *Computer Music Journal*, *14*(1), 42–47. https://doi.org/10.2307/3680115

Kozak M., Nymoen K., & Godøy R. I. (2012). Effects of spectral features of sound on gesture type and timing. In E. Efthimiou, G. Kouroupetroglou, & S. E. Fotinea (Eds.), *Gesture and sign language in human–computer interaction and embodied communication* (pp. 69–80). *Lecture Notes in Computer Science*, Vol, 7206. Berlin, Germany: Springer. o

Lago, N. P., & Kon, F. (2004). The quest for low latency. In *Proceedings of the International Computer Music Conference* (pp. 33–36). Retrieved from http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.10.1143

Lee, M., Freed, A., & Wessel, D. (1991). Real-time neural network processing of gestural and acoustic signals. In *Proceedings of the International Computer Music Conference* (pp. 277–281). Retrieved from http://hdl.handle.net/2027/spo.bbp2372.1991.064

Leman, M. (2008). *Embodied music cognition and mediation technology*. Cambridge, MA, USA: The MIT Press. https://doi.org/10.7551/mitpress/7476.001.0001

Li, X., Zhou, P., & Aruin, A. S. (2007). Teager–kaiser energy operation of surface emg improves muscle activity onset detection. *Annals of Biomedical Engineering*, *35*(9), 1532–1538. https://doi.org/10.1007/s10439-007-9320-z

Maes, P.-J., Leman, M., Lesaffre, M., Demey, M., & Moelants, D. (2010). From expressive gesture to sound. *Journal on Multimodal User Interfaces*, *3*(1-2), 67–78. https://doi.org/10.1007/s12193-009-0027-3

Magnusson, T. (2019). *Sonic writing: Technologies of material, symbolic, and signal inscriptions*. London, UK: Bloomsbury Academic. https://doi.org/10.1080/14794713.2020.1765577

Martin, C. P., Glette, K., Nygaard, T. F., & Torresen, J. (2020). Understanding musical predictions with an embodied interface for musical machine learning. *Frontiers in Artificial Intelligence*, *3*, Art. 6. https://doi.org/10.3389/frai.2020.00006

Martin, C. P., Jensenius, A. R., & Torresen, J. (2018). Composing an ensemble standstill work for myo and bela. In *Proceedings of the International Conference on New Interfaces for Musical Expression* (pp. 196–197). Blacksburg, VA, USA: Zenodo. http://doi.org/10.5281/zenodo.1302543

Martin, C. P., & Torresen, J. (2019). An interactive musical prediction system with mixture density recurrent neural networks. In *Proceedings of the International Conference on New Interfaces for Musical Expression* (pp. 260–265). Porto Alegre, Brazil: Zenodo. http://doi.org/10.5281/zenodo.3672952

Mendoza Garay, J. I., & Thompson, M. (2017). Gestural agency in human-machine musical interaction. In M. Lesaffre, P.-J. Maes, & M. Leman (Eds.), *The Routledge companion to embodied music interaction* (pp. 431–439). New York, NY, USA: Routledge. https://doi.org/10.4324/9781315621364

Nojima, I., Watanabe, T., Saito, K., Tanabe, S., & Kanazawa, H. (2018). Modulation of EMG-EMG coherence in a choice stepping task. *Frontiers in Human Neuroscience*, *12*, Art. 50. https://doi.org/10.3389/fnhum.2018.00050

Nymoen, K., Caramiaux, B., Kozak, M., & Torresen, J. (2011). Analyzing sound tracings: A multimodal approach to music information retrieval. In *Proceedings of the International ACM Workshop on Music Information Retrieval with User-centered and Multimodal Strategies*, (pp. 39–44). New York, NY, USA: ACM. https://doi.org/10.1145/2072529.2072541

Næss, T. R. (2019). *A physical intelligent instrument using recurrent neural networks* (Master’s thesis, University of Oslo). Retrieved from http://urn.nb.no/URN:NBN:no-73901

Paine, G. (2009). Towards unified design guidelines for new interfaces for musical expression. *Organised Sound*, *14*(2), 142–155. https://doi.org/10.1017/S1355771809000259

Pakarinen, J., Puputti, T., & Välimäki, V. (2008). Virtual slide guitar. *Computer Music Journal*, *32*(3), 42–54. https://doi.org/10.1162/comj.2008.32.3.42

Pham, H. (2006). *Pyaudio: Portaudio v19 python bindings*. Retrieved from https://people.csail.mit.edu/hubert/pyaudio

Phinyomark, A., Campbell, E., & Scheme, E. (2019). Surface electromyography (EMG) signal processing, classification, and practical considerations. In G. Naik (Ed.), *Biomedical signal processing* (pp. 3–29). Singapore: Springer. https://doi.org/10.1007/978-981-13-9097-5_1

Pizzolato, S., Tagliapietra, L., Cognolato, M., Reggiani, M., Müller, H., & Atzori, M. (2017). Comparison of six electromyography acquisition setups on hand movement classification tasks. *PloS One, 12*(10), e0186132. https://doi.org/10.1371/journal.pone.0186132

Purwins, H., Li, B., Virtanen, T., Schlüter, J., Chang, S.-Y., & Sainath, T. (2019). Deep learning for audio signal processing. *IEEE Journal of Selected Topics in Signal Processing*, *13*(2), 206–219. https://doi.org/10.1109/JSTSP.2019.2908700

Rowe, R. (1992). *Interactive music systems: Machine listening and composing*. Cambridge, MA, USA: The MIT Press. Retrieved from https://wp.nyu.edu/robert_rowe/text/interactive-music-systems-1993/

Santello, M., Flanders, M., & Soechting, J. F. (2002). Patterns of hand motion during grasping and the influence of sensory guidance. *Journal of Neuroscience*, *22*(4), 1426–1435. https://doi.org/10.1523/JNEUROSCI.22-04-01426.2002

Schacher, J. C., Miyama, C., & Bisig, D. (2015). Gestural electronic music using machine learning as generative device. In *Proceedings of the International Conference on New Interfaces for Musical Expression* (pp. 347–350). Baton Rouge, Louisiana, USA: Zenodo. http://doi.org/10.5281/zenodo.1179172

Schaeffer, P. (2017). *Treatise on musical objects* (C. North & J. Dack, Trans.). Oakland, CA, USA: University of California Press. https://doi.org/10.1525/9780520967465-001 (Original work published in 1967)

Schober, P., Boer, C., & Schwarte, L. A. (2018). Correlation coefficients: Appropriate use and interpretation. *Anesthesia & Analgesia*, *126*(5), 1763–1768. https://doi.org/10.1213/ANE.0000000000002864

Schubert, E., Wolfe, J., & Tarnopolsky, A. (2004). Spectral centroid and timbre in complex, multiple instrumental textures. In *Proceedings of the International Conference on Music Perception and Cognition* (pp. 654–657). Retrieved from http://newt.phys.unsw.edu.au/~jw/reprints/SchWolTarICMPC8.pdf

Selesnick, I. W., & Burrus, C. S. (1998). Generalized digital Butterworth filter design. In *IEEE Transactions on Signal Processing*, *46*(6), 1688–1694. https://doi.org/10.1109/78.678493

Serra, X. (2005). Towards a roadmap for the research in music technology. In *Proceedings of the International Computer Music Conference.* Retrieved from https://repositori.upf.edu/handle/10230/34486?locale-attribute=en

Smalley, D. (1997). Spectromorphology: Explaining sound-shapes. *Organised Sound*, *2*(2), 107–126. https://doi.org/10.1017/S1355771897009059

St-Amant, Y., Rancourt, D., & Clancy, E. A. (1996). Effect of smoothing window length on rms emg amplitude estimates. In *Proceedings of the IEEE Annual Northeast Bioengineering Conference* (pp. 93–94). New Brunswick, NJ, USA: IEEE. https://doi.org/10.1109/NEBC.1996.503233

Tahiroğlu, K., Kastemaa, M., & Koli, O. (2020). Al-terity: Non-rigid musical instrument with artificial intelligence applied to real-time audio synthesis. In *Proceedings of the International Conference on New Interfaces for Musical Expression* (pp. 331–336). Birmingham, UK: Birmingham City University.

Tanaka, A. (1993). Musical technical issues in using interactive instrument technology with application. In *Proceedings of the International Computer Music Conference* (pp. 124–126). Tokyo, Japan: ICMC. Retrieved from http://hdl.handle.net/2027/spo.bbp2372.1993.023

Tanaka, A. (2015a). Intention, effort, and restraint: The EMG in musical performance. *Leonardo Music Journal*, *48*(3), 298–299. https://doi.org/10.1162/LEON_a_01018

Tanaka, A. (2015b). *Myogram* [Music composition and performance]. Retrieved from https://youtu.be/G6H1J2k--5I

Tanaka, A. (2019). Embodied musical interaction. In S. Holland, T. Mudd, K. Wilkie-McKenna, A. McPherson, & M. Wanderley (Eds.), *New directions in music and human−computer interaction* (pp. 135–154). Cham, Switzerland: Springer. https://doi.org/10.1007/978-3-319-92069-6

Tanaka, A., Donato, B. D., Zbyszynski, M., & Roks, G. (2019). Designing gestures for continuous sonic interaction. In *Proceedings of the International Conference on New Interfaces for Musical Expression* (pp. 180–185). Porto Alegre, Brazil: Zenodo. http://doi.org/10.5281/zenodo.3672916

Tanaka, A., & Donnarumma, M. (2018). The body as musical instrument. In Y. Kim & S. L. Gilman (Eds.), *The Oxford handbook of music and the body*. Oxford, UK: Oxford University Press. https://doi.org/10.1093/oxfordhb/9780190636234.013.2

Valles, M. L., Martínez, I. C., Ordás, M. A., & Pissinis, J. F. (2018). Correspondence between the body modality of music students during the listening to a melodic fragment and its subsequent sung interpretation [Abstract]. In *Proceedings of th 15th International Conference on Music Perception and Cognition & the*

10th Triennial Conference of the European Society for the Cognitive Sciences of Music (p. 308). Retrieved from http://sedici.unlp.edu.ar/handle/10915/70462

Van Nort, D., Wanderley, M. M., & Depalle, P. (2014). Mapping control structures for sound synthesis: Functional and topological perspectives. *Computer Music Journal*, *38*(3), 6–22. https://doi.org/10.1162/COMJ_a_00253

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., Carey, C. J. Polat, I., Feng, Y., Moore, E. W., VanderPlas, J., Laxalde, D., … SciPy 1.0 Contributors. (2020). SciPy 1.0: Fundamental algorithms for scientific computing in Python. *Nature Methods*, *17*, 261–272. https://doi.org/10.1038/s41592-019-0686-2

Visi, F., Coorevits, E., Schramm, R., & Miranda, E. R. (2017). Musical instruments, body movement, space, and motion data: Music as an emergent multimodal choreography. *Human Technology*, *13*(1), 58–81. https://doi.org/10.17011/ht/urn.201705272518

Waisvisz, M. (1985). The hands, a set of remote midi-controllers. In *Proceedings of the International Computer Music Conference* (pp. 313–319). Burnaby, BC, Canada: ICMC. Retrieved from http://hdl.handle.net/2027/spo.bbp2372.1985.049

Ward, M. R. (1971). *Electrical engineering science*. New York, NY, USA: McGraw-Hill.

Winges, S. A., Furuya, S., Faber, N. J., & Flanders, M. (2013). Patterns of muscle activity for digital coarticulation. *Journal of Neurophysiology*, *110*(1), 230–242. https://doi.org/10.1152/jn.00973.2012

Authors' Note

The authors thank the participating musicians, as well as Victor Evaristo González Sánchez and Julian Führer, for their contributions during the data collection and modeling processes. This work was supported in part by the Research Council of Norway (Project 262762) and NordForsk (Project 86892).

All correspondence should be addressed to
Çağrı Erdem
RITMO Centre for Interdisciplinary Studies in Rhythm, Time and Motion
Department of Musicology
University of Oslo
Postboks 1133 Blindern 0318 Oslo, Norway
cagri.erdem@imv.uio.no

Human Technology
ISSN 1795-6889
www.humantechnology.jyu.fi

HUMAN TECHNOLOGY

ISSN: 1795-6889

www.humantechnology.jyu.fi Volume 16(3), November 2020, 348–371

CREATING DIGITAL MUSICAL INSTRUMENTS WITH AND FOR CHILDREN: INCLUDING VOCAL SKETCHING AS A METHOD FOR ENGAGING IN CODESIGN

Kjetil Falkenberg[1]
KTH Royal Institute of Technology
Sound and Music Computing
Stockholm, Sweden

Hans Lindetorp
KMH Royal College of Music
Music and Media Production
Stockholm, Sweden
and
KTH Royal Institute of Technology
Sound and Music Computing
Stockholm, Sweden

Adrian Benigno Latupeirissa
KTH Royal Institute of Technology
Sound and Music Computing
Stockholm, Sweden

Emma Frid
KTH Royal Institute of Technology
Sound and Music Computing
Stockholm, Sweden

Abstract: *A class of master of science students and a group of preschool children codesigned new digital musical instruments based on workshop interviews involving vocal sketching, a method for imitating and portraying sounds. The aim of the study was to explore how the students and children would approach vocal sketching as one of several design methods. The children described musical instruments to the students using vocal sketching and other modalities (verbal, drawing, gestures). The resulting instruments built by the students were showcased at the Swedish Museum of Performing Arts in Stockholm. Although all the children tried vocal sketching during preparatory tasks, few employed the method during the workshop. However, the instruments seemed to meet the children's expectations. Consequently, even though the vocal sketching method alone provided few design directives in the given context, we suggest that vocal sketching, under favorable circumstances, can be an engaging component that complements other modalities in codesign involving children.*

Keywords: *vocal sketching, digital musical instruments, codesign, children, performance, prototype building.*

DOI: https:/doi.org/10.17011/ht/urn.202011256768

VERTAISARVIOITU KOLLEGIALT GRANSKAD PEER-REVIEWED www.tsv.fi/tunnus

INTRODUCTION

Applications of sound and music computing increasingly are becoming available to the general public, as demonstrated widely in both academic assets and commercial products. However, these tools largely capitalize on personalized content, from both user and provider perspectives (Mason, Jillings, Ma, Reiss, & Melchior, 2015; Riedmiller, Mehta, Tsingos, & Boon, 2015). For instance, streaming services for music have intricate recommendation systems (Song, Dixon, & Pearce, 2012), and connected playback devices adjust to users' relocation and surroundings (Francombe et al., 2017). Increasingly, end users also can change their sonic environments to their liking (Eriksson, Atienza, & Pareto, 2017). These tendencies toward personalization pose several challenges to design tasks focusing on music technology, one of them being how to allow end users to convey sonic preferences for the purpose of system design. One research method proposed in this context is *vocal sketching*, that is, to sketch sounds using one's vocal apparatus (Ekman & Rinott, 2010). Vocal sketching, along with verbal descriptions of nonvocal sounds, previously have proved effective in, for instance, recognition tasks (Lemaitre, Houix, Voisin, Misdariis, & Susini, 2016; Lemaitre & Rocchesso, 2014). In recent studies (e.g., Panariello, Sköld, Frid, & Bresin, 2019), researchers have approached aspects of communicating sound preferences using a vocal sketching methodology.

The purpose of this work is to explore whether vocal sketching techniques can be applied successfully in codesign situations with very young users who have no formal experience in expressing sounds for design purposes. Specifically, we set out to probe whether preschool children are capable of describing "fantasy musical instruments" with sufficient sonic detail for developing new digital musical instruments (DMIs) or new interfaces for musical expression (NIMEs). The study involves master of science (MSc) students' interactions with children in a codesign workshop environment. In this research design, we could collect valuable information from both the work process and the results, as well as observe how the vocal sketching contributed to the design as one of several possible design components.

The inclusion of children in design processes in technology- and engineering-based learning activities (Harriman, 2015) and elementary school pedagogy (Rosas, Behar, & Ferreira, 2016) is a growing trend. However, little research has been published on children as designers of DMIs or NIMEs. Early work in this field focused on expressive DMIs *for* children as a way of engaging in learning activities (see, e.g., Weinberg, 1999). Researchers have dedicated considerable energy and resources toward the design of instruments for novices. However, participatory design methods have not been so prominent (see, e.g., an overview of recent work in McPherson, Morreale, & Harrison, 2019). In a study by Mazzone, Iivari, Tikkanen, Read, and Beale (2010), three different design activities were carried out for the design of a musical device for children. The authors explored outcomes of the employed methods to understand the activities from two perspectives: whether the methods contributed to the design of a mobile music application and whether they suitably involved children in the process. They presented general considerations about conducting design sessions with children that can also be applied to other design contexts. Their findings suggested important design considerations: (a) involve teachers or education experts before and during the design sessions, (b) use props (and media) to initiate engagement by the children, (c) vary the communication channels using expressive tools (e.g., drawings, prototyping, acting, storytelling, or playing),

(d) record the progress and end results, and (e) allow experts from different disciplines to analyze the results.

In the study described in this paper, we invited 15 students enrolled in an MSc course on musical communication and music technology to codesign novel DMIs together with five preschool children. The students initially interviewed the children in a workshop after which they had 2 weeks to build the children's envisioned instruments. A central task given to the students was to engage the preschool children in vocal sketching to describe the sounds produced by their envisioned instrument. The produced instruments were showcased at the Swedish Museum of Performing Arts, where the children performed the specific instrument they helped design.

The pedagogical perspective of the study presented here is primarily concerning higher education learning—and not that of preschoolers. This perspective was reported in Hansen, Latupeirissa, Frid, and Lindetorp (2020). Below, we describe the perspectives of our study focusing on a codesign workshop employing participatory methods, in which DMI design was conducted with children as informants. The purpose of the described work was not to conduct a controlled experiment for assessing vocal sketching independently or to produce excellent DMIs; the aim communicated to the students was to explore instrument design methods. In essence, we were interested in exploring how the students could engage the children in using vocal sketching in a casual but inviting environment and whether the children could produce design instructions by imitating, imagining, and portraying sounds. To establish the context as casual and inviting, the students were instructed to allow for any method and means initiated by the children and in any modality (gestures, drawing, movement, verbal descriptions, among others).

Vocal Sketching

Vocal sketching involves the use of the voice and body to demonstrate the relationship between action and sonic feedback (Rocchesso, Serafin, & Rinott, 2004). The method of sketching sounds using the voice has grown as a field of audio research. Although researchers employ various definitions of the concept of vocal sketching, the method could be described simply as a counterpart of sketching by drawing on paper (Delle Monache, Rocchesso, Baldan, & Mauro, 2015), where the intended output is an imitation or demonstration of a nonvocal sound (Lemaitre et al., 2016). The term vocal sketching, coined by Bencina, Wilde, and Langley (2008) and presented in work by Ekman and Rinott (2010), was inspired by the notion of "vocal prototyping" for vocalized bodily movements in artistic performances or installations. It can be considered as a methodology for addressing sound design, alleviating the challenges inherent for nonexperts when thinking and communicating about sound in early design stages (Ekman & Rinott, 2010). Ekman and Rinott emphasized the importance of paying attention to social barriers to vocal interaction and ways to overcome obstacles such as shyness in a group setting when using vocal sketching. In their study, they included a warm-up task, a series of design tasks, and a concluding session for reviewing the results and share reflections on the process.

Vocal sketching can be effective particularly when describing sounds that do not have clearly agreed-upon symbols in language, for example, when the source of the sound cannot be identified or when communicating sound characteristics that are ambiguous, such as pitch or temporal qualities (Ekman & Rinott, 2010; Lemaitre, Dessain, Susini, & Aura, 2011). The European research project SkAT-VG (Sketching Audio Technologies using Vocalizations and

Gestures; Rocchesso, Lamaitre, Susini, Ternström, & Boussard, 2015), explored vocal sketching as a method for design and included the results in a sound design toolkit (Baldan, Delle Monache, & Rocchesso, 2017). The SkAT-VG project explored not only vocalization but also complementing gestures (Delle Monache et al., 2018). The project focused partly on sound classification and recognition from sketches and partly on design and synthesis. Vocal sketching as a method has been proved to be feasible, enjoyable, and applicable to design processes even without prior vocal training (Ekman & Rinott, 2010). However, little work in this domain has focused on how it can be used in early design processes with children, which is our focus in the current work.

Children as HCI Designers

Although numerous studies report on designing musical instruments *for* children, very few papers have addressed the topic of designing *with* children. Typically, studies involving musical instrument design target children older than the preschool age or children with special needs, and often also indicate a resolute constraint in technology, such as constructing with, for example, LEGOs. In an overview, Muller (2007) noted that some design approaches historically have been more common when working with children in the design loop. Nesset and Large (2004), for instance, described several design methods adopted for working with children: user-centered, participatory, contextual, and scenario-based. In user-centered design, the developer considers how the product is best adjusted to the user, while in participatory design (or codesign) the intended user is included throughout the design and development processes. In contextual design, on the other hand, developers aim to understand user needs, mainly through observation. Finally, scenario-based design presents a user situation as a fictitious story.

Another approach has been to investigate the four distinct roles a child can take on during design tasks: user, tester, informant, and partner (Druin, 2002). When the child acts as an informant, the design team can choose to include the child at various stages. When the child acts as a partner in design, the role is intentionally more active throughout the process: Ideas typically develop through a collaboration between adults and children. We treated the children primarily as informants, where they contributed with sketches, descriptions, and ideas at the beginning of the design process.

According to Druin (2002), working with children as informants can be done, appropriately, either in their school setting or in environments that support the project goals. Although many previous studies have focused on children aged 7–11 years, we applied similar concepts to younger children.

Research Questions

We did not anticipate that children would be able to envision completely new instruments; rather, we expected them to describe ideas based on previous experience. Thus, we investigated the agreement between the child's description of an instrument and the students' proposed DMI prototype through analysis of annotated recordings from the workshop and the presentation of the instruments. In particular, we looked for traits of how sound preferences were communicated (by the children) and realized (by the students) in order to answer the following questions:

- Can vocal sketching be used by young children to communicate their ideas of new musical instruments via sounds?
- How suitable is vocal sketching when it comes to engaging children as active participants in a design process?

METHODS

The study described in this paper took place at the Swedish Museum of Performing Arts in Stockholm. We followed the five guidelines suggested by Mazzone et al. (2010): (a) involve teachers in the study preparation, (b) engage children using props, (c) allow multimodal and expressive communication, (d) record the progress and end results, and (e) involve experts from other disciplines in the analysis.

Participants

A group of five children from a local preschool (3 F, 2 M, 54–77 months of age, all with good verbal skills), were invited to participate in the study. The children had previously taken part in experiments with sound in interaction at the same museum (Frid, Lindetorp, Hansen, Elblaus, & Bresin, 2019). They also had participated in two lessons at their preschool focusing on musical instruments. During the first session, the children tried about 50 musical instruments (primarily percussive, classical, traditional, and popular instruments, but no synthesizers or other DMIs). As a creative task, they also created a double-reed oboe-like instrument from a drinking straw. During the second session, they were introduced to amplification, synthesizers, and audio effects. The preschool involved in this project had no particular focus on music pedagogy; however, the children participated regularly in singing activities and a pianist visited them biweekly.

Procedure

We carried out the study within the framework of an MSc course on musical communication and music technology; to involve students is an attested approach in our research (Hansen et al., 2019). The task of the students was to develop DMIs for the children, as described in the previous section. Fifteen students prepared for the study through a range of activities. They participated in lectures on vocal sketching, sound synthesis, the role of sound quality in interaction, NIMEs and DMIs, and mapping of an instrument’s control and sound parameters (Hunt & Kirk, 2000). Additionally, they received practical exercises and introductions to music programming in Pure Data (Puckette, 1997), sound design, and interface building using Bela boards (McPherson, 2017). From their study programs, most or all students previously completed courses involving participatory design methods. For this course, they planned a workshop focusing on designing musical instruments with the preschool children.

The children were not explicitly introduced to the concept of DMIs during the project. However, one week prior to the children’s meeting the students, the preschool teacher started discussing traditional musical instruments with the children, and they were encouraged to draw “fantasy instruments” on paper. The day before the workshop, one of the authors demonstrated vocal sketching for the children at the preschool and engaged them in a playful activity of producing sounds vocally. In other words, vocal sketching was not presented as a design method. The exercise involved examples of both portraying the sounds of animals and musical instruments onomatopoeically and of imitating such sounds more realistically.

During the workshop at the museum that the 15 students organized as part of their course work, each child was paired with one representative from a group consisting of between two and four students. The workshop was organized in three steps, as follows.

Step one: The child–student pairs visited a museum exhibition called Sound Check. This exhibition presented 17 novel musical instruments that visitors were invited to try; many of these were DMIs. The presented interfaces ranged from interactive sonification of radiation to a pinball machine with repurposed bass guitars. The children observed and tested these musical instruments for 45 minutes. This activity was important in the workshop for two reasons: This session allowed the children to become familiar with their student-group representative and vice versa, while it also allowed the students and children to establish a common point of reference in terms of musical interfaces. Many of the creations presented at this exhibition demonstrated how various mapping strategies impacted how the instrument is experienced by the player, thus providing a basis for discussion and learning material for the students.

Step two: Following the introductory activity, the student-group representatives presented their child-informants with the task of envisioning and describing a fantasy musical instrument. However, they did not explicitly suggest that vocal sketching or any other method should be used by or with the children when they gathered information.

Step three: Each child–student pair completed an interview session. The students had prepared for their interview sessions based on guidelines for interviewing children (Cohen, Manion, & Morrison, 2002, p. 528). For the purpose of the study, students were encouraged prior to the workshop to conduct practice interviews with children from their social sphere, if possible. The students had been instructed to attempt to lead the discussion about musical instruments if the child would not express design ideas spontaneously, which is common for the partners in a design approach. Although the reference literature (e.g., Cohen et al., 2002) recommended conducting interviews with children in a familiar setting, we decided (in consultation with the preschool pedagogues) to carry out the interviews at the museum. The location was useful for practical reasons but also because we considered it worthwhile to bring the children to a novelty-inspiring environment and build trust through accompanied exhibition visits. All the interviews took place simultaneously in a workshop room set up at the museum. The room was equipped with one video camera and one Zoom HD4 Pro audio recorder per pair for recording and documentation. The entire session was scheduled to take approximately 30 minutes, but the groups were allowed to allocate more time if needed. The teacher and the parents waited by the open door to the room; they could decide to stay within or out of sight.

Design Process

To assist in the group work, we provided access to the children's videos and audio recordings through a nonpublic encrypted streaming service. The students designed the DMI for their respective child based only on the above-described workshop and interview session. In cases where the students needed technical support, for instance, with programming or hardware issues, a teaching assistant was available. The students built their interfaces in a well-equipped prototyping lab located at KTH Royal Institute of Technology. All DMIs were built using Bela boards, sensors (e.g., accelerometers, buttons, sliders, conductive fabric, piezo elements, proximity sensors), and external loudspeakers. Pure Data was the primary programming method. A significant design requirement imposed on the students, apart from following the guidelines defined by the children, was that the instrument should not involve any computer technology beyond the Bela (or a similar microcontroller) and should emit sound through embedded loudspeakers. Preferably, the DMIs should not depend on producing mechanical sounds or be made to resemble acoustic instruments. In addition, design decisions should be motivated by and considered in relation to the course literature and explicated in the form of a written project report that followed a conference proceedings paper template.

Although not specifically explained to the students, all the groups followed a scenario- and contextual-inspired design approach. In this approach, they needed to suggest a target solution quickly, grounded on a basic understanding of the child and his/her goals and presumed actions. Thus, the child acted mainly as an informant in this context and no user-centered techniques or participatory steps were applied in the development process. Instead, the students interpreted the child's description of the final product and added functionality to the design accordingly. Instrument development was completed without intervention from teachers or parents. With only 2 weeks for building the instruments, students had little time for prototype iterations.

Presentation and Showcasing the Instruments

The instruments were showcased at a performance event scheduled 2 weeks after the workshop. At this point, the children revisited the museum to demonstrate their respective instruments together with their student group. The children were provided 45 minutes to try out their instruments together with the students in an informal session (i.e., rehearsal). This period also allowed the children to demonstrate the instruments for each other. After rehearsal, the children and their parents revisited the museum exhibition and had a break. Then, an audience of around 30 persons—consisting mainly of the parents of the children, friends of the students, museum representatives, and colleagues of the authors—attended a public performance showcasing the new instruments. During this performance, each child was allotted several minutes to demonstrate his/her instrument. A microphone and small PA system were used to amplify the output from the instrument's loudspeaker. We video recorded the entire performance for subsequent analysis and documentation.

After the performance, the children and the audience were invited to try all the instruments in an open hands-on session. The researchers took the opportunity to collect informal feedback from the children's parents, the preschool teacher, and the audience members at this point. In total, the children played with their instrument for about 90 minutes, encompassing the rehearsal, presentation, and the concluding open session.

Analysis of Results

Two researchers from KTH, who were familiar with the sketching method but not yet involved in the study until this point, were invited to actively observe the public presentation and then watch, annotate, and analyze the recorded material from both the workshop and musical performances. Specifically, they were instructed to annotate instances of vocal sketching or other articulations of sound preferences, as well as descriptions of gestures or other design input provided by the children. In addition, they attentively observed how the children interacted with and appropriated the instruments. Following the steps of observation and annotating, the researchers compared their annotations from the workshop recordings with the instrument design as performed by the children in the concluding musical performance recordings. The goal of their analysis was to assess how vocal sketches and other information gathered in the workshops seemed to affect the instrument outcome. We include their insights along with our own within the Results section.

Compliance with Ethical Standards

The study did not involve any stressful procedures; thus, no ethics approval was required for the behavioral studies reported in this paper. We conducted the study in accord with the KTH Royal Institute of Technology's ethical policy. For the management of participants' personal data, we followed regulations supervised by KTH's Data Protection Officer. Parents of the preschool children were required to sign a consent form prior to their children's participation in the study. The informed consent included information about the study, the task, and the analysis. All parents and students consented to possible future publication of photo and video material from the experimental session. The participating teacher and all parents received a post-review version of this manuscript and have read and approved the information presented here.

RESULTS

Five new DMIs, as performed by five children, were included in the analysis from the workshop and presentation. Here we report on some general results from the workshop and DMI development before describing each instrument and the embodied experiences of the children performing with them. We also describe the role and impact of vocal sketching for the development of each instrument.

Interview Sessions and Workshop

All interviews were conducted as planned. Only one of the five children asked to have her parent with her; the rest of the children were alone with their student-group representatives during the interviews while the parents waited out of immediate sight. The teacher did not have to intervene to comfort or support the children. All the children were engaged in the conversation, so the noise level in the workshop room was quite high. Nevertheless, recorded sound and video material were of sufficient quality for annotation and analysis of interview dialogues. We note that the workshop session did not generate a large number of vocal sketches. As such, we found no need for a quantitative analysis based on high fidelity audio recordings.

The children used various methods for describing their imagined instruments, including verbal descriptions, playing techniques demonstrations and gestures, sketches and drawings on paper, as well as comparisons with existing instruments. They also used vocal sketching, vocalizations, and singing to a certain extent, and created sounds with the help of their body and nearby objects. Figure 1 shows an example from the workshop where the child described details in a drawing for the envisioned instrument. (The built instrument is presented in Figure 6). We had encouraged the students leading the workshop to allow for and engage in such multimodal activities. Some students

Figure 1. A child describing the pipes of her imagined pan flute-like instrument to the student through drawings during the workshop interview. The final version of the instrument is shown in Figure 6. The image is used with permission.

resorted to vocally sketching themselves in order to encourage the child to behave similarly. Other students gave examples or other suggestions to which the child could respond and react.

The limited number of vocal sketching recordings obtained from the workshop resulted in our decision not to consider quantitative audio analysis as worthwhile. Instead, we transcribed the video and audio recordings to extract the children's descriptions of their envisioned instrument, regardless of descriptive method or presence of vocal sketching. We report on the sound descriptions per respective DMI below.

Instrument Development

During the 2-week instrument development period, we recommended and facilitated that students create the DMIs using the Bela board hardware and the Pure Data software; other resources could be added, if needed. We also strongly encouraged the students to use Bela with batteries and an integrated loudspeaker in order to be able to create small and autonomous instruments. The primary expectation, however, was for each group to manufacture its child's described instrument as accurately as possible.

The instrument interface building took place in the campus prototyping lab that provided materials and technologies such as a laser cutter for wood, a soldering station, a sewing machine, and a 3D printer. Three interfaces were assembled from the ground up, while two were reengineered from existing toys (see, e.g., Figure 2). For the most part, the audio synthesis relied on playback and manipulation of sampled sounds using Pure Data. None of the project groups seemed to run into performance issues with either Pure Data or Bela. The main challenge appeared to be to obtain useful senor input in instrument design. All students were new to the Bela platform, but many had experience with Arduino, which has some commonalities.

Each DMI in its final configuration is presented below together with references to the child's original description, observations from the performance, excerpts from the transcribed material, and, when appropriate, additional insights from the written project reports completed by the students. We present the instruments in the order of youngest child (54 months) to oldest child (77 months).

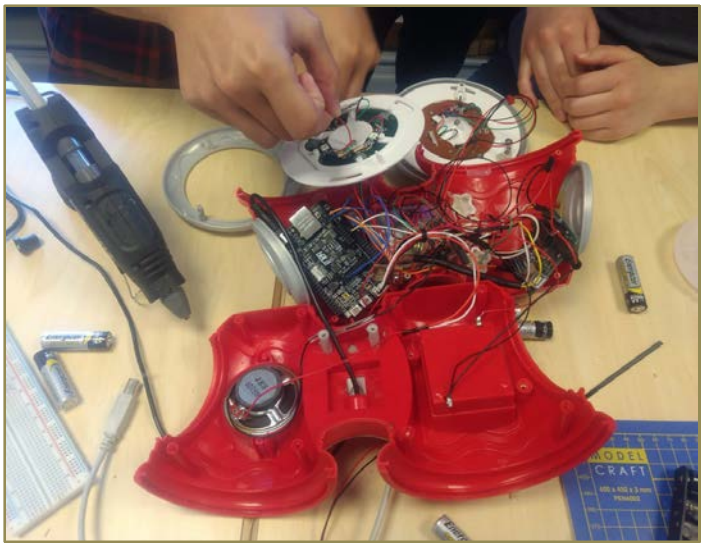


Figure 2. Students repurposing a toy drum for quick prototyping by adding a Bela board and sensors to construct the "Maja" drum (shown in Figure 5).

Instrument 1: "I Pour Soy in the Piano"—A Sushi Table Keyboard Instrument

The child did not use vocal sketching when describing this envisioned instrument; however, he described the sonic outcome of the instrument in detail using a narrative of the interaction with the instrument, which also explains the very explicit name. Interestingly, the proposed design of the envisioned instrument was clearly inspired by real experiences, thus reflecting the transformation of everyday experiences into design ideas, something previous researchers have mentioned as a challenge for children participating in codesign tasks (e.g., Vaajakallio, Lee, & Mattelmäki, 2009). The instrument was to consist of a piano-like keyboard that could be played in a traditional manner using the fingers. The boy described the keyboard layout in a drawing: The keys were supposed to have alternating sounds of eating sushi and, in his own words, of "*drinking soy sauce that is mixed with carbonated water*." The instrument should look like a piano (although the black keys were not included in the drawing), but it should not sound like one. One specific gestural control was demonstrated in the workshop: It involved pouring soy sauce into the piano keyboard. Moreover, the boy specified including a big button that could be triggered through a motion similar to "*a dog digging a hole in the ground*" (as demonstrated with a scratching gesture). A key feature of the instrument was that one should be able to eat while making music.

Much of the interview with this child focused on describing the physical characteristics of the instrument. Consequently, the visual representation corresponded well to the child's description. The final design of the instrument, shown in Figure 3, consisted of a low table intended for eating sushi as well as an integrated MIDI piano keyboard. (It would be venturesome to build a new keyboard without black keys.) As the descriptions of the sonic output were verbal rather than from vocal sketches, the students decided to use sampled sounds of pouring and chewing. Thus, the piano keys trigger various audio recordings of eating sushi and drinking soy.

The child had requested to be able to literally pour soy into the piano as a playing gesture. However, for a number of practical reasons, the students decided to adapt this feature to interaction with a soy bottle placed on the table. The bottle was situated on a pressure sensor that

Figure 3. Playing the keyboard interface of the "I pour soy in the piano" instrument. On the table is also a soy bottle on a pressure sensor and a rectangular cushion with hidden buttons for drumming; the other items on the table were props. The image is used with permission.

would, depending on the amount of soy left, control the playback and sound level of a Japanese-themed music soundtrack. The design instruction of pouring soy into the piano was thus abandoned, to the child's expressed disappointment, but the soy bottle feature was kept. A large padded cushion on the table surface with integrated buttons could be pressed to produce drum-like sound samples and was intended to support the suggested "a dog digging a hole in the ground" gesture.

The instrument was designed with the child's body in mind: The table size was adjusted so that a small person could reach it and play on the instrument from all directions. The design combines familiar experiences from both playing piano and passing a bottle of soy at a dinner table. Interestingly, the child attempted to explore the device's possibilities and limits in terms of embodied musical actions. For example, he sat on the floor to play all the keys simultaneously, using his arms. He also looked underneath the table to see if he could examine the technical construction. Interaction-wise, the keyboard was played using the fingers, but the interaction described as "pouring soy into the piano" and "digging like a dog" were approximated by the soy bottle's weight and hitting the cushion, respectively. In the students' written report, they described how they tackled the challenge of turning the unmusical and mundane eating-sound experience into a musical one. In particular, they explored the cartoonification methods developed in the Sounding Objects project (Rocchesso, Bresin, & Fernström, 2003).

Instrument 2: "Elsa"—A Unicorn Instrument

The child who described this instrument was rather shy and did not engage much during the interview. She used no vocal sketching or paper drawing, instead she asked immediately for a purple unicorn. The further descriptions that resulted from the interview were presented by the student as various options, such as "big or small," "high or low," "one or two buttons," and so on, from which the child chose. The child did not give any instructions regarding the instrument's sounds apart from wanting a "*unicorn sound*," but she described that it should make several different sounds, rather than just a single one. When asked how the instrument should be played, the child made a violin-bowing-like gesture. Back at the preschool later that day, the girl started on her own initiative to experiment in order to find sounds for her instrument, and she asked her teacher to call and tell the students that she wanted the unicorn to sound like *"shaking a pearl in a glass."* This was included in the design directives and was the only interaction between the students and children after the interview. According to the interview and this additional description, the instrument was interpreted to take the shape of a purple unicorn sitting on the ground, be about 30 cm tall, have multiple sounds, and one button.

The final instrument was a repurposed purple unicorn plush toy, shown in Figure 4. An accelerometer, a button, and a loudspeaker were held by the unicorn in a hugging fashion. When tilted, the instrument plays the sound of a glass pearl bouncing rhythmically in a glass container. However, the tilt is measured using the accelerometer and not from a bowing-like gesture. The sounds resemble a glockenspiel and were programmed to allow for playing melodies in three octaves. The button on the back plays a "unicorn sound" that was designed as a sparkling sound of several glass pearls bouncing more energetically in a glass container. According to the students' project report, they decided early on that having several sounds connected to the button would not be intuitive. To open up the design space, they added the accelerometer control for making it possible to produce melodies by tilting around two axes.

When trying out the instrument, the child held it in front of herself with her arms straight and then tilted it back and forth (see Figure 4a). She did not hug the instrument much, contrary to what children typically do with a soft stuffed animal. After receiving instructions from the students, she also tried tilting it sideways, as well as pushing the button (Figure 4b). The instrument design offered an interesting embodied musical experience where the child had to learn how to tilt the unicorn in order to control the pitch.

(a) (b)

Figure 4. Playing the Elsa instrument which consists of a unicorn plush toy equipped with an accelerometer, a loudspeaker, and a big button placed on its back. It plays pitched sounds similar to bouncing glass pearls when tilted. The images are used with permission.

Instrument 3: “Maja”—A Bongo Drum Set

This child used vocal sketching to some extent. Specifically, she produced the sound of a drum by saying “*boom*/*poom,*” and she sang the *Pippi Longstocking* theme song. Moreover, she specified that the drum should “*decide for itself*” what kind of music it should play, as well what it “*is saying*.” In addition, the drum should play the song “Let It Go” from the Disney movie *Frozen*. She indicated that it should be played with drumsticks. The students struggled a bit during the workshop interview because the child had very clear ideas about some of the properties of the instrument, while also responding “*I don’t know*” when asked questions.

The developed instrument corresponded well to the child’s description, apart from the fact that it was not played using drumsticks. Figure 5 shows the final instrument: a repurposed and hacked bongo toy drum with two drumheads emitting red or blue light. The drum “decides for itself” which sound to play on one of the drumheads, using a random function, and plays the theme songs from *Frozen* and *Pippi Longstocking* on the other drumhead. The colored lights (red for *Pippi Longstocking* and blue for *Frozen*) were added to provide a way of learning how to keep a steady beat. The instrument was played using a hitting gesture captured by a piezo element and a pressure sensor. In the written report, the students argued, based on principles for designing for children (Resnick & Silverman, 2005), that it is better to observe and meet children’s needs than to give them what they want. Although a drum that plays random sounds is easy to implement, the design would be in conflict with fundamental instrument mapping strategies of predictability (Jordà, 2004); such an implementation results in the inability to anticipate the sound that will be produced next.

Drum instruments typically involve vivid body movements. However, the child did not approach the instrument as though it was a bongo drum. Instead, she kept a distance from the instrument, trying out one sound at a time with her left hand while pressing gently on the drumhead. The students did not try to move her closer to make it easier for her to use combinations of the sensors and controllers, and therefore, the embodied musical experience did not meet its full potential.

Figure 5. The “Maja” instrument is a repurposed toy drum with a piezo element and a pressure sensor below the drumheads. The drum lights up in blue or red according to which song is being played, and the lights also flash to accentuate the rhythm. The image is used with permission.

Instrument 4: "Jomäo"—A Novel Pan Flute

The student-representative asked this child to mention her favorite instrument and color. She described a red traditional pan flute. When asked to draw a picture, the child partly stepped away from the original idea and added buttons. Through a dialogue with the student, the buttons were mapped to playing tones following her description of *"dark"* to *"light."* In the design process, a button was provided to allow switching between a piano- and a flute-inspired sound. Yet, one explicit directive expressed by the child was that the instrument should *"not sound like a pan flute."* The only example of vocal sketching was when the child quietly said *pling* and *ding* in response to the student asking how she felt a piano sounded. The child seemed to prefer drawing rather than using her voice. One detail that was particularized in this instrument was the enigmatic name, "Jomäo" (for which she offered no explanation). The creative process was driven primarily by the student, while the child would approve or disapprove ideas and suggestions. The instrument description could thus evolve toward a design instruction that conformed to the underlying specifications of the project work.

The developed instrument incorporated the visual sketch and the sound design by resembling a pan flute (see Figure 6). It is played using both hands and can be held like a pan flute with four buttons on each side that can be played using the thumbs. In terms of the produced sounds, the child did not want the instrument to sound like a pan flute. Thus, a player can alternate between a processed flute-inspired sound and a processed piano sound. Additional characteristics—such as octave, scale tones, and timbre—were later introduced by the students. The child-directed concept of "dark" to "light" were interpreted as low to high pitch, respectively (in line with, e.g., Dubus & Bresin, 2013).

Played tones were visualized with red and green lights emitted from the otherwise only decorative pipes. In the students' project report, they discussed how the design instruction of a pan flute conflicted with a wish to produce something novel. The liberty taken here was, for instance,

Figure 6. The instrument "Jomäo" has a pan flute-imitating appearance and consists of a button interface. LED lights in red and green indicate which tones that are being played. The photo shows the instrument during the presentation with a microphone to project the sound. The image is used with permission.

to implement polyphony and consider alternative methods for interaction. One design constraint was that the instrument should be easy for the child to hold. As a result, the size- efficient box could only fit a very small loudspeaker placed in the bottom left under the casing with sound propagating from the rounded kerf-cut corners. The small speaker resulted in a distorted sound which further distanced "Jomäo" from acoustical instrument sounds.

Played tones were visualized with red and green lights emitted from the otherwise only decorative pipes. In the students' project report, they discussed how the design instruction of a pan flute conflicted with a wish to produce something novel. The liberty taken here was, for instance, to implement polyphony and consider alternative methods for interaction. One design constraint was that the instrument should be easy for the child to hold. As a result, the size-efficient box could only fit a very small loudspeaker placed in the bottom left under the casing with sound propagating from the rounded kerf-cut corners. The small speaker resulted in a distorted sound which further distanced "Jomäo" from acoustical instrument sounds.

The size of the instrument was appropriate for a child and, when trying the instrument, she held it with both hands, exploring the sounds using her thumbs. To amplify the sound in the performance situation, a microphone was placed near the instrument. One of the students supported the instrument with her hand and positioned the internal speaker close to the microphone (see Figure 6). These unintentional attributes of the performance situation were observed to reduce the child's possibilities in terms of interaction with the instrument and of the embodied experience.

Instrument 5: "Christiano Ronaldo"—A Button Interface

The child who described this instrument wanted to have soft buttons that could be pushed to generate a range of sounds. Additionally, he successfully vocally sketched multiple times throughout the interview session, producing sounds such as "*click clock,*" "*ding,*" and blowing sounds. This child further tapped the table in order to describe a drum sound. The rows of buttons should play, respectively, melodies, drum sounds, and effect sounds (but no animal sounds). Finally, the child mentioned that the sound should come through headphones only (i.e., no speaker option) and that the volume should be adjustable.

In terms of final design outcome, the instrument corresponded well to the child's description. The instrument is a yellow and green box with three rows of eight buttons and two sliders (see Figure 7). However, the interviewing student had to ask many leading questions in order to discern how the instrument should actually function. The students also made several suggestions during the design phase about how the instrument could work in terms of incorporating both melody and drum sounds, which might have influenced the sonic outcome. The final design contains two sliders that control a sound-altering effect and the overall sound level. The students' report mentioned that they had to interpret which sound effects to include in the third row of the interface and to avoid animal sounds. The students also described how they responded in their design process to the child's vocal sketches with their own vocal sounds, entering a kind of imitation dialogue.

When trying out the instrument, the child initially explored the sounds with one hand. After about a minute, he added the other hand and started combining two sounds. Contrary to the students' expectations, he did not move the sliders and press the buttons simultaneously. Instead, he used one hand for support or to stabilize the instrument while the other hand was moving a slider. When the child enjoyed the sound he was making, he started to swing his body to follow along with the produced sounds.

Figure 7. The instrument “Christiano Ronaldo” consists of a wooden box with three rows of eight soft rubber buttons and two sliders. The image is used with permission.

DISCUSSION

Collaborating creatively with young children involves specific challenges that may differ from codesigning creative interfaces with adults. Interestingly, our work highlights that although vocal sketching was not used to any great extent by all children in the described codesign task and challenges occurred with this proposed method, it is potentially practicable as a tool with others in a user-centered design task focused on developing DMIs for children. We present here a detailed discussion regarding the use of vocal sketching as a prototyping method with young children, along with a summary of limitations of the presented work and suggestions for improvements in future work.

Our findings suggest that young children should be allowed to use a modality of their own choice when describing instrument designs and exploring potential sound scenarios. Also, we claim that a combined methodology, allowing for multiple modes and means of interaction (drawing, describing sounds and shapes verbally, using props, playing, singing, and acting), provides information that is rich enough for successful DMI design. In our research, based on the children’s reactions and responses in terms of satisfaction and joy, we conclude that they were intrigued by having an instrument built specifically for them. We did not evaluate these observations formally through follow-up interviews. However, even a year later, the teachers and parents have noted repeatedly that the children still remember and talk about their involvement in the project in great detail. In particular, the children still refer to “their own” instruments. With the unreserved confirmation from teachers and parents who expressed a high appreciation of the whole experiment, we are certain that the children felt safe and engaged in the design process.

Using Vocal Sketching in DMI Design with Children

Can vocal sketching be used by young children to communicate sounds? We observed that all the children managed to produce vocal sketches in the preparatory tasks conducted before the workshop.

However, during the workshop, only a couple of them used vocal sketching despite being encouraged to do so. The task of describing a "fantasy instrument" is understandably abstract for young children, and the task of describing or mimicking the sound of an instrument they have never heard is even more difficult. Nevertheless, children are commonly known for their vivid imaginations, boundless play, and immersion in fictional realities; not the least, children perform sound imitations already from prelingual age as part of human cognitive development. Thus, they are indeed capable of relating to abstract concepts that involve sound making at an early age. It is worth noting that the children in the study appeared to have no problems when it came to using other means of communication to describe their envisioned instruments. Most participants provided drawings, gestures, and/or verbal descriptions. A combined methodology allowing for multiple means of communicating their design ideas seemed to be fruitful in terms of exploring potential design solutions, perhaps because this allowed the children to use a modality of their own choice.

One may ask why some children produced vocal sketches while others did not. The children had, allegedly, comparable musical interest and training. However, a span of 22 months is significant in terms of learning, skills, experience, and development for children of this age. However, the low number of participants complicates our ability to address age as a contributing factor. Moreover, the results indicate that the younger participants provided descriptions and imaginative instrument concepts that were on par with the older children. Presumably, individual development and experience could account for larger variations than age and training in this case. We also want to acknowledge the impact that the interviewer (an element that was not controlled in this study) had on the creative setting: We could not predict the child's reaction to the interview situation or how the student interviewer managed in a fluid situation. As such, the above-described factors and the study design involving only a small test group can conceivably explain the differences we experienced.

Regarding communication of sound preferences, at least one child expressed a desire for the envisioned instrument to be able to produce low-frequency sounds. In the context of vocal sketching, we note the constraints in terms of frequency range capabilities of children's voices. The physical properties of the vocal apparatus can thus impact vocal sketches. For example, most children cannot produce low-frequency sounds in the same way as adults. Nevertheless, considering that children are capable of mimicking sounds of big trucks, lions, and other low-frequency sounds, this might not have had a significant effect on the feasibility of using vocal sketching as a design method.

For our study, the user-centered approach involved in converting sketching into functional prototypes could not rely solely on an analysis of the vocal sketches found in the recordings: The quality of recordings obtained from the design workshop were insufficient for building sound models from signal analysis alone or from computed audio features. Ideally, the students would have plentiful vocal sketches within a high sound-to-noise ratio, but the spoken interviews and drawing material obtained from the workshop drew more of their attention. In general, the majority of the DMIs focused on sample-based sound generation, not real-time synthesis, based on verbal descriptions of intended sounds. In future studies, it would be interesting to collect more high-quality recordings of the vocally produced sounds. It would be interesting to explore how existing technologies could be used to support the prototyping sessions, for example, by providing examples of similar sounds, by direct manipulation of the voice, or even real-time synthesis. One could imagine solutions in which machine learning methods could be used to identify sounds similar to the vocal sketches or a scenario in which the researchers engage in an iterative loop that supports the child presenting alternative or clarifying examples of proposed audio effects. Additionally, the software tools for prototyping developed in the above-mentioned SkAT-VG project (Baldan, et al., 2017) perhaps could be adapted to support children as designers. A challenge in this context is that most of the design and prototyping procedures ideally would have to be done in real time.

Engagement in the Task

How suitable is vocal sketching when it comes to engaging children as active participants in a design process? The students reported that the material collected during the workshop was somewhat inadequate in terms of providing guidelines for the DMI designs. However, all groups still managed to interpret the given descriptions and declared concise musical ambitions with their creations. Considering that vocal sketching was not introduced explicitly as a design method to the children and that multimodal descriptions in general appeared to come easily, we believe that the children would not identify one modality as less agreeable than another.

As a confirmation of the correspondence between the description and the final product, we observed during the musical performances that the instruments largely seemed to match the children's expectations. Our interpretation is that even if the communication between students and children could be improved to some extent, some fundamental aspects of the design were salient. What is interesting in this context is how important the sound was for the child in the design process. For example, would a child describing a guitar be satisfied if the students had provided a ukulele, and if not, would this be because of the size, the number of strings, or the sound of the instrument? An immediate related interpretation is that the students could possibly have produced the same instruments without using voice sketching at all as a data-gathering method. This speculation is neither supported nor dismissed from our study. However, it is not unlikely that the whole design process (and hence, the students' learning outcomes) would have been very different without this mode of research.

Another perspective that has not been studied directly here is whether the children felt they could convey their ideas using their voice. As this project was, with all probability, the first time the children were asked to imagine and design a new instrument, we have no reason to believe that they would have been in a position to reflect much on the various methods that were used. The goal of their work—that is, what the instrument would sound and look like—was too abstract.

Perhaps the design instructions could have improved if a teacher had been more involved in the workshop preparation: We observed that the children could perform vocal sketching in the preparatory session where one of the authors participated as a teacher. Moreover, elements of play with the creation of sounds, in which the children easily started to imagine and produce different types of sounds with various tools (including their voice) or probes, potentially could have been beneficial in this case (cf. Mazzone et al., 2010).

Motivation and Bias

From the students' perspective, each team was required to present a DMI that was to be both showcased publicly and assessed as a learning activity in the course. In the interest of deriving appreciation, they possibly focused particular effort on making the design immediately appealing to the child who was to perform with it. However, during the demonstration and performance session, we observed that the students' desire to make a good presentation for the course examiner took away some of the focus that they should have had on the child exploring the instrument. Even in the project reports, the students generally tended to focus on presenting the DMI from a perspective of possibilities of sound production and technical possibilities, rather than discussing the general artistic value of playing music with the instrument or the specific use case of the child for whom it was built. Because this learning activity and the project report were part of a formal course examination, we knowingly introduced a bias in which the students' motivation was not necessarily intrinsic to that of codesigning with children. Thus, even though the course results were good, and the children appeared satisfied, we cannot conclude that vocal sketching is an effective method following the procedure outlined here. Therefore, we suggest changes in the methodology for future work (see below).

In the course's project instructions, we presented vocal sketching as a useful method to gather design instructions from preschool children. However, we took into account the risk that young children might not provide sufficient information. Therefore, we designed both the experiment and the learning activity in such a way that the students would not need to rely solely on the children's vocal sketching. Analysis of the recorded interviews revealed that the students often tried to initiate a discussion with the children by encouraging them to speak about their favorite instrument or favorite color. The reason why these questions were brought up was probably that the children typically did not provide spontaneously any design directives for the instruments or their sounds. In addition, expectations from the participants involved in the workshop—students who hoped to get material, and possibly children who hoped to respond correctly—can become a hindrance to creativity rather than a boost when time is constrained. Thus, the tension from unmet expectations likely affected the interview process and consequently the DMI design.

A Safe and Accommodating Environment

Producing imitations of sounds in front of a stranger can be assumed to be a reasonable challenge, for anyone. Even experienced adult researchers can be challenged when trying to make subjects feel comfortable enough to perform vocal sketching. In the case of vocal sketching with adults, Ekman & Rinott (2010) indicated a warm-up exercise was crucial to making the participants more comfortable with the task, as well as possibly repeated examples in which the instructor demonstrated vocally sketched sounds.

Encouraging participants to use vocal sketching for design also is not trivial and requires preparations, sufficient communication skills, and persistence. The need for inspiring conditions in this context is therefore of great importance, particularly so when designing with children. Consequently, we suggest that in encouraging vocal sketching with untrained or even introverted or shy persons, it is imperative that the setting is well prepared, inviting, creative, and safe. Interestingly, none of the children seemed to be bothered by or even aware of the video cameras, perhaps because of the omnipresence of cameras among parents and in schools nowadays.

The most difficult part of this project appeared to have been the communication of sound preferences between the children and students. One remaining question is to what extent the child and the student-group representative were involved in creating a shared understanding of the sound design. One way to address that is through a combination of multiple methods used when working with children as informants. This way, a child's preferences can be taken into account and the task of designing can be adapted for each child.

Limitations and Suggestions for Future Work

It is possible that the methodology used to introduce vocal sketching in the design workshop could have been improved and that more design directives for the envisioned sounds for the respective instrument could have been collected if certain methodological shortcomings had been addressed. Because the children all managed to sketch sounds vocally in the preparatory session held at the preschool, in which one of the authors and all the pedagogues there participated, it is likely that the interview context with students—outside the children's place of familiarity—had a significant, negative impact on the success of the process. Therefore, future design sessions with children should emphasize developing a safe and encouraging environment for the children and involve more careful preparation of warm-up tasks, possibly with singing and elements of play developed together with music pedagogues who know the children. Moreover, other prototyping methods that incorporate props and sound examples should be explored.

The procedure allowed the child to play the new instrument for only a limited time, approximately 90 minutes, which encompassed the rehearsal, presentation, and the open hands-on session. Additionally, the students knew from the task description that little opportunity was available for the children to develop playing skills. Nevertheless, the designs most likely presented a steeper learning curve than expected, resulting in lower potential for musical expression (see Jordà, 2004), a situation that might have been resolved if more time had been scheduled, particularly for the child's practice with a unique instrument.

It is worth noticing that some of the envisioned sounds described by the children in this study were not necessarily tightly linked to the common notion of what a traditional tonal musical instrument should sound like. For example, the child who created the instrument "I pour soy in the piano" described everyday sounds of eating and the child who created the unicorn instrument "Elsa" described the sound of shaking a pearl in a glass container. These results highlight the potential for children to act as designers of sounds that go beyond the notion of a traditional instrument. The fact that the children used metaphors of sounds from everyday interactions shows the potential for exploring how young children want to engage with musical instruments and what actually can be considered a musical instrument for them. This is an interesting finding, considering that commercial musical instruments for young children commonly provide a rather limited set of tonal sounds and provide little possibility for customization or exploration of complex sound synthesis methods.

Much is still unknown about vocal sketching, especially for particular groups such as young children. Some future directions and developments for experiments could be to study the various research modalities individually, to approach vocal sketching systematically, and to find any correspondence between a design description and the end-product by means of qualified interpretation or experience from parents and teachers. Also, it would be valuable to use a range of methods to produce materials for the design tasks and to educate interviewers who set out to use vocal sketching with children. More studies on communicating sound properties nonlingually with embodied sketching methods should be done in order to get a deeper understanding of how to best approach the task of codesigning with children.

CONCLUSIONS

In this paper, we described a study in which students enrolled in an MSc course on musical communication and music technology engaged preschool children in codesign of novel DMIs. The children used a variety of modalities and techniques to provide design descriptions that were then transformed into working musical prototypes. The prototype instruments were constructed with Bela boards and a range of materials and sensors, depending on the design instructions and the children's needs. The purpose of this study was not to evaluate the DMIs. Rather, it was to explore how university students engaged young children in exploring vocal sketching as a design method—that is, as a means for the children to describe envisioned sounds for a new instrument—as well as a means to assess whether this method is effective in instrument design.

In the current study, we considered not only vocal sketches of envisioned instruments but also other sound-producing actions as well as verbal descriptions of sounds. We found that even if vocal sketching might serve as a potential solution when it comes to communicating sound preferences of young children, ideally the method should be combined with other methodologies for exploring sounds. In this study, vocal sketches provided few design directives for the student instrument builders. Nevertheless, it appears the instruments met the expectations of the children, based on their reactions and those of the teachers and parents present during the workshop. Interestingly, the developed DMIs appeared successful even without some children being able to provide a detailed description of the intended sound, perhaps thanks to the richness of other material collected during interviews with the children, such as paper sketches, gestures, and verbal descriptions of sounds.

The study tasked the MSc students with engaging the children in employing vocal sketching in the challenging setting of an unfamiliar environment outside their preschool. Our findings cannot point to any specific rationale for why, despite the children's successful practice of vocal sketching prior to the instrument design workshop, the method was infrequently applied by the children in the workshop setting, even when prompted. Nevertheless, we suggest that it should be possible to use vocal sketching as a method for engaging in a DMI codesign activity, especially when taking into account that sound imitation is a natural and important part of a child's play from an early age (see, e.g., Kuhl & Meltzoff, 1996). Supporting the findings of Mazzone et al. (2010), we emphasize that vocal sketching should be employed together with a range of other methods in design tasks to accommodate the needs and preferences of the children. Thus, it is essential to prepare the interviewers who set out to encourage children to use vocal sketching in a design task.

IMPLICATIONS FOR RESEARCH, APPLICATION, OR POLICY

Our research raises several implications for future studies employing the vocal sketching method. The importance of training participants to more effectively embody and communicate the methods is a key consideration for future research. Additionally, our study suggests that a variety of methods (e.g., drawing and verbal descriptions) should be used to supplement material when engaging children in a sound-design task. Finally, because sketching instrument sounds vocally is not a simple task, neither for children nor adults, the methodology is particularly dependent on favorable circumstances to work well; thus, additional attention to context, environment, and interaction would be key areas to identify and qualify in future research.

ENDNOTE

1. This author has published previously under the name Kjetil Falkenberg Hansen.

REFERENCES

Baldan, S., Delle Monache, S., & Rocchesso, D. (2017). The sound design toolkit. *SoftwareX*, *6*, 255–260. https://doi.org/10.1016/j.softx.2017.06.003

Bencina, R., Wilde, D., & Langley, S. (2008). Gesture ≈ Sound experiments: Process and mappings. In *Proceedings of the 2008 International Conference on New Interfaces for Musical Expression* (pp. 411–414). https://doi.org/10.5281/zenodo.1179490

Cohen, L., Manion, L., & Morrison, K. (2002). *Research methods in education* (5th ed.). London, UK: Routledge. https://doi.org/10.4324/9780203224342

Delle Monache, S., Rocchesso, D., Baldan, S., & Mauro, D. A. (2015). Growing the practice of vocal sketching. In *Proceedings of the International Conference on Auditory Display* (pp. 58–65). Graz, Austria: Institute of Electronic Music and Acoustics (IEM) and University of Music and Performing Arts Graz (KUG). https://iem.kug.ac.at/fileadmin/media/institut-17/proceedings150707_3.pdf

Delle Monache, S., Rocchesso, D., Bevilacqua, F., Lemaitre, G., Baldan, S., & Cera, A. (2018). Embodied sound design. *International Journal of Human–Computer Studies*, *118*, 47–59. https://doi.org/10.1016/j.ijhcs.2018.05.007

Druin, A. (2002). The role of children in the design of new technology. *Behaviour & Information Technology*, *21*(1), 1–25. https://doi.org/10.1080/01449290110108659

Dubus, G., & Bresin, R. (2013). A systematic review of mapping strategies for the sonification of physical quantities. *PLoS ONE*, *8*(12), e82491. https://doi.org/10.1371/journal.pone.0082491

Ekman, I., & Rinott, M. (2010). Using vocal sketching for designing sonic interactions. In *Proceedings of the 8th ACM conference on Designing Interactive Systems* (pp. 123–131). New York, NY, USA, ACM. https://doi.org/10.1145/1858171.1858195

Eriksson, M. L., Atienza, R., & Pareto, L. (2017). The sound bubble: A context-sensitive space in the space. *Organised Sound*, *22*(1), 130–139. https://doi.org/10.1017/s1355771816000418

Francombe, J., Mason, R., Jackson, P. J. B., Brookes, T., Hughes, R., Woodcock, J., Franck, A., Melchior, F., & Pike, C. (2017). Media device orchestration for immersive spatial audio reproduction. In *Proceedings of the 12th International Audio Mostly Conference on Augmented and Participatory Sound and Music Experiences* (AM ’17, pp. 1–5). https://doi.org/10.1145/3123514.3123563

Frid, E., Lindetorp, H., Hansen, K. F., Elblaus, L., & Bresin, R. (2019). Sound Forest: Evaluation of an accessible multisensory music installation. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (pp. 1–12). https://doi.org/10.1145/3290605.3300907

Hansen, K. F., Bresin, R., Holzapfel, A., Pauletto, S., Gulz, T., Lindetorp, H., Misgeld, O., & Sköld, M. (2019). Student involvement in sound and music research: Current practices at KTH and KMH. In *The 1st Nordic Sound & Music Computing Conference* (pp. 36–41). https://doi.org/10.5281/zenodo.3755824

Hansen, K. F., Latupeirissa, A., Frid, E., & Lindetorp, H. (2020). Unproved methods from the frontier in the course curriculum: A bidirectional and mutually beneficial research challenge. In *Proceedings of the 14th International Technology, Education and Development Conference* (INTED2020, pp. 7033–7038). http://dx.doi.org/10.21125/inted.2020.1862

Harriman, J. (2015). Start ’em young: Digital music instruments for education. In *Proceedings of the International Conference on New Interfaces for Musical Expression* (pp. 70–73). https://www.nime.org/proceedings/2015/nime2015_218.pdf

Hunt, A., & Kirk, R. (2000). Mapping strategies for musical performance. In M. M. Wanderley & M. Battier (Eds.), *Trends in gestural control of music* (pp. 231–258). Paris, France: Ircam - Centre Pompidou.

Jordà, S. (2004). Instruments and players: Some thoughts on digital lutherie. *Journal of New Music Research*, *33*(3), 321–341. https://doi.org/10.1080/0929821042000317886

Kuhl, P. K., & Meltzoff, A. N. (1996). Infant vocalizations in response to speech: Vocal imitation and developmental change. *The Journal of the Acoustical Society of America*, *100*(4), 2425–2438. https://doi.org/10.1121/1.417951

Lemaitre, G., Dessein, A., Susini, P., & Aura, K. (2011). Vocal imitations and the identification of sound events. *Ecological Psychology*, *23*(4), 267–307. https://doi.org/10.1080/10407413.2011.617225

Lemaitre, G., Houix, O., Voisin, F., Misdariis, N., & Susini, P. (2016). Vocal imitations of non-vocal sounds. *PLoS ONE*, *11*(12), e0168167. https://doi.org/10.1371/journal.pone.0168167

Lemaitre, G., & Rocchesso, D. (2014). On the effectiveness of vocal imitations and verbal descriptions of sounds. *The Journal of the Acoustical Society of America*, *135*(2), 862–873. https://doi.org/10.1121/1.4861245

Mason, A., Jillings, N., Ma, Z., Reiss, J. D., & Melchior, F. (2015). Adaptive audio reproduction using personalized compression. In the *Audio Engineering Society 57th International Conference: The Future of Audio Entertainment Technology—Cinema, Television and the Internet* (pp. 1–7). http://eecs.qmul.ac.uk/~josh/documents/2015/Mason%20et%20al%20-%20AES57%20-%202015.pdf

Mazzone, E., Iivari, N., Tikkanen, R., Read, J. C., & Beale, R. (2010). Considering context, content, management, and engagement in design activities with children. In *Proceedings of the 9th International Conference on Interaction Design and Children* (IDC ’10, pp. 108–117). https://doi.org/10.1145/1810543.1810556

McPherson, A. (2017). Bela: An embedded platform for low-latency feedback control of sound. *The Journal of the Acoustical Society of America*, *141*(5), 3618–3618. https://doi.org/10.1121/1.4987761

McPherson, A., Morreale, F., & Harrison, J. (2019). Musical instruments for novices: Comparing NIME, HCI and crowdfunding approaches. In S. Holland, T. Mudd, K. Wilkie-McKenna, A. McPherson, & M. M. Wanderley (Eds.), *New directions in music and human-computer interaction* (pp. 179–212). Cham, Switzerland: Springer. https://doi.org/10.1007/978-3-319-92069-6_12

Muller, M. (2007). Participatory design: The third space in HCI. In A. Sears & J. Jacko (Eds.), *The human–computer interaction handbook* (Vol. 20071544, pp. 1061–1081). Boca Raton, FL, USA: CRC Press. https://doi.org/10.1201/9781410615862

Nesset, V., & Large, A. (2004). Children in the information technology design process: A review of theories and their applications. *Library & Information Science Research*, *26*(2), 140–161. https://doi.org/10.1016/j.lisr.2003.12.002

Panariello, C., Sköld, M., Frid, E., & Bresin, R. (2019). From vocal sketching to sound models by means of a sound-based musical transcription system. In *Proceedings of the Sound and Music Computing Conference* (pp. 1–7). https://doi.org/10.5281/zenodo.3249298

Puckette, M. S. (1997). Pure data. In *Proceedings of the International Computer Music Conference* (pp. 269–272). Hong Kong: ICMA.

Resnick, M., & Silverman, B. (2005). Some reflections on designing construction kits for kids. In *Proceedings of the 2005 Conference on Interaction Design and Children* (pp. 117–122). https://doi.org/10.1145/1109540.1109556

Riedmiller, J., Mehta, S., Tsingos, N., & Boon, P. (2015). Immersive and personalized audio: A practical system for enabling interchange, distribution, and delivery of next-generation audio experiences. *SMPTE Motion Imaging Journal*, *124*(5), 1–23. https://doi.org/10.5594/j18578

Rocchesso, D., Bresin, R., & Fernström, M. (2003). Sounding objects. *IEEE Multimedia*, *3*, 42–52. https://doi.org/10.1109/MMUL.2003.1195160

Rocchesso, D., Lemaitre, G., Susini, P., Ternström, S., & Boussard, P. (2015). Sketching sound with voice and gesture. *Interactions, 22*(1), 38–41. http://dx.doi.org/10.1145/2685501

Rocchesso, D., Serafin, S., & Rinott, M. (2004). Pedagogical approaches and methods. In K. Franinovic & S. Serafin (Eds.), *Sonic interaction design* (pp. 135–150). Cambridge, MA, USA: MIT Press.

Rosas, F. W., Behar, P. A., & Ferreira, G. (2016). DMIE: A learning object to design digital musical instruments for education. In *Proceedings of E-Learn: World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education 2016* (pp. 1073–1082). https://www.learntechlib.org/primary/p/174045/

Song, Y., Dixon, S., & Pearce, M. (2012). A survey of music recommendation systems and future perspectives. In *9th International Symposium on Computer Music Modelling and Retrieval* (CMMR 2012, pp. 395–410). London, UK: Queen Mary University of London.

Weinberg, G. (1999). *Expressive digital musical instruments for children* (Doctoral dissertation, Massachusetts Institute of Technology). https://dspace.mit.edu/handle/1721.1/62942

Vaajakallio, K., Lee, J. J., & Mattelmäki, T. (2009). “It has to be a group work!” Co-design with children. In *Proceedings of the 8th International Conference on Interaction Design and Children* (pp. 246–249). https://doi.org/10.1145/1551788.1551843

Authors’ Note

The authors thank the participating children and their parents, teacher Chrissie Engström Elers, and the students of DT2213 Musical Communication and Music Technology at KTH for their contributions to this research. The study was conducted in collaboration with the Swedish Museum of Performing Arts, under the leadership of Lars Annersten. A paper that focused on the pedagogical perspective of this study was presented at the 14th International Technology, Education and Development Conference (see Hansen et al., 2020). The DMIs were built in the MIDDLA, a prototyping lab at the division of Media Technology and Interaction Design, KTH.

All correspondence should be addressed to
Kjetil Falkenberg
KTH EECS
Sound and Music Computing Group
Lindstedsvägen 3, 10044 Stockholm, Sweden
kjetil@kth.se

Human Technology
ISSN 1795-6889
www.humantechnology.jyu.fi

Human Technology

ISSN: 1795-6889

www.humantechnology.jyu.fi

Volume 16(3), November 2020, 372–405

MUSIC, VIBROTACTILE MEDIATION AND BODILY SENSATIONS IN ANOREXIA NERVOSA: "IT'S LIKE I CAN REALLY FEEL MY HEART BEATING"

Gabriela Patiño-Lakatos
Lutheries-Acoustique-Musique team (LAM)
Jean Le Rond d'Alembert Institute
Sorbonne University-CNRS
and
Center for Research in Psychoanalysis,
Medicine and Society (CRPMS)
University of Paris Diderot
France

Hugues Genevois
Lutheries-Acoustique-Musique team (LAM)
Jean Le Rond d'Alembert Institute,
Sorbonne University-CNRS
France

Benoît Navarret
Institute for Research in Musicology
(IReMus)
Sorbonne University-CNRS
France

Irema Barbosa-Magalhaes
Center for Research in Psychoanalysis
Medicine and Society (CRPMS)
University of Paris Diderot
France

Cristina Lindenmeyer
Center for Research in Psychoanalysis,
Medicine and Society (CRPMS) University
of Paris Diderot
France

Maurice Corcos
University of Paris Descartes
and
Research Unit of the Adolescent and Young Adult
Department of Psychiatry
Institut Mutualiste Montsouris
France

Aurélie Letranchant
Research Unit of the Adolescent and Young Adult Department of Psychiatry
Institut Mutualiste Montsouris
France

Abstract: *This article presents the theoretical, scientific, and methodological foundations for the design and implementation of an innovative technological and clinical platform that combined sound, music, and vibrotactile mediation used in a therapeutic setting by adolescents suffering from anorexia nervosa. In 2019, we carried out a pilot experiment with a group of 8 adolescent patients hospitalized in the Eating Disorders Unit of the Department of Adolescent and Young Adult Psychiatry of the Institut Mutualiste Montsouris in Paris. Within this clinical framework, we aimed to create conditions suitable for patients to* *reinvest in their "disaffected" bodily zones and internal experiences through reflecting on the sensations, emotions, and ideas generated by the sensory experiences created when sound and musical stimuli are transmitted through vibrations. The findings demonstrate the ways in which adolescent patients made use of the platform's audiovibrotactile mediating objects to express a personal associative process through speech during their exchanges with clinical psychologists.*

Keywords: *music, vibrations, therapy, anorexia, body, emotions.*

DOI: https:/doi.org/10.17011/ht/urn.202011256769

VERTAISARVIOITU
KOLLEGIALT GRANSKAD
PEER-REVIEWED
www.tsv.fi/tunnus

INTRODUCTION

This article presents the design and creation of an original mediation framework combining sound, music, and vibrotactile stimuli to aid in the treatment of adolescents suffering from anorexia nervosa, part of the "eating disorders" category of the DSM-5.[1] In 2019, we conducted a sound, music, and vibrotactile mediation workshop over five sessions to support adolescents hospitalized for anorexia within the Eating Disorders Service of the Adolescent and Young Adult Department of Psychiatry at the Institut Mutualiste Montsouris (IMM) in Paris.

The Problem of Anorexia Nervosa

Among young persons affected by an eating disorder, particularly those who develop anorexia nervosa during adolescence, the profound physical and mental transformations of puberty can result in dependency issues. At a time when the individual is tackling the process of mourning his/her childhood and starting on the path of adolescence, a largely unconscious struggle arises regarding the internalization process of separation from his/her parents (Jeammet & Corcos, 2010; Kaplan, 1984/1995). In this context, an eating disorder can function as a defense against internal and external strains. The young subject tends to create relationships of dependence with objects in the outside world, particularly in the form of deprivation or excessive feeding. The subject also overinvests in the world of perception and motor action, while simultaneously trying to maintain strict control over his/her own bodily sensations and functions. Control of the mouth and the digestive apparatus can be overemphasized, to the detriment of the other zones of the body, as well as the relationship with others, which then become "disaffected" or disinvested affectively, emotionally, and representationally (Corcos, 2008, 2011; Kestemberg, Kestemberg, & Decobert, 2005; McDougall, 1989). The excessive physical activity and demand for powerful sensations—as well as the desensitized, seemingly abandoned, body—can be understood as forms of defense against the affect that threatens to overwhelm the adolescent. The symptoms of dependence and control emerge at the cost of the subject's relationship to his/her inner psychic life (i.e., sensations, profound affect, emotions, ideas), which also is experienced as threatening. Working with these patients, therefore, means helping them reappropriate their inner experience and emotions so that they gradually become capable of reinvesting in their psychic life and an affective relationship with others (Logak & Barbosa, 2016).

Obstacles to Therapy

Clinicians frequently are challenged with the difficult process of constructing a therapeutic alliance with a patient who regularly attacks relationships and struggles to verbally engage in the process of association (Lindenmeyer, 2019). The process of association indicates the successive connections produced by the subject among his/her emotionally charged psychic representations (Freud, 1895/1950). Given these challenges, what type of therapeutic work can a clinician offer to open a possible pathway? In the field of clinical work with anorexia patients, various mediation alternatives have been explored since the 1950s. Some of these mediation practices have been consolidated and/or diversified in recent decades (Brun, Chouvier, & Roussillon, 2019; Corcos, Gumy, & Loisel, 2019). Among the various techniques explored has been the therapeutic process of using music, dance, massage, balneotherapy, and painting as mediating supports.

In this perspective, we questioned whether an original clinical framework of bodily and artistic practice adapted to these young patients could overcome these therapeutic challenges. Thus, in an exploratory study with eight participants, we implemented an innovative mediation platform to assist in the therapeutic treatment of adolescent anorexic patients by generating profound bodily sensations through sound and music vibrations conveyed via various mediating objects. The design and implementation of this platform resulted from a collaborative and interdisciplinary vision. Our general research question was whether this mediation platform could lead anorexia patients to explore their bodily sensations and verbalize their inner experiences. Explicitly, then, how would these patients come into contact with the mediating objects and express their relationships to these objects? How would they use the sounds, and what types of representation and affect would they express in relation to these sounds? Would these sounds arouse in the patients any representations and emotions linked to their past and current experiences? Moreover, would this mediation platform reveal important contents of patients' psychic lives, which possibly could be elaborated in a full therapeutic process?

The goal of our study was to encourage patients to explore their bodily sensations and perhaps express the affect and ideas generated by these feelings during their discursive exchanges with two psychologists. Given the exploratory nature of the study, our two hypotheses were deliberately broad. Based on previous research work on mediation practices, especially in the therapeutic treatment of anorexia, our main assumption was that the bodily sensations generated by our technological devices in the clinical context would mobilize affect, emotions, and representations (Brun et al., 2019; Corcos et al., 2019; Hypothesis 1). Expressing bodily sensations and associative processes verbally represents an important step for patients that subsequently can be elaborated in a specific therapeutic setting, one distinct from the mediation platform. We expected to find signs of these associative processes in the patients' expressions, both verbal and nonverbal: observed behavior and utterances referring not only to immediate bodily sensations but also to the ideas, personal experiences, and emotions linked to them (Hypothesis 2). In the Methods section, we formulate some additional hypotheses on the potential function of sound excerpts and mediating objects for patients. These additional hypotheses oriented the design of the platform but are outside the scope of this particular paper.

Music, New Technologies, and Vibrotactile Mediation

Thanks to its specific properties and methods, music is a field of experience and practice greatly suited to therapeutic work. Some authors have pointed out that a musical experience is fundamentally embodied and emotional (Cox, 2016; Sacks, 2007). Experiencing music, both its production and reception, requires and sustains the subject's bodily engagement in terms of feelings and various modes of motor activity. Likewise, musical experience awakens and mobilizes affect, from which it typically cannot be dissociated. According to the musicologist Arnie Cox (2016, p. 177), affect in music includes "everything that might be described in terms of feeling, including emotions, moods, desires, and urges, as well as the feelings of exertions, balance, alertness, warmth, and other sensory experiences … Musical affect is a special case of affective life generally: every experience has an affective dimension, which is simply what an experience feels like, or 'the feeling of what happens.'"

Moreover, psychologists, psychiatrists, and psychoanalysts have shown that musical experience is rooted in an infant's first sensory and affective experiences and in his/her relationships with others and the surrounding environment. This musical experience specifically relates to the rhythms, intonations, and melodies transmitted to him/her by others at an early stage and through which the infant learns to move his/her body, giving form to ongoing muscular effort (Didier-Weill, 1995, 1998). On the other hand, the phenomenon of sound and music is dynamic and evanescent: It arises, unfolds, and ceases after a variable period of time, leaving behind mnesic (psychosomatic) traces within the subject and reactivating former traces, some of them related to early experiences. From the perspective of subjective experience, music can be remembered or transformed, especially because of modern recording techniques and digital media. In therapeutic work, therefore, music becomes a highly plastic experiential object (Pankow, 1977).

In its multimodal and embodied aspects, receiving music through vibrations seems particularly relevant when working with anorexic patients because sound waves can be “touched” and, in turn, touch the subject’s body profoundly and extensively. According to Merleau-Ponty, “All tactile perception, while opening itself to an objective ‘property’, includes a bodily component: the tactile localisation of an object, for example, assigns to it its place in relation to the cardinal points of the body image” (1945/2002, p. 370). Further, Merleau-Ponty wrote that “to touch is to touch oneself … The touching oneself and the touching have to be understood as each the reverse of the other” (1964/1968, p. 255). We argue that the phenomenological reversibility of “touching/touching oneself” echoes the psychic reversibility of “touching/being touched,” following the connections between the activity and passivity characteristics of the drive (Freud, 1915/1957).

The vibrotactile modality forms a specific part of the general sense of touch, which is the first sense connecting the human fetus to the mother’s body and, through her, to the environment (Golse, 2010). This first fetal sense usually falls into the perceptive background with the appearance of hearing and sight later in development. However, it remains fundamental to the experience of one’s own body, of others, and the world (Fröhlich, 2000; Merleau-Ponty, 1945/2002). The concept of the “skin ego,” developed by the psychoanalyst Didier Anzieu, aimed to account for the role of skin in the psychic construction of the ego. This development mainly forms from the contact with the mother’s body and the “sound envelope,” a feeling created specifically by the early experience of the mother’s voice among other surrounding sounds (Anzieu, 1985/2016). In fact, the phenomenon of vibration, even though it is accessed through the skin, reaches deep parts of the entire body (i.e., the musculature, skeleton, and internal organs).

New technologies have altered contemporary musical practices significantly, as evidenced by the boom in artistic performances, activities, and computer-music training courses (Genevois & de Vivo, 1999). Alongside the development of new digital tools, growth is apparent over the past decade in the amount of scientific and technical research on the perception of vibrations in the field of musical experience (e.g., Gandhi, Sesek, Tuckett, & Bamberg, 2011; Giordano & Wanderley, 2015; Hopkins, Maté-Cid, Fulford, Seiffert, & Ginsborg, 2016; Merchel & Altinsoy, 2013; Wollman, Fritz, & Frelat, 2015). The results have concerned primarily persons with hearing or visual impairment and have been used in different ways, for example, to aid in everyday activities, leisure or education, art, and therapy. In Europe, particularly in Norway, Finland, and England, vibroacoustic methods have been used in music therapy since the 1980s to improve somatic and mental disorders (Jacobsen, Pedersen, & Bonde 2019; Skille, 1989). A qualitative research based on “vibroacoustic therapy” was carried out in Estonia with adolescent girls suffering from heightened anxiety combined with low self-esteem and/or body image problems (Rüütel, Ratnik, Tamm, & Zilensk, 2004). This experiment intended to enhance the young people's ability to cope with stress through relaxation and positive bodily experiences. In general, new technologies and their present-day practices have transformed indeed the processes of the subjective construction of bodily experience. Thanks to their potential in terms of the synthesis, editing, recording, broadcasting, and real-time transformation of sound, new digital tools offer unprecedented possibilities of exploring the emotions and representations linked to the bodily sensations provoked by sound and musical events (stimuli). In this way, the research carried out in this area of clinical work can be expanded for and adapted to other forms of mental distress.

In this context, the bodily mediation framework we designed had a number of specificities. First, the vibration was not an independent stimulus associated with the sound and music event; rather, it was directly generated by it. Second, patients with anorexia could actively play with and transform music in real-time through the digital tools. Third, they could explore sound and music vibrations through their body contact with "mediating objects" made of various materials. In this way, the framework transformed the reception of sounds and music through vibrating mediating objects that the patient could invest in (Brun, 2019). Additionally, the presence of two psychologists was key in creating a situation that functioned as a mediator between the patient's internal and external worlds. One psychologist played a restrained active role with patients, while the other took on an observation-based role with limited possibilities of interaction. We will specify in the Methods section the experiment setup, the roles of psychologists, and the instructions given to the patients. From a psychoanalytic perspective, our framework aims to encourage the patient to engage in a process of association and verbal expression of his/her deep inner experiences, based on his/her bodily sensations.

We ground our current research in three previous experiments. First, Hugues Genevois and Errika Manta carried out in 2012 a vibrotactile exploration experiment with Θέατρο Κωφών Ελλάδος [Greek Deaf Theatre], a sign-language theatre company based in Athens, in coordination with the group's interpreter and facilitator, Sophia Roboli. This experiment demonstrated the importance of one's relationship toward the social other in the construction of the perception of vibrotactile stimulation.

Second, Hugues Genevois developed a "sound table" and other Max-based digital tools[2] for the *Histoires sensibles* project. This vibrasonorous pedagogic experiment, carried out in 2012–2013, was designed by the French composer Pascale Criton and Elsa Falcucci, a lecturer at the National Institute for Deaf Youth. The experiment was intended primarily for deaf students but also included mixed audiences with different sensory conditions (hearing, deaf, sighted, or blind). It showed that the reception of vibrotactile signals, although seemingly intuitive and immediate, requires a certain level of learning by the subject to consciously elaborate his/her bodily sensations through technological devices (Criton, 2014; Criton, Genevois, Falcucci, & Patiño-Lakatos, 2014; Patiño-Lakatos, 2015).

Third, in 2015, Gabriela Patiño-Lakatos, Benoît Navarret and Hugues Genevois carried out an empirical study looking at the perception of vibration signals transmitted via a prototype vibrotactile bracelet designed by Genevois.[3] This prototype used a transducer (i.e., Dayton Audio 13 mm 8 ohms NXT) controlled through an amplifier by software that synthetized and transformed signals in real time (Max). The experiment involved 10 subjects presenting various sensory conditions: 6 nonimpaired men, a nonimpaired woman, a blind man, a visually impaired man, and a hearing-impaired woman. Throughout four stages, the participants expressed their perception of the vibrotactile signals from the bracelet worn on the wrist of their dominant arm. This experiment resulted in a better understanding of the meaning and possible communication functions of vibrotactile signals in the context of collective play for shared musical practice (Patiño-Lakatos, Navarret, & Genevois, 2019, 2020). Moreover, it showed that participants possessed significant potential to detect consciously, differentiate, and recognize vibration signals. They could organize their perceptions into categories according to their sensory situation, as well as their personal and professional backgrounds. Indeed, as the sense of touch is generally little educated in an explicit and conscious manner, the verbal description of vibrotactile perception was an unusual and astonishing situation for the subjects. Therefore, we concluded, a vibrotactile event mobilizes the recall of personal experiences and metaphorical language.

Taken together, these three previous studies demonstrate the relevance of vibrotactile mediation for the subjective construction of bodily sensations. Yet we believe additional benefits remain unknown.

METHODS

Sample

During a 2-month period, we conducted an experiment with eight patients—one boy and seven girls—aged 14–19 years and hospitalized with the symptoms of either the restricting or restricting–purging types of anorexia nervosa. All had a body mass index of 14–18 (kg.m-2). We used the diagnostic criteria of anorexia nervosa established by the DSM-5. The patients were hospitalized in the Adolescent and Young Adult Department of Psychiatry of the IMM for between 3 and 10 months. After reading and signing the information and consent form, patients were free to either accept or refuse taking part in the workshop without any impact on their usual medical treatment at the institute. This project was approved by the Ethics Committee of the University of Paris: CER-PD: 2019-39-BARBOSA.

Workshop Setting

We implemented the mediation platform in collaboration with the clinical team of the institution responsible for patient care. We first presented to the psychiatrists, who spoke to their patients about the possibility of participating in this experiment. The platform was designed as a workshop, similar to the usual mediation activities offered at the IMM. The workshop involved a short research experiment in a therapeutic context in order to test the mediation platform as a support for a full therapy process. Two clinical psychologists among us conducted the workshop—they are coauthors of this paper. The psychologists were not on the staff of the IMM nor worked with the patients in the institution.

The workshop took place in a room at IMM dedicated to the installation. The psychologists conducted five individual workshop sessions with each patient, at a frequency of one session per week. Each session included a 30-minute exploration time followed by a 20-minute discussion with both psychologists attending. One of the psychologists, always the same person, contacted all the patients in the institution and brought them into the workshop room for each session. The other psychologist was always present in the room when the patient arrived. The patient and the psychologists would take off their shoes before starting the workshop session; this instruction was meant to create a welcoming setting that would mark the transition between outside and inside, between before and after, with respect to the time and space of the workshop. In the first session, one of the psychologists introduced the platform to the patient, but she described neither the sounds nor the possible uses of the mediating objects (see below the description of sounds and mediating objects). At the beginning of each session, one psychologist—again, always the same person—gave the patient the following instruction: “You can do anything you like as long as you take care of the objects.” The patient would then be able to freely explore the objects and express his/her associations in a verbal exchange with the psychologists. In addition to the patients’ free speech, three open projective-type questions were asked systematically at the end of each session: What did you feel? Which of the sounds and objects were the most interesting to you? What did they make you think of? The patients’ comments throughout the workshop sessions and their final responses were written down by the observer psychologist (always the same one) who is trained for this task.

The presence of two psychologists in the workshop sessions followed the clinical requirements of this study, based on the importance of maintaining a frame of human contact rather than leaving the patient alone with the objects. From a clinical point of view, it was essential that this type of platform is led by professionals trained to recognize and explore unconscious processes with patients. The interventions of the psychologists were restricted to giving instructions regarding the workshop, providing a reassuring, containing presence—akin to the function of holding (Winnicott, 1971a)—and stimulating the vibrasonorous play and the patient’s verbal associations. Their presence fulfilled a containing function insofar as they supported the patients by intervening only in what was necessary to facilitate the experience. The goal was not to engage in a therapeutic process as such but encourage (or, occasionally, refocus) the patient in the options within the workshop space. Moreover, the functions of the psychologists were differentiated with respect to the patients. One psychologist was responsible for a more active discursive exchange with the patient; the other served a more distant function as an observer, interacting with the patient only periodically by asking questions about his/her sensory experience with stimuli and objects, as well as responding to specific requests or making suggestions depending on the patient’s situation. Second, two psychologists were necessary to the research protocol: The use of patient recordings in a clinical setting with vulnerable adolescents is a sensitive issue and we were not able to use recording devices. Thus, to obtain detail from the events within each workshop, one psychologist was responsible specifically for observing and taking accurate notes on what happened and what was said during each session. She noted the patient’s speech and use and duration of the contact with the objects (timing and verbatim). This technique introduced a limitation in the data collection. Therefore, in addition to the notes taken during the session by the observing psychologist, both psychologists would transcribe and compare their observations at the end of the session. This two-step procedure was implemented to control possible biases (e.g., omission and interpretation) in data collection by the researchers.

Mediating Objects: Sound, Music, Contact, and Vibration Media

The sound, music, and vibrotactile mediation framework we designed comprised 16 audiovibrotactile stimuli as well as several devices—four audiovibrotactile mediating objects, two microphones, a gesture control MIDI interface, a computer, a software tool, a mixing console, and an amplifier. We describe the features and functions of both the stimuli and the devices in the following subsections.

Stimuli

In previous studies, we observed the semiotic and sensory-feedback properties of a vibrotactile stimulus as part of our research on the use of digital interfaces in musical practice, auditive supplementation, and the sharing of a new type of "sensory listening." In a 2015 study (Patiño-Lakatos et al., 2019, 2020), the stimuli were speech, complex composed sounds (i.e., musical excerpts), or basic sounds (pulse-like patterns, drone effects, and homogeneous tone colors) used to evaluate sensory thresholds (frequency and dynamic range sensibilities). In the current study, the patients could experience the sound and music both auditorily and vibrotactilely. They had a tactile experience of sound and music stimuli through body contact with the vibrating objects, as the technical system converted sound signals to vibrations. Simultaneously, the actual audio component was available in the stimuli, as the various vibrating objects functioned like loudspeakers, producing an aerial acoustic radiation of the sound. Thus, the patients could clearly hear the sound of the stimuli in an auditive way, without necessarily entering into physical contact with the vibrating objects.

Unlike previous studies, the present research was carried out in a clinical setting, that is, not in a purely artistic context. The stimuli were chosen with regard to the characteristics of patients hospitalized for eating disorders and based on information provided by Professor Dr. Maurice Corcos and his team at IMM. In this sense, we formulated a series of additional hypotheses on the possible psychological functions of different types of stimuli. Some sound stimuli were familiar (e.g., soundscapes, soundscenes from everyday life, a heartbeat) in order to induce a sense of relief or control (Hypothesis 3). Previous work on the psychosomatic approach suggest that familiar sounds have a containing function in that they underpin a sense of continuity and sharing in the lived experience (Baruch, 2009; Lauras-Petit, 2009). Other stimuli were abstract sounds, such as unidentifiable musical instruments and less common musical languages (e.g., a composition process based on writing *sounds* rather than writing *notes*). Our assumption was that abstraction could generate emotional states through the effect of surprise, leading to a renewed interest and encouraging patients to express their unique personal imagery and deep inner experiences (Hypothesis 4). The final, more involved, stimulus comprised vocal sounds captured during the sessions by two handheld microphones, one for the patient, another one for the interacting clinician. Patients could speak and explore their own voice while fully in control of the sound production process (e.g., intensity of sound, duration, meaning, sensations). Thus, their voices were emitted from inside their bodies and then returned to them from the outside, auditorily and vibrotactilely, through the vibrotactile objects: This can be described as an audiovibrotactile feedback (see below). As for the interacting clinician (always the same one), she could invite patients to speak with her or play with her voice as part of their experience. Patients also could feel the vibrations of the clinician's voice through the audiovibrotactile feedback. Following the work of Didier Anzieu on the sound envelope, we assumed that the audiovibrotactile sensation of the voice could trigger representations and affect related to the construction of patients' body and psyche (Hypothesis 5). Furthermore, the range of therapies provided at the IMM includes massage therapy and "packing" therapy[4] (Corcos et al., 2019) as ways of satisfying the need for an envelope felt by certain patients. We took these patients' needs into consideration in our selection of the stimuli, assuming that some of them could satisfy this need (Hypothesis 6), and in the design of the vibrotactile objects (see below).

The technical configuration of the entire clinical framework presented two additional unique features that affected the patients' sensory experience, both while exploring and manipulating the stimuli and in reflecting on their experiences. First, complex sensory experience can arise from both vibrotactile and auditive perceptions because some of the vibrating objects produce enough acoustic radiation to result in audible sounds—especially the wooden table and the foam ball. In this sense, we could describe these objects as audiovibrotactile.[5] The range of frequencies they are able to reproduce makes it possible to both hear vibrations and feel them through the body. Secondly, various designs of the vibrating objects can induce complementary perceptive modalities, meaning that, depending on the transmitting object, a single stimulus can potentially produce different sensations.

We selected and edited 16 sound and music sequences (each music sequence is an excerpt of a larger work), looking at their potential to elicit physical and psychic sensations that encourage free-association and subjective interpretation (Table 1). Eight sound sequences evoked natural or industrial environments, whereas the remaining eight were pieces of music played in a particular context. These comprised rhythmic sequences of impulse-like signals whose dynamic envelope could be identified based on the instruments or lesser known acousmatic music compositions. This variety of stimuli was intended to allow patients to choose from and interact with a variety of sounds. In addition, the corpus of stimuli remained strictly the same from session to session with each patient. We therefore could observe changes in listening and practices over time.

Audiovibrotactile Mediating Objects

We developed four audiovibrotactile mediating devices: a low, oval-shaped wooden table (length: 47 in. [119.4 cm]/width: 23.5 in. [59.7 cm]/height: 14 in. [35.6 cm]), a medium-sized foam ball (diameter: 8 in. [20.3 cm]), a headrest pillow (width: 12 in. [30.5 cm]) and a blanket (originally measuring 59 in. x 90.5in. [150 x 230 cm] but folded in half and sewn to a size of 29.5 in. x 90.5 in. [75 x 230 cm]). These devices served as media to transmit to the adolescents the vibrotactile stimuli induced by sound and music. Each item offered various kinds of contact with the vibratory phenomenon and enabled us to study the different relationships with the objects in which patients can engage. We formulated, as follows, a series of additional hypotheses on the possible psychological functions of these devices.

First, the table is a hard, semifixed mediating object used previously in experiments with children and adolescents. It is symbolically associated not just with eating—and thus with food as an object—but also with social sharing and exchange. As an object evocative of community and communion, the table fosters the mediated relationship to the other and the communication of bodily sensations around a shared object. In our framework, we deliberately opted for a low, oval-shaped table, which is more ambiguous and less determined by cultural conventions than the rectangular shape that frequently functions as a dining table. This object can also fulfill the supporting function of a table-bench, on which the patient can, if he/she chooses, rest or sit down, lie down, or stand (Hypothesis 7). Second, the medium-sized headrest pillow is a soft mediating object, the style of which usually allows the patient to wear it around his/her neck and encourages relaxation (Hypothesis 8). However, because of its shape and texture, it can be placed on or held

Table 1. The 16 Edited Sound and Music Stimuli Used in the Mediation Platform with Hospitalized Adolescent Anorexia Nervosa Patients.

	SOUND STIMULI
1	Water flow (BBC, n.d.d)
2	Water waves (BBC, 1981)
3	Wind and sandstorm (BBC, n.d.a)
4	Footsteps on grass (BBC, n.d.b)
5	Human heartbeat (BBC, n.d.c)
6	Steam locomotive (excerpt, Schaeffer, 1948/2010a, CD 1, track 1)
7	Water drops (Genevois, 2019)[6]
8	Bubbling water (Genevois, 2019)
	MUSIC STIMULI
9	Traditional music of Bali (Sentana, 1991, track 4)
10	Traditional music of Benin (Rouget, 1990)
11	"Bilude" (Schaeffer, 1979/2010c, CD 3, track 17)
12	"De Natura Sonorum" (Parmegiani, 1974/2008, CD 5, track 9)
13	"Étude aux Tourniquets" (Schaeffer, 1948/2010b, CD 1, track 2)
14	"Pacific Tubular Waves" (Redolfi, 1979/1988, track 6)
15	"Presque Rien N°1" (Ferrari, 1970/2009a, CD 3, track 1)
16	"Presque Rien N°4" (Ferrari, 1990–1998/2009b, CD 3, track 9)

Note. The parenthetical information indicates the source or creator of each stimulus.

against the body in various other ways. Third, the cotton blanket is a flexible and enveloping object, fulfilling a containing and intimate psychic function (Anzieu, 1985/2016; Bion, 1962; Winnicott, 1971a; Hypothesis 9). Lastly, the ball is a round and mobile mediating object, soft but firmer than the headrest pillow. As such, the ball supports the functions of control, of grasping, manipulating, and moving with the hands in relation to other parts of the body, such as the chest, face, stomach, or legs (Hypothesis 10).

In order for these objects to function as an audiovibrotactile mediating device, they needed to be fitted with specific technology. We therefore employed transducers to transmit the sounds through the mediating objects. These transducers function rather like loudspeakers: They receive an audio signal through a Behringer NX1000D amplifier. For technical reasons that involve, among other things, the density of the vibrating objects, we used two types of transducers in the installation. We affixed a Clark Synthesis TST329, a bigger and more powerful unit, under the table. Then, several Dayton Audio DAEX25FHEs were embedded in the other objects: four in the blanket, three in the headrest pillow and one in the ball.

Each object—including the two hand microphones—was placed inside a space delimited by a carpet (Figure 1), which the patient would enter. We considered this material, albeit soft, a demarcation of the object exploration space that was clinically relevant to symbolically institute the frame and the limits of a "playing" setting for the patients (Winnicott, 1971a). Establishing such parameters also helped to spatially differentiate the positions of the patients and psychologists during the workshop sessions. The two psychologists remained half-way across the room, outside the frame delimited by the carpet, although they could sometimes enter the space to modify something, provide support, or help the patient continue.

The elements of music and sound transmitted by these mediating objects encouraged exploration without necessarily calling upon pre-established interpretations. Instead, they were intended to stimulate the patient's associative process while taking bodily sensations as a starting point. By engaging various parts of the body—such as the upper and lower limbs, back, neck, torso and face—we expected this framework to elicit a reinvestment of the body regions that previously have been disinvested defensively by these young anorexic patients, that is, areas other than the mouth and the digestive tract (Figure 2). In this way, the technological tools used in the body mediation framework could bring to light free associations between (a) sound and music, (b) bodily sensations originating through touch, and (c) the events of the patient's life history.

Figure 1. The mediating objects placed inside the framework space, delimited by a carpet. Other materials for the patients' use were included.

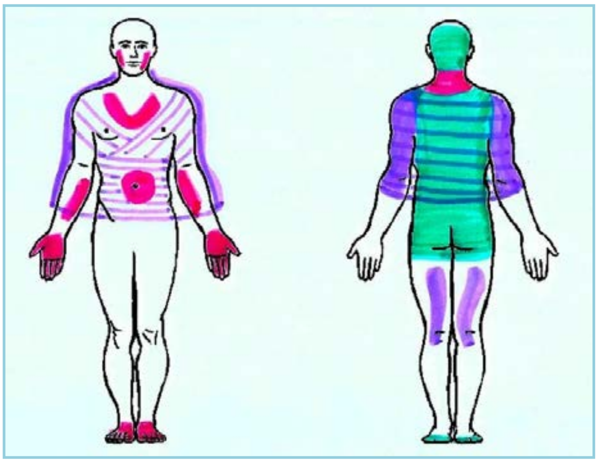


Figure 2. Body zones affected by contact with the framework's mediating objects. Colors represent the body contact of the mediating objects except the microphone: blanket (purple), table (green), ball, and/or headrest pillow (fuchsia).

Technical Installation and the Software Tool

The installation was controlled by a software tool developed at LAM (Lutheries-Acoustique-Musique [Lutheries-Acoustics-Music]) and used the Max programming language. This software enabled the transmission of a selection of sound sequences and musical excerpts to the four audiovibrotactile objects. To do this, we provided the patients with a Korg NanoKontrol2 MIDI[7] interface that allowed them to control the sensory aspects of the vibrations and sounds, such as intensity level and playback speed (Figure 3). This MIDI interface sent commands from the patient

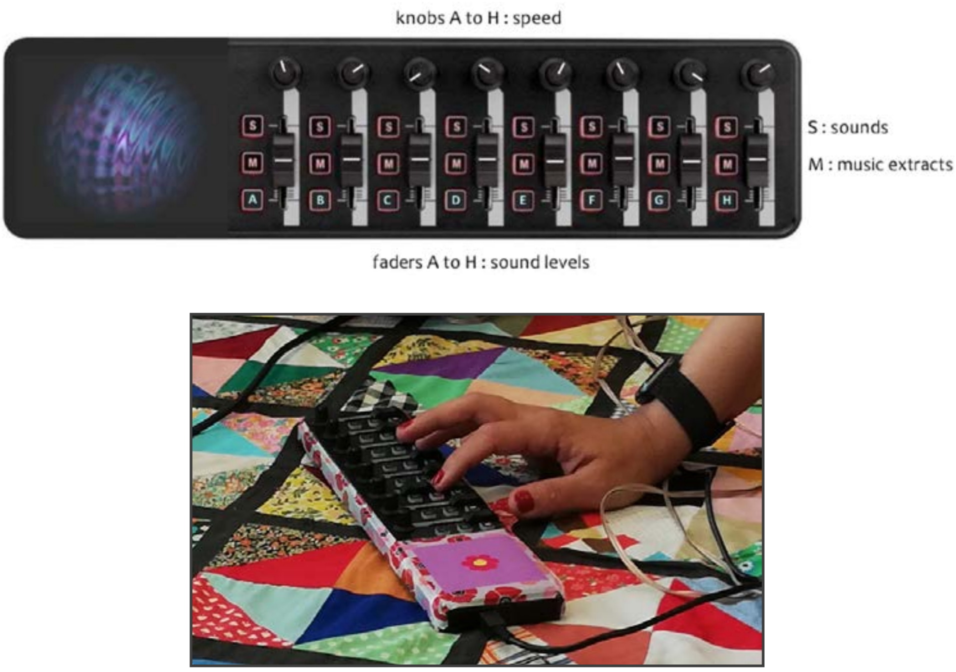
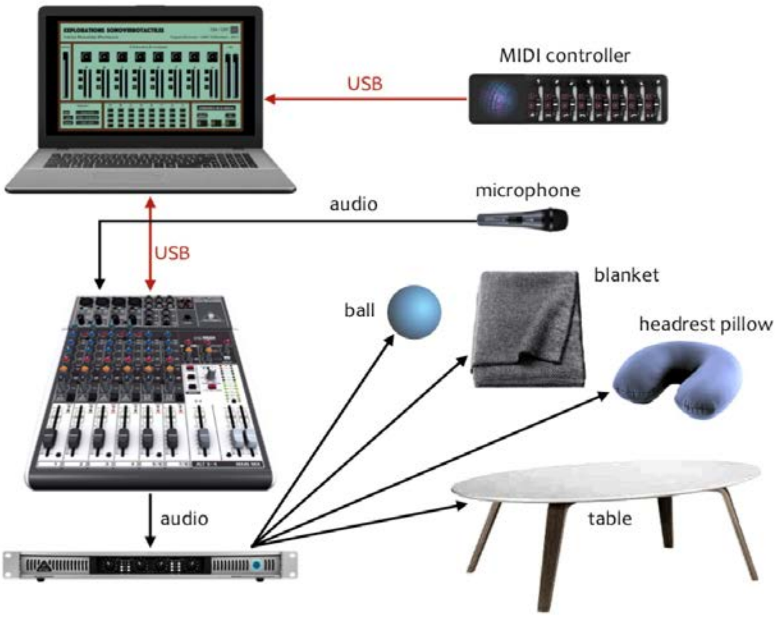


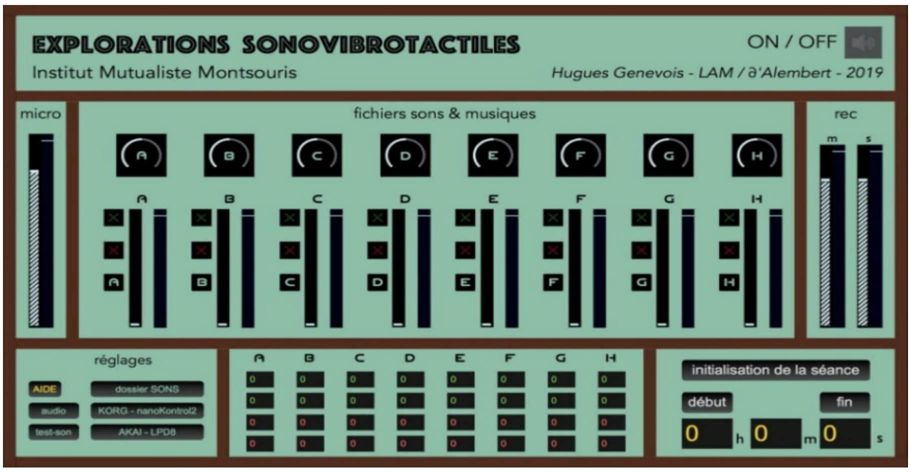
Figure 3. The MIDI interface allowed patients to manually control the stimuli, both the choice of sound and the character of that sound (i.e., sound level and speed).

to the computer via a USB interface. A small mixing console (Behringer Xenys 1204USB) functioned as a digital–analog converter and fed the various mediating devices via a multichannel amplifier. It also powered the microphones during the sessions. For the purposes of analysis, the software tool recorded the patients' choices of sounds and the subsequent audio signal transmitted to the mediating objects (Figure 4).

Stimuli were activated almost exclusively by the patients, who had constant control over the MIDI interface. If the patients asked the clinician to select or modify stimuli for them, they could give access to the interface to the by placing it outside the carpet, which was a clinically significant



(a)



(b)

Figure 4. The technical installation components (a) of the framework for audiovibrotactile interaction by anorexia patients via mediating objects. Item (b) provides a closer image of the software tool.

action. The length of the cables connecting the devices to the computer and amplifier was sufficient to allow for these movements. All mediating objects were permanently activated and it was not possible to disconnect them during a session. Patients could choose the objects through which they wanted to explore stimuli. In this sense, the gesture control MIDI interface allowed them to personally manipulate the mediating objects and vibrasonorous phenomena in order to modulate their own bodily sensations. Using the MIDI interface to gesturally control the sound vibrations and music meant that the subject must assume, psychically and corporeally, the position of the agent—that is, to avoid passively receiving stimuli from the outside—in order to modulate his/her sensations based on his/her experience. This possibility of modulation transforms the vibration and sound elements into plastic objects (Pankow, 1977).[8] For anorexic patients, what is at stake is being able to modulate the bodily sensations that produce anxiety: The deep connections with the human drives are difficult to contain yet without having to use control as a defense to detach from the drives completely. This therapeutic mediation framework, therefore, helps the adolescent find a balance between staying in control and letting go when experiencing these sensations.

Data Systematization and Analysis

The observing psychologist took verbatim notes of patients' speech during the workshop sessions, both their free comments during exploration time and their responses to the final questions. The observing psychologist took notes also of patients' behavior related to selected sounds, as well as the use and duration of the contact with the objects (description and timing with chronometer). She noted body postures when the patients entered in contact with the objects. These notes of patients' speech and behavior were systematized, that is, coded and quantified. The patients' body contacts with each mediating object were counted and their body postures with these objects were classified and counted. The patients' verbal utterances were numbered. The total number of sentences uttered directly referring to the sounds (all sessions and all patients together) were classified by sound stimuli. For each sentence, the significant semantic units (differentiated as nouns, verbs, adjectives, syntagma, etc.) used to name, describe, and qualify each sound were indexed and counted.

Additionally, the events of the workshop and the patients' speech were transcribed after each session by both psychologists in order to compare their perceptions of the observed situations and the verbal expressions. We could then analyze patients' verbal and nonverbal expressions, contextualizing their behavior and speech.

Furthermore, in developing the software tool, we employed a "chirographic" recording method to record the patients' actions on the gesture control MIDI interface. Such recorded data involved the choice of specific sound and music sequences, how long they were played (start/stop), and the changes in playback levels and speed. This automatized digital recording could measure the gestural data generated by the patient's handling of the interface during the workshop sessions, allowing quantitative gestural data analysis. In this way, the patients' actions and certain types of data on their bodily activity were recorded in a noninvasive manner, that is, without a need for body sensors.

Data systematization grids were designed specifically for this experiment to capture verbal and nonverbal (i.e., movements, body positions) data. We defined systematic analytic categories that were extracted from the data collected by the experiment in order to address the research questions formulated at the beginning of the study. Thus, we could quantify—mainly in terms of percentage, but also with mean, median, mode, variance, and standard deviation—the patients' verbal and nonverbal data related to the stimuli and mediating objects. In this way, the quantitative analysis complements both the qualitative and quantitative data gathered for this mixed methods research study.

RESULTS

The young patients showed good adherence to the workshop: Only one patient of the eight decided to stop the experiment after the first session. All other patients attended the entire series of five planned workshop sessions.

Use of the Mediating Objects

In the presence of the clinical psychologists, each of the patients approached the setting in a highly individual manner, exploring his/her sensations through gestures and movements, and expressing them verbally. Each object lent itself to different types of use, following the patients' personal preferences and the suggestions made by the psychologists when patients requested it, depending on the conditions of each session. In regard to the data collected, we made two kinds of measurements: the number of occurrences of each object's use by each patient and total use by all patients together. The overall data of the objects selected and used by the patients during the five workshop sessions show that the ball was the most frequently used object (28% during all workshop sessions, all patients together), followed by the table (21%) and the headrest pillow (20%). As for the remaining objects, the microphone was used relatively rarely (13%). However, the data show that none of the objects was completely ignored (Table 2).

We chose the number of occurrences of use as a unit of measure to mark the patients' engagement with the mediating objects. The number of occurrences of object use during all workshop sessions indicates one dimension of real use (i.e., how many times all patients turned to and grasped an object). In this sense, it is an indicator of the degree of patients' activity in relation to the objects. It can express an inner state, a search, or an interest of patients that needs to be analyzed in relation to other indicators in each situation. Indeed, a patient could take an object and drop it quickly. However, returning to it several times, instead of ignoring or abandoning it, indicated a certain engagement by the patient in that specific situation, a search of some sort in

Table 2. The Percentage of Mediating Objects Use During All Workshop Sessions for All Patients.

Object	Occurrences of use
Ball	28%
Low, oval table	21%
Headrest pillow	20%
Blanket	18%
Microphone	13%

relation to this object, even if he/she could not always settle in a long or deep exploration. The value of this indicator represents, therefore, only one dimension of behavior that should be interpreted in the light of other data collected in the context of the workshop.

Another indicator of patients' activity with the objects was the duration of use, that is, how much time patients spent in contact with an object. From this standpoint, the total duration of use of each object showed that the patients stayed generally longer with the headrest pillow (34% during all workshop sessions, all patients together), than the blanket (26%), the table (19%), the ball (18%), and the microphone (13%). This measurement also indicates a prolonged body contact with some objects, perhaps with significant subjective engagement, even though those objects were chosen less frequently. Duration of use can express patients' inner state that should be analyzed in relation to other elements gathered in each situation. Therefore, the value of this indicator is not absolute and calls for interpretation within the context. For example, a patient could sometimes take an object and simply forget that he/she is holding it, without exploring or even paying attention to it. The psychologists observed this phenomenon particularly with two patients and in relation to changes in their inner state and the dynamic of the clinical situation during workshop sessions. When they felt depressed or anxious, particularly when talking about their life, they less actively engaged in the exploration of the objects. This happened more often with the headrest pillow and the blanket, which are objects that can be put on the body without needing to be actively held or engaged by the subject. In a future experiment, the frequency and the duration of use should be more thoroughly compared and discussed in the light of qualitative case studies.

What the patients said about their preferences regarding the objects, due to their properties or the sensations they procured, is significantly related to their actual use of them, indicated by the frequency of use (i.e., number of use occurrences), when considering all patients together. However, these data also show a gap between the subjective appreciation of an object (i.e., the value assigned to it by a patient) and the actual use he/she made of it within the situation. Thus, 67% of the patients said they preferred the table because it provided more powerful vibrotactile sensations; 50% said they liked the ball because it induced localized inner sensations, for example, when placed against one's stomach or chest. Likewise, 50% liked the blanket's soft and enveloping texture when lying down on it or wrapping it around their shoulders or waist, even though the object generated fewer vibrotactile sensations. The headrest pillow and microphone represented interest by 33% of patients. The patients were not asked to choose a single preferred object exclusive of others. Moreover, from one session to the next, patients would speak openly about the changes in the way they appreciated the vibrotactile objects. Therefore, the total percentage of object preferences sums up to more than 100% (Table 3).

The difference between the subjective appreciation and the actual use of the object could be due to different reasons. First, as we noted in the Introduction, anorexic adolescents often overinvest in the world of perception and action. Discussions with the IMM clinical team led us to consider that, depending on their inner states, patients with anorexia tend to seek either strong sensations or holding. In this sense, the wooden table and the compact foam ball could appeal to patients looking for sensations because these objects allow a better reception of vibratory sensations. Additionally, as we indicate below, both objects, probably due to their form, allowed a variety of modes of grasping and use, and this could facilitate their psycho–corporeal appropriation by patients. The preference of half of the patients for the blanket could be explained by their need of holding, according to their inner state at a given time. Second, the real use can indicate the patients' search and exploration of the objects. However, it does not

Table 3. Percentages for Object Preferences for Specific Audiovibrotactile Mediating Objects as Expressed by All Patients Over All Workshop Sessions.

Object	Occurrences of use
Low, oval table	67%
Ball	50%
Blanket	50%
Headrest pillow	33%
Microphone	33%

directly express their subjective experience of the objects, such as the way the subject represents the qualities of an object according to the sensations, the affect, and the memories aroused by contact.

Body Positions Adopted with the Mediating Objects

Certain objects at times were handled in a highly personal way, particularly the ball and the table. The various postures adopted in relation to these objects pointed toward, on one hand, the framework's high degree of plasticity and adaptability to personal preferences. On the other hand, they also suggested that most patients preferred or decided to stimulate different areas of the body in a range of positions and postures and in relation to the sensations generated by each object. Moreover, the observation of the uses of the mediating objects revealed a wide range of possible configurations, showing each patient's unique preferences, depending on his/her psychic availability and state, which varied from session to session.

Furthermore, patients typically explored spontaneously each mediating object individually. However, they engaged at times two or more objects simultaneously during the sessions on their own initiative and based on their own preferences—as well as with the suggestions from the psychologists, according to the situations at hand. The most frequent combinations of the mediating objects were the pillow and the ball and the pillow and the blanket (both 30% of use occurrences, all patients together). The next most common combination was the simultaneous use of the ball, the pillow, and the blanket (i.e., the same three objects used together). On two occasions, all four mediating objects were used simultaneously by one female patient.

The ball and the table encouraged the most diverse and surprising positions, postures, and physical handling the objects, with six and five categories of use, respectively. (These categories are presented in the figures associated with the subsections below.) They were followed in the classification by the headrest pillow, the blanket, and the microphone, with four different categories of use, detailed below. The pillow and the microphone were associated with more exclusive uses in regard to the possibilities explored by the patients; the table, the pillow, and the blanket led to more equally distributed types of use.

The Table

All patients explored the table without showing any apprehension, either of their own initiative or with the psychologists' suggestion. They had no difficulty coming near it or even lying down on it, except for one female patient who chose to sit rather than lie down. Some patients explored the table spontaneously and frequently, in a highly personal way from one session to the next—by touching it with their hands and arms; with their forehead, cheek, or chest; by sitting in front of it; by sitting, lying down, or standing on it. The most frequent uses of the table were in the reclining position, on one's back or side in a fetal position (83% of patients, 30% of use occurrences) and in a sitting position (83% of patients, 28% of use occurrences). Half of the patients (50%) would sit next to the table to touch and explore it, as well as climb upon it. One particular female patient chose this last position regularly during her five workshop sessions, staying on the table in standing position, sometimes touching it with the sole of her foot, sometimes only with her toes as in classic dance movements (Figure 5).

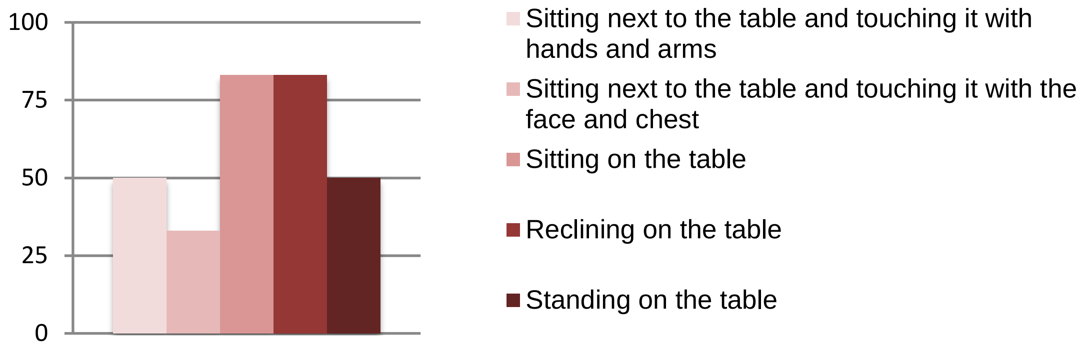


Figure 5. Percentage of uses of the low, oval table by all the patients over all workshop sessions.

The Ball

The ball was most often grasped by the hands (100% of patients, 47% of use occurrences, for all patients). To a lesser extent, the ball was touched without being grasped (83% of patients, 15% of use occurrences), or grasped and placed close to the stomach (50% of patients, 15% of use occurrences) or the chest (33% of patients, 9% of occurrences). When not grasped with hands, patients put one or both hands atop the ball, which was located on the carpet. Even though this interaction with the ball was less frequent, these uses were significant insofar as they presented repeatedly and were common to several patients. Some patients explored also other uses that did not involve grasping with the hands. Other areas of the body and positions were explored, such as placing the ball under the toes, on top or under the head, using this spherical and soft object to feel the vibrations in other places of the body (Figure 6).

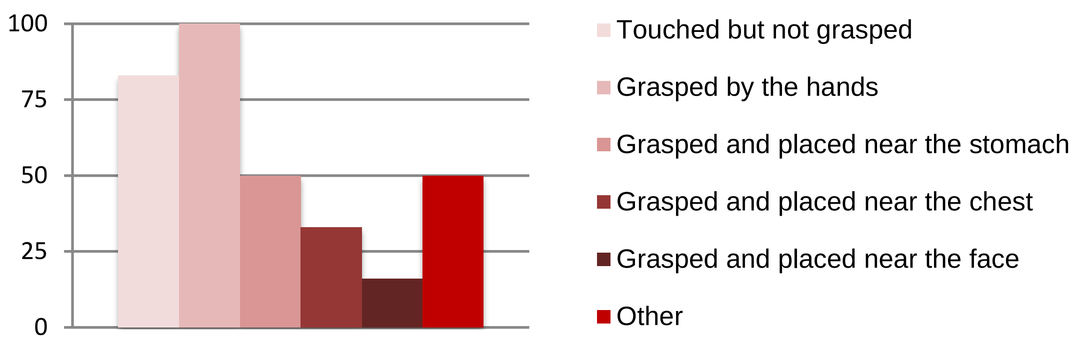


Figure 6. Percentage of uses of the ball by all the patients over all workshop sessions.

The Headrest Pillow

The patients placed the headrest pillow mainly around their necks while sitting or lying down (100% of patients, 83% of use occurrences). However, to a lesser degree, other unique uses were explored: grasping and placing the pillow near other parts of the body, such as the legs, thighs, or the palm of the hand (33% of patients, 10% of occurrences; Figure 7).

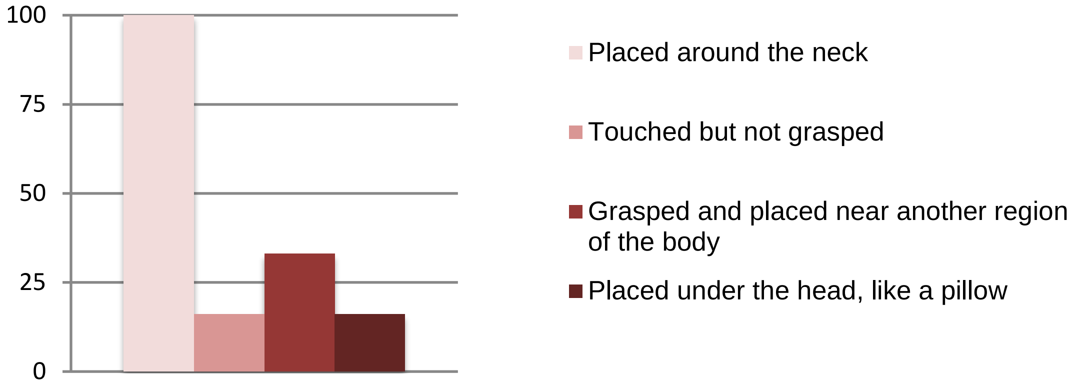


Figure 7. Percentage of uses of the headrest pillow by all the patients over all workshop sessions.

The Blanket

The use of the blanket fell into four categories, equally distributed among the patients. It was used by patients mainly as an enveloping object, wrapped around their shoulders or waist, when sitting or lying down (83% and 67% of patients, 40% and 18% of occurrences of use, respectively). Significantly, it was used for other explorations as well, especially as a contact surface spread out on the floor, for manual exploration or to lie on (67% of patients, 21% of use occurrences). This object therefore invited other types of exploration beyond those for which the framework was originally designed (stimulation of the neck, shoulders, or shoulder blades), which suggests that its vibrotactile mode encouraged playful exploration due to its softness, flexibility, and ability to cover large areas of the body (Figure 8).

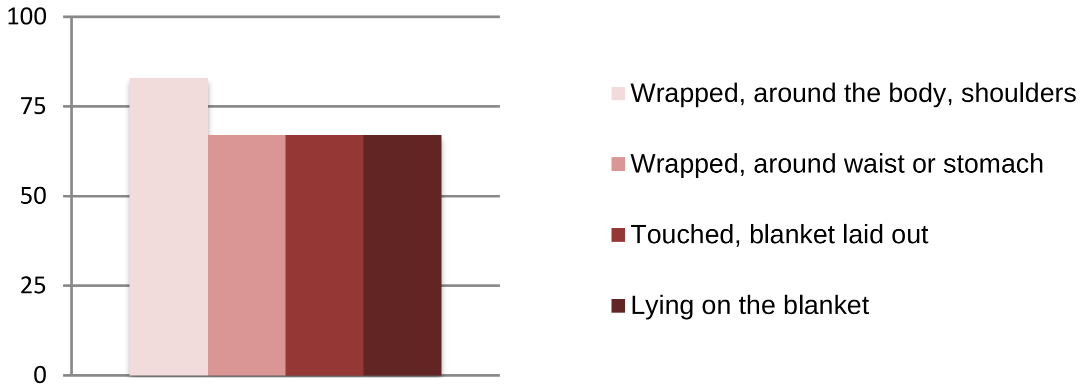


Figure 8. Percentage of uses of the blanket by all the patients over all workshop sessions.

The Microphone

Although it was used relatively less frequently than the other objects, the microphone was explored in one way or another by all patients (100% of patients, 54% of occurrences), especially to speak to the interacting psychologist (only one of the psychologists used the microphone for a regular exchange). In this sense, the sessions showed that the microphone helped create a verbal exchange between the patient and the psychologist. The microphone is not the same type of mediating object as the table, blanket, pillow, or ball, which all transmitted sound via their vibrating properties. Rather, it functioned similarly to the MIDI interface: It transmitted an acoustic signal to the mediating objects, but does not itself vibrate. However, it differed from the MIDI interface's role of transmitting recorded sounds in that it captured the acoustic signal produced by the patient in real time and transmitted it to the vibrating mediating objects. The microphone thus served as an unprogrammed sound source, enabling patients to express themselves spontaneously by using their voice, but also by rubbing the microphone or producing nonvocal sounds—and receive sensory feedback.

Certain patients expressed, at a particular moment, an uncanny feeling on hearing their own voice or feeling its vibrations (e.g., "*It's bizarre;*" "*strange,*" "*like a story within a story*").[9] Yet half of them (50%) also used the microphone to generate percussion sounds, to accompany impulse-like sounds such as the heartbeat (BBC, n.d.c) or to produce the sound of a galloping horse in imaginary sound scenes. A female patient repeatedly explored vocal sound effects such as breathing, sighing, and tongue-clicking to create scenes with other sounds, for example, the sound of stone skipping (Figure 9).

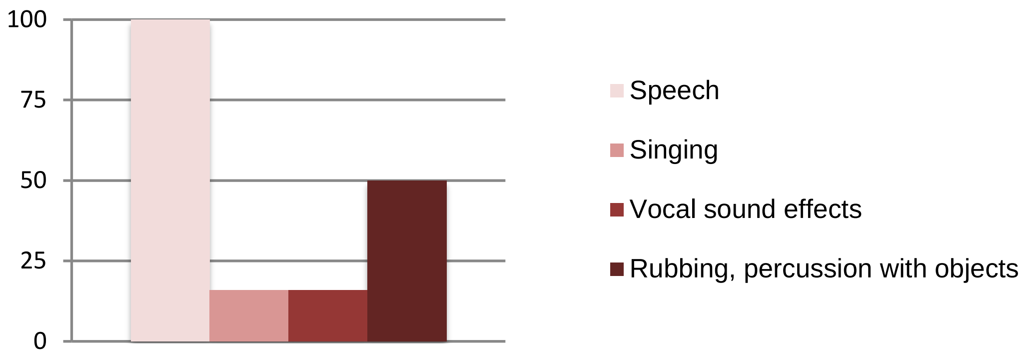


Figure 9. Percentage of uses of the microphone by all the patients over all workshop sessions.

The Choice of Sound and Music Sequences

All the sounds were explored by the patients more or less repeatedly according to their personal preferences. On the one hand, the sounds were mostly chosen by the patients. However, in some circumstances—on the patient's request, depending on his/her mental state or a difficult associative moment—certain sounds were suggested to patients by the psychologists in order to (re)start the process of association. On the other hand, the patients listened to the sounds generally through body contact with the audiovibrotactile mediating objects, and, thus, they simultaneously felt sound vibrations. They sometimes verbally expressed their feelings about music and vibrotactile mediation: "*I like the vibrations. I prefer them to sounds*"; "*I like the sensations—with music, I'm more used to it … but the sound also helps because it gives you ideas at the beginning*"; "*It was* *nice, the sounds brought different scenes into my mind—I like the vibrations; "I find it more pleasant when the vibrations are stronger."* Nevertheless, the patients sometimes explored sounds or music without necessarily using an audiovibrotactile mediating object. They would perceive them in a purely auditive way (through the aerial acoustic radiation produced by the vibrating mediating objects). The framework thus enabled a type of back-and-forth movement between tactile sensations and auditory perceptions.

The sounds that patients most often identified as their favorites were the sound of waves related to the sea (BBC, 1981; 83% of patients), heartbeats (BBC, n.d.c; 83%), and the musical excerpt "Bilude" (Schaeffer, 1979/2010c, CD 3, track 17; 33%), according to the responses the patients gave to one of the questions posed at the end of each session. Other sounds were considered pleasing and calming because of the scenes or sensations they evoked, such as the sounds of flowing water (BBC, n.d.d), of footsteps on grass (BBC, n.d.b), or the two excerpts from the soundscape music piece "Presque Rien N°1" (Ferrari, 1970/2009a, CD 3, track 1, 1990-1998/2009b, CD 3, track 9). Certain sounds, such as the train (Schaeffer, 1948/2010a, CD 1, track 1), the bubbling water (Genevois, 2019), or isolated water drops (Genevois, 2019), were deemed pleasant by some patients but unpleasant by others. However, feelings about particular sounds could change during the course of a single session or from one session to another, depending on the patient's mental state, the moment in the verbal discussion with the psychologists, or the process of modulating and combining the selected sounds. Based on what the patients expressed verbally during the sessions, sound sequences such as "Étude aux Tourniquets" (Schaeffer, 1948/2010b, CD 1, track 2), "De Natura Sonorum" (Parmegiani, 1974/2008, CD 5, track 9) and "Étude aux chemins de fer" (sound of trains; Schaeffer, 1948/2010a) were perceived as unpleasant and even anxiety producing by some patients because of the psychic associations and bodily sensations they induced: "*It makes me think of home, with strange beings, strange beings who ring bells*" and "*I'm afraid of the train because on the train you often have to eat. I can't wait to satisfy my hunger, so I eat everything very quickly, and suddenly it's complicated.*"

Moreover, the same sound could refer to a variety of ideas and emotions (of pleasure or displeasure), depending on the playback speed and sound level. These two sound parameters could lead the patients to contrasting affective and emotional situations and positions, revealing an affective ambivalence. Depending especially on the playback speed, sounds perceived as pleasant and calming could become unpleasant and frightening: an accelerated heartbeat (BBC, n.d.c), slowed-down dripping water (Genevois, 2019), the "Bilude" (Schaeffer, 1979/2010c, CD 3, track 17) piece played more slowly or more quickly. For example, several patients liked the "Bilude" excerpt because it alternated between light piano melodies and the sounds of everyday life. However, when played more slowly, the same excerpt led to the following associations in one female patient: "*It's a little more morbid", "It sounds a bit sinister", "…a haunted castle, a great mansion or a cemetery, there are monsters and …it's scary and … it's as if there was a huge skeleton playing the piano", "It's all white and … it's scary I think and… it's enormous."* The sound of heartbeat could be felt as pleasurable at a normal or slow speed (e.g., "*I like the heartbeat*"; "*You need to hear it, don't you?*") and disagreeable or anxiety-provoking at a higher speed ("*It's stressing me out—I don't like it so fast"*). Therefore, the very identity of the sound, not only the ideas and emotions the patient associated it with, would sometimes change depending on the sound levels and playback speed variations. For example, when slowed down, the sound of the train in "Étude aux chemins de fer" (Schaeffer, 1948/2010a, CD 1, track 1) led some patients to the impression of entering an underwater world (e.g., "*It's like going into the water*"; "*… a bit like we were underwater—it's a little blurred*") or being at the heart of a storm (e.g., "*It looks like a thunderstorm—I have a little storm*"; "*I imagine a wind and it moves an antenna—a lighthouse antenna by the sea*").

The analysis of the verbal associations and the vibrasonorous play (gestural control of the MIDI interface by the patients) suggested that the manifest preferences for certain sounds often corresponded with the recall of these sounds during the session. So, the analysis of the total accumulated duration of each of the sounds triggered, when considering all patients and all sessions together, shows that the sound of heartbeat (BBC, n.d.c; 14.8%), flowing water (BBC, n.d.d; 10.3%), footsteps on grass (BBC, n.d.b; 9.3%), and the excerpt from "Bilude" (Schaeffer, 1979/2010c, CD 3, track 17; 9.1.%) were activated for the longest period of time. Sounds of bubbling water (Genevois, 2019; 3.2%) and isolated water drops (Genevois, 2019; 3.1%) were explored to a lesser extent. However, patients also spent significant time listening to the extracts from "De Natura Sonorum" (Parmegiani, 1974/2008, CD 5, track 9; 7.2%), "Étude aux Tourniquets" (Schaeffer, 1948/2010b, CD 1, track 2; 4.8%), and "Étude aux chemins de fer" (Schaeffer, 1948/2010a, CD 1, track 1; 4.4%), which were nevertheless perceived as unpleasant or stressful by certain participants. These sounds were not explored to a lesser degree than others, such as the traditional music from Benin or Bali (Rouget, 1990; Sentana, 1991, track 4) or the excerpt from "Presque Rien N°4" (Ferrari, 1990–1998/2009b, CD 3, track 9; Figure 10). Therefore, sounds that had not been explicitly perceived as preferable or pleasant by patients were still used nevertheless and also generated associations, as expressed by the patients.

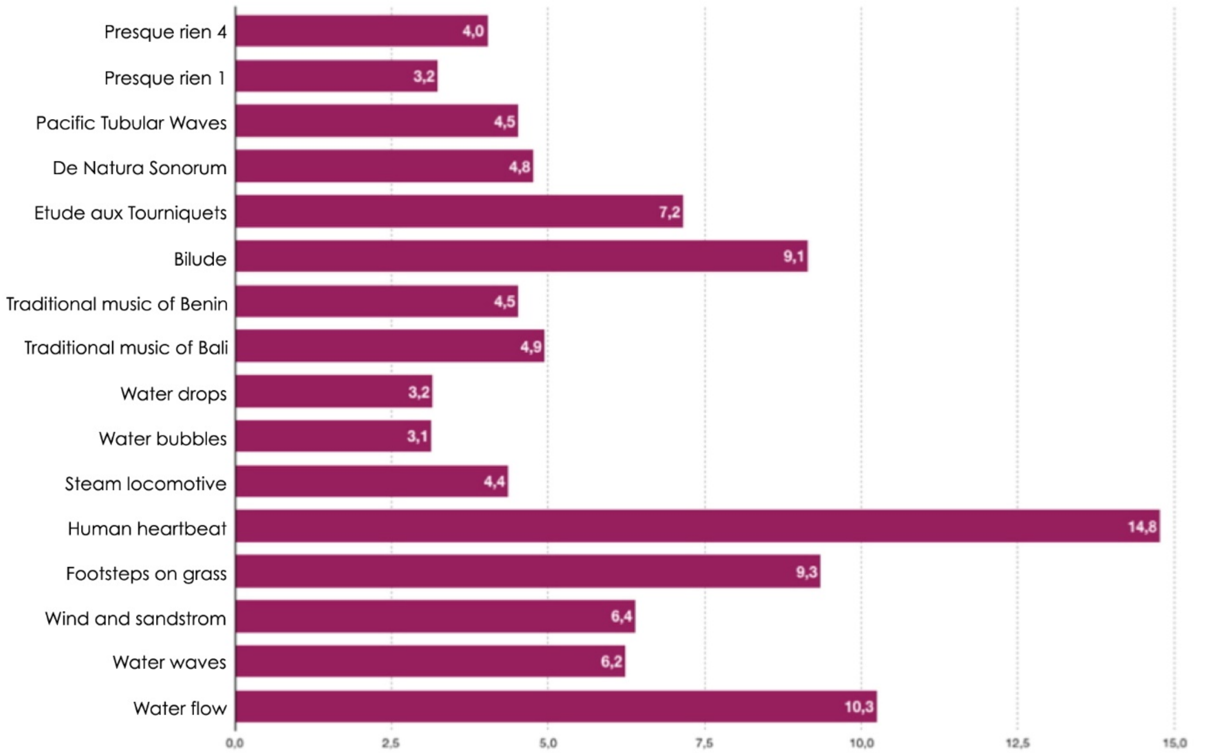


Figure 10. Histogram of time spent on each sound, all patients combined during all workshops. The x-axis expresses the percentage of activation time of each sound in relation to the cumulative duration of all sounds.

The relative differences between the verbally expressed “preference” for certain sounds and the actual use of these sounds in the sessions by patients and psychologists required us to consider the associative relevance of the different sounds, that is, their ability to mobilize and highlight deep affect and representations, which involve the psychic conflicts to be elaborated through therapeutic work. Just because a patient perceived and identified a sound as unpleasant or strange does not mean that it cannot generate associations indicative of psychic contents or help elaborate them.

Although each sound was explored separately, at times they also were activated together by the patients—and sometimes by the psychologists when patients asked them to do so. Consequently, it was frequently a combination of sounds and music that led to verbal associations with various sensations, ideas, and emotions. In particular, the sounds evoking water and the sea (i.e., the sound of flowing water in a natural environment [BBC, n.d.d], the water waves [BBC, 1981], or the wind and the sandstorm [BBC, n.d.a]) were activated and grouped together by the patients as sounds linked to “the sea” (Figure 11).

Verbal Expressions Linked to Sound and Music Sequences

Patients used sounds to nonverbally express their singular mental images and affect. Sometimes, they verbally expressed these representations and affect associated with sounds. We analyzed the semantic content of their verbal utterances directly referring to the sounds. The percentages below represent the number of occurrences of the semantic units relative to the total number of these units counted for all of the sounds.

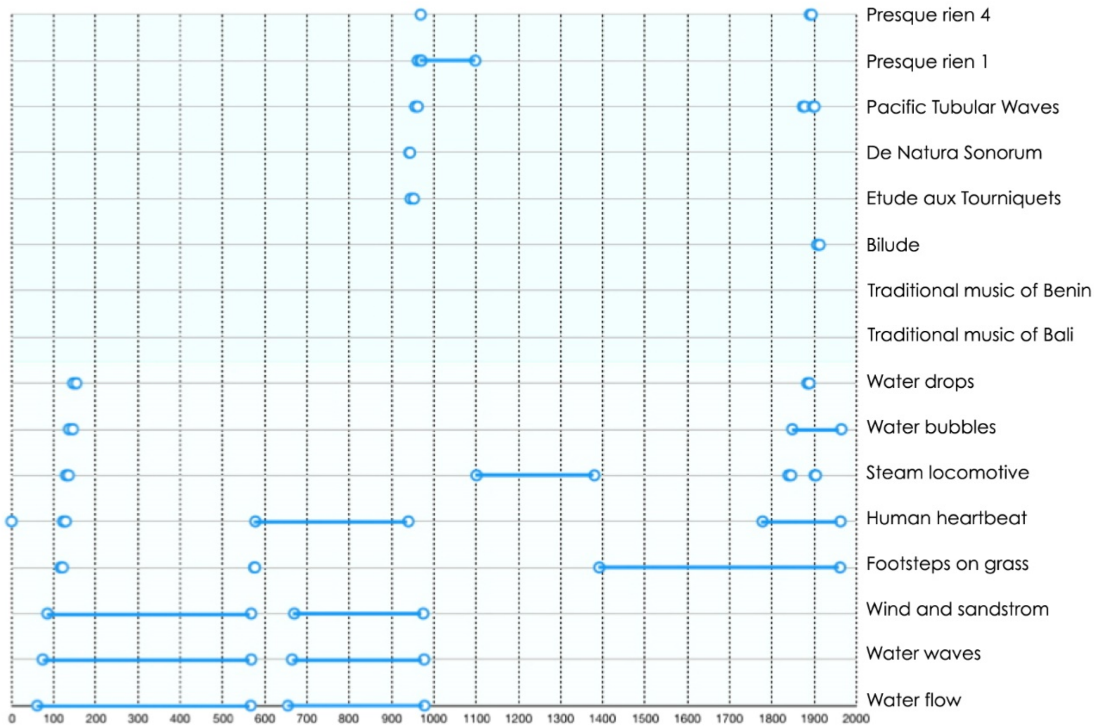


Figure 11. Temporal representation of the sounds triggered by a patient during one session. The x-axis expresses time, marked here by a vertical line every 100 seconds (temporal representation scale of the duration of the session).

The sounds that provoked the most verbal associations directly linked to their sensory, sonorous, and vibrotactile properties were the sounds of the waves related to the sea (BBC, 1981; 12% of occurrences); the sounds of flowing water (BBC, n.d.d; 10.6%); the second excerpt from Ferrari's "Presque Rien N°4" (1990–1998/2009b, CD 3, track 9), which was associated with the countryside (10.6%); the sounds of footsteps on grass (BBC, n.d.b; 9.3%); the extract from "Bilude" (Schaeffer, 1979/2010c, CD 3, track 17), including piano music and sampled everyday sounds (8.8%); the wind and sandstorm (BBC, n.d.a; 8.4%); and the sounds of heartbeat (BBC, n.d.c; 7.5%). These results are presented in Table 4. The discourse analysis of the verbal associations suggests that the available sounds and music lent themselves to a wide range of unique associative chains, specific to each patient, beyond the few common themes that we could identify: groups of unfamiliar beings (e.g., "*the tribe*"; "*strange beings*"; "*nonhumans*"; or "*extra-terrestrials*"), bathing, swimming on the surface of or under the water, horses and galloping, the storm and the tempest, and travel and movement. The motifs of water—as the sea, rain, storm, or tempest, and including situations such as swimming on the surface or under water, or even drowning—appeared very frequently (36% of occurrences). Sounds linked to

Table 4. Percentage of Verbal Associations Produced by All Anorexic Patients to the Stimuli During All Workshop Sessions.

Sound/music stimuli	Verbal occurrences (%)
Water waves (BBC, 1981)	12.0
Water flow (BBC, n.d.d)	10.6
"Presque Rien N°4" (Ferrari, 1990–1998/2009b, CD 3, track 9)	10.6
Footsteps on grass (BBC, n.d.b)	9.3
"Bilude" (Schaeffer, 1979/2010c, CD 3, track 17)	8.8
Wind and sandstorm (BBC, n.d.a)	8.4
Human heartbeat (BBC, n.d.c)	7.5
"Presque Rien N°1" (Ferrari, 1970/2009a, CD 3, track 1)	5.7
"De Natura Sonorum" (Parmegiani, 1974/2008, CD 5, track 9)	4.8
Steam locomotive (Schaeffer, 1948/2010a, CD 1, track 1)	4.4
Water drops (Genevois, 2019)	3.5
Music of Benin (Rouget, 1990)	3.5
Bubbling water (Genevois, 2019)	3.1
"Étude aux Tourniquets" (Schaeffer, 1948/2010b, CD 1, track 2)	3.1
Music of Bali (Sentana, 1991, track 4)	2.2
"Pacific Tubular Waves" (Redolfi, 1979/1988, track 6)	1.7

the images of the sea also brought up associations with family holidays. The sounds of water waves related to the sea (BBC, 1981; 18.5%), flowing water (BBC, n.d.d; 17%), footsteps on grass (BBC, n.d.b; 14%), and the wind (BBC, n.d.a; 10.5%), therefore, induced associations with the names of places linked to individual memories. For example, the sound of waves related to the sea evoked precise geographical locations such as Nice, Playa del Carmen in Mexico, and Morbihan in Brittany, Biarritz, or Dunkirk, referring to the patients' life experiences. Other sounds also fulfilled this function, though to a lesser extent.

After having spontaneously identified the sounds via their referents (the sea, the storm, dry leaves, a train, etc.)—at times using objectifying language in the third person ("*It's a…*"; "*It could be called…*") and at other times more subjectivizing language in the first person ("*It makes me think of…*")—most patients would let themselves engage, with more or less difficulty, in more personal free association, often encouraged by the psychologists. In other words, for a given patient, a particular sound could refer to a general object or phenomenon in the outside world, as well as to a unique situation. For example, for one female patient, the "Bilude" piece (Schaeffer, 1979/2010c, CD 3, track 17), the sound of footsteps on grass (BBC, n.d.b), and the wind and sandstorm sound (BBC, n.d.a) led to the following association: "*A boy is trying to escape—he's running away from his parents.*" And even though all of the patients identified the railway sounds (Schaeffer, 1948/2010a, CD 1, track 1) as a train, for one patient, this was specifically the train "*to go to the university.*" The sounds of flowing water (BBC, n.d.d), water waves (BBC, 1981), and the wind (BBC, n.d.a) universally evoked imagery of the sea but, for one patient, these and other sounds generally reminded her of tempests and storms, repeatedly featuring the theme of drowning: "*We are at the seaside and there is some music and a storm is coming," "It's going to break the piano," "It's broken," and "I'm drowning."*

The direct reference to the sounds did not exclude another level of verbal associations linked to the patient's life experience and memories that were not based on or directly refer to these sounds. This level of associations linked to patients' life experiences was very present. The sounds and their vibrations could therefore embody and support the elements of the patients' experiences, facilitating mental representation and expression. As a result, sounds could function as a kind of background projective material (e.g., elements of the soundscapes and imaginary scenes), helping patients express their experiences—affect and representations—without having to explicitly refer to the perceived sound (i.e., the description and qualification of the stimulus).

DISCUSSION

The pilot study investigated how patients suffering from an anorexia nervosa explored sound-initiated vibrotactile mediating objects within a space delimited by a carpet, in a process of associative expression. The experiment raised questions that should be investigated further in the future. Some of these questions relate to the clinical function of the mediation framework with respect to the patients' responses to it. We can address several topics for a general analysis in regard to the results.

In terms of the physical and mental appropriation of the mediating objects, the bodily positions and the use of the various audiovibrotactile objects highlighted the particular psychic positions of each patient, depending on the specific aspects of his/her history. For instance, we could see that the data gathered in study could support the probability of the patients seeking to contain and relieve threatening emotions; expressing a need for control exercised through demanding bodily postures; demonstrating intense bodily activity expressing fantasy scenarios; and/or searching for a reassuring object in order to feel one's body exist. These clinical categories were qualitatively drawn from the intersection of various kinds of information: how patients used the sounds and the objects during the sessions; what patients said in each workshop session about themselves, their situation, and emotions; and general information about patients' medical treatment communicated by the clinical team at the beginning and the end of the experiment. Thus, some patients spoke about their feelings related to their past and present lives, expressing anxiety, anger, or sadness. These patients sought contact with the mediating objects through quiet body postures, such as lying down, placing the ball near their stomach or chest, or wrapping themselves in the blanket. They often expressed a search for relaxation: "*The session today helped me relax*"; "*It feels like a warm container and at the same time covering, it helps me relax.*" Patients also sought to control their emotions through demanding bodily postures, such as standing on the table upright on the toes for long minutes, or by looking for strong sensations (e.g., "*With the feet, it spread all over the body while with the hands it spread differently—it's because I practice dancing*", "*It reminds me of memories, with cousins—it was nice—my heart was beating very fast, I was stressed*"; "*A little bit like heartbeats, it makes you feel something, we live a little like that*", "*I was a bit into the sensations and the vibrations, because in the end I tried to put all the music together*"). Moreover, a female patient used the objects to bodily enact her own stories and fantasy scenarios (e.g., "*There we will have to clean up. We are going to try to make a house ... here we are, now we are going to build a huge bed, a bed with 30 places ... we must build the table*"). In these various situations, patients used objects to reassure themselves. This analysis on the clinical functions of the uses of the platform for therapy should be developed more fully in a future experiment.

Contrary to our initial hypotheses, the patients did not express verbally or nonverbally (i.e., via speech, smile, eye contact, body posture, initiating use, or by frequency/duration of use) any significant difficulties or concerns in entering into contact with mediating objects. In the case of the low, oval table, most seemed pleased to explore it by using a wide variety of bodily positions, that is, patients could lie or sit on the table as on a large bench. Perhaps this is because, outside its usual social context, the low, oval table has a deliberately ambiguous shape and thus evokes functions other than that of the symbolic object associated with food. Thus, the table provided an opportunity and possibly encouragement for patients to invest in their psycho–corporeal relationship with the object in other ways. Because of its physical properties, the table also transmitted the most powerful vibrotactile sensations. On the opposite end of the usage scale, the blanket was less effective in this sense, although patients did like its soft and enveloping texture, which fulfilled a containing function. A future study will require improving the technical implementation of this device.

The patients did not avoid any of the sound sequences. However, the sounds and vibrations associated with the sea and the heartbeat were particularly meaningful for a number of patients (e.g., "*It's like I can really feel my heart beating.*"). These particular sounds encouraged the emergence of sensations, affect, and memories, some perhaps buried deep within the body. However, this outcome was not due to the sound phenomenon per se, but rather to the potential of transmitting sound through vibrations in the clinical framework as a whole, which facilitated the process of association.

The use of the gesture control MIDI interface was clinically significant in the workshop. Allowing patients full control in turning on and off and modulating sound excerpts was intended to prevent the adolescents being passively exposed to stimuli, given the dependency and control issues associated with anorexia nervosa. As noted in the Introduction, anorexia sufferers' need for control to cope with internal strains, albeit at his/her own expense, is a symptom of the disorder. Thus, our clinical approach to give these young patients the possibility to control some parameters of sound and vibration—and to manipulate the mediating objects themselves in order to modulate their sensations—made good sense. We believe the patients' appropriation of the mediation platform was a condition for them to appropriate elements of their unique personal experiences. In this sense, the use of the gesture control MIDI interface by the patients offered a condition to work with them on more flexible modes of control, where the issue is appropriation and modulation of sensations and emotions. The psychologists observed that patients more often used the binary control option of the MIDI interface to turn sound sequences on and off. They were generally less inclined to spontaneously use the modulation function of the interface. However, when invited to explore this function, some of them used it to explore their feelings, talk about their experiences, or imagine associative scenes. Nevertheless, the clinical function of the gesture control interface should be more systematically assessed in a future experiment.

Although the framework was particularly effective with some patients, others found it difficult to express their associations. Two female patients in particular brought forth relatively few associative elements. They opted for descriptive and operative statements about the sounds, objects, and vibrotactile sensations, using objective language; alternatively, they would simply say whether they found the sensations pleasurable or not (e.g., "*I don't like it*", "*It's not satisfying*"). As a result, the psychologists sometimes were confronted with the patients' difficulties in entering into the audiovibrotactile play due to inhibition, resistance, or by attacking the therapeutic alliance (e.g., "*I don't know what to do anymore*"). In these situations and respecting the process of each patient, the psychologists used techniques of punctuation, reformulation, and relaunch by making open suggestions.[10] In other words, they had to find a way of intervening in the setting to overcome these problems, to restart the process of play, and encourage association. This observation suggests that the sound, music, and vibration mediation framework requires a clinical setting that should include psychologists as mediators between the patient and the objects. The clinical setting of this experiment, especially with vulnerable patients, thus demands the presence of trained professionals.

Moreover, despite its facilitating function, the mediating framework cannot entirely replace therapy. It is inspired by therapeutic methods, such as the Squiggle Game (Winnicott, 1971b) and the reconstruction of body image through clay modeling (Pankow, 1977). However, the framework was not conceived as a form of therapy in itself; rather, the framework serves explicitly as a facilitator of therapeutic work conducted in a separate and specific setting. It therefore should be used in conjunction with the work of an institutional team.

The results of this pilot experiment were presented and discussed with the IMM clinical team. The feedback was positive, but the short duration of the exploratory study does not allow us to account for its possible medium- and long-term therapeutic effects. The study aimed to assess patients' appropriation of the mediation platform and their responses to it. We assessed the evolution of patients' state from the beginning to the end of the workshop on the basis of their behavior and speech during the workshop sessions. We compared the observations of the psychologists with information communicated by the clinical team on the patients' medical treatment. No patient psychologically decompensated during the experiment. Some of them were able to talk about their symptoms, family situations, and emotions during the workshop sessions.

The evidence from this pilot study seems promising; nevertheless, the study raises questions about the potential impact of this workshop on the usual follow-up treatment of patients by the hospital team. We also need clarity on the issue of how long this therapeutic mediation framework should be employed: Could it be used for more than five sessions with each patient? And if so, would it interfere with other modes of treatment within the institution? A new study could help in understanding whether long-term use might be beneficial, might exhaust the framework's possibilities, or become an object of resistance to the ongoing therapeutic work. Our assumption is that the therapeutic efficacy of this platform is linked to its ad hoc, occasional application: It was designed as a short-term workshop (five sessions) and all but one of the patients agreed to this duration. Its precise beneficial length is to be determined through further research. Moreover, in longer term experimentation in the future, we intend to formalize a questionnaire to assess more thoroughly how this platform influences the inner experience of the patients with respect to their clinical situation.

CONCLUSIONS

The specificity of this platform lies in its ability to offer pliable objects that function as springboards for the inner experience of sensations, emotions, and ideas. The sound stimuli and the mediating objects of this framework have demonstrated a high degree of plasticity, lending themselves to various uses and singular associations. The sounds and their vibrating properties were appropriate as projective materials, which could adopt a different sound identity for each patient, or even for the same patient within the space of a session or from one session to another. This was enabled by the questions, follow-ups, and play suggestions made by the clinical psychologists.

For the patients, the platform fulfilled certain essential clinical functions: relaxing or calming bodily sensations; investing in certain zones of the body by seeking powerful feelings; expressing emotions and associative processes by either creating stories (fantasy scenarios), or recalling memories and talking about current experiences. The analysis of the verbal data centered on the patients' relationship to their sensory experience has shown that the setting encouraged the expression of these associative movements in speech, while providing the patients with a containing framework. In this sense, the mediation platform satisfies the objective of this study.

Its scientific value and originality also lie in its ability to bring together several disciplinary approaches, whose respective abilities and expertise were necessary to implement this research. Specifically, this clinical therapeutic work could not have been carried out without the involvement of musicologists and musical acoustics engineers.

IMPLICATIONS FOR THEORY AND APPLICATION

The framework has demonstrated its potential when working therapeutically with young anorexic patients, whose mental life is organized defensively by exerting strict control over emotions and sensations. This mediation platform gives an important place for body exploration as a means of self-exploration, that is, working with these patients on the contents of their psychic lives from their immediate bodily experiences. Thus, from a research perspective, this study contributes important findings regarding the role of vibrotactile and audiovibrotactile interventions and, especially, the benefit of specific mediating objects, results that can benefit from future studies. The design of this type of mediation platform—intended to not to expose these patients to stimulation in a passive and solitary position—supports future investigation into how to provide patients the possibility to be active in regard to the mediating objects and to allow them to make a transition between control, modulation, and release regarding their inner experiences. Moreover, the actual experimentation with patients and the results indicate that the support of trained psychologists is essential for the clinical implementation of this type of mediation platform. It is important for patients to give a place to the meaning of bodily experience through the associative process expressed verbally. Thus, if adapted to any given clinical context, this experiment could help guide the design of similar mediation platforms for working with adolescents driven to extreme modes of defense or with major risks of violent or self-destructive behavior.

ENDNOTES

1. DSM-5 is the fifth edition of the *Diagnostic and Statistical Manual of Mental Disorders*, published by the American Psychiatric Association in 2013.
2. Max is a visual programming language for the specialized needs of artists, educators, and researchers working with audio, visual media, and physical computing. Max is developed and distributed by Cycling ’74, an American software development company, and is infinitely flexible for creating interactive media software. See https://cycling74.com/products/max-features for more information.
3. The experiment *Histoires sensibles* (2012−2013) and the empirical study with the vibrotactile bracelet (2015) were part of the PANAM project funded by the French National Research Agency.
4. In the “packing” method, a patient’s body is wrapped in wet fabric within a psychotherapeutic context. This method is used at the IMM to reconstitute the psycho−corporeal envelop of the patient and is built upon the human relationship between the patient and the hospital practitioner.
5. More of the stimuli were wide-frequency range sources (from 30 Hz to 20 kHz), as were some music selections and soundscapes. A filtering was added due to the specifications of each transducer (i.e., some with full-range frequency response) and the coupling effects of each vibrotactile object (depending on its size, materials, and the level of stimuli). In actual use, the measurement of cut-off frequencies is critical, as this also depends on the area of the body being stimulated.
6. Both the water drops and bubbling water sounds were created by H. Genevois for an immersive installation commissioned by the Centre National de Création Musicale (GMEA) for “The Week of Sound” in 2019, in Albi, France. These sounds are not copyrighted. Genevois created them from two programs he wrote in Max. Both sound excerpts are derived from random generative algorithms. The idea behind these creations was not to make “real sounds” but rather to create imaginary drops and bubbles to question human perception.
7. MIDI (short for Musical Instrument Digital Interface) is a technical standard that describes a communications protocol, digital interface, and electrical connectors that connect a wide variety of electronic musical instruments, computers, and related audio devices for playing, editing, and recording music.
8. The expression “plastic object” refers to an object’s capability for being molded or receiving form—materials such as, for instance, clay, wax, and polymers—especially during a creative process related to art. This type of object can be also referred to as a pliable object because it is flexible and easily adaptable. The German and French psychoanalyst Gisela Pankow (1977) employed this notion to

explain the use of modeling clay in her therapeutic work with patients. Thus, in therapy, a plastic object refers to various mediating external supports that help patients to transform their inner experiences (Brun, 2019).

9. The qualitative data such as the patients' speech were gathered in French. Data quotes were translated into English for publication by a translator who discussed specific wording with the authors of this paper. Additionally, for clarity in this text, commas are used to separate multiple comments from a single patient; a semicolon separates quotes from multiple patients.
10. Punctuation, reformulation, and relaunch are verbal techniques used by the clinician in psychoanalytic therapy and clinical interview to help the patient engage in the expression of an associative process, as well as to recognize some contents within his/her own speech. Punctuation is a way of intervening at certain time to introduce pauses and articulations in the rhythm of the speech; reformulation is when a clinician presents what the patient has said in alternative wording; and a relaunch is to renew the associative process by asking a question, making a comment, or providing a suggestion.

REFERENCES

Anzieu, D. (2016). *The skin-ego* (N. Segal, Trans.). London, UK: Karnac. (Original work published 1985).

Baruch, C. (2009). Les pulsations de l'inconscient [The pulsations of the unconscious]. *Champ psychosomatique, 54*(2), 63–70.

BBC [British Broadcasting Company]. (1981). *Seawash at river estuary; mid perspective waves on single - 1981* (1S4, reprocessed; 07053100.wav). Retrieved from BBC Sound Effects library at bbcsfx.acropolis.org.uk

BBC [British Broadcasting Company]. (n.d.a). *Desert wind - dry, hot, sandy, constant* (07061036.wav). Retrieved from BBC Sound Effects library at bbcsfx.acropolis.org.uk

BBC [British Broadcasting Company]. (n.d.b). *Footsteps, one pair of feet walking in grass* (07041193.wav). Retrieved from BBC Sound Effects library at bbcsfx.acropolis.org.uk

BBC [British Broadcasting Company]. (n.d.c). *Human heartbeat, 58 bpm (average for fit young person)* (07070016.wav). Retrieved from BBC Sound Effects library at bbcsfx.acropolis.org.uk

BBC [British Broadcasting Company]. (n.d.d). *Lapping water in marshes, St Simon's Island, Georgia* (07070227.wav). Retrieved from BBC Sound Effects library at bbcsfx.acropolis.org.uk

Bion, W. R. (1962). *Learning from experience*. London, UK: Tavistock.

Brun, A. (Ed.). (2019). *Les médiations thérapeutiques* [Therapeutic mediations]. Toulouse, France: Erès.

Brun, A., Chouvier, B., & Roussillon, R. (Eds.). (2019). *Manuel des médiations thérapeutiques* [Therapeutic mediations manual]. Malakoff, France: Dunod.

Corcos, M. (2008). *L'anorexie mentale. Déni et réalités* [Anorexia nervosa: Denial and realities]. Rueil-Malmaison, France: Doin éditions.

Corcos, M. (2011). *Le corps insoumis. Psychopathologie des troubles des conduites alimentaires* [The rebellious body: Psychopathology of eating disorders]. Malakoff, France: Dunod.

Corcos, M., Gumy, M., & Loisel, Y. (Eds.). (2019). *Pratiques des médiations corporelles à l'adolescence* [Practices of bodily mediation in adolescence]. Malakoff, France: Dunod.

Cox, A. (2016). *Music and embodied cognition: Listening, moving, feeling, and thinking*. Bloomington, IN, USA: Indiana University Press.

Criton, P. (2014). Listening otherwise: Playing with vibrations. In *Proceedings of International Computer Music Conference and Sound Music Computing Conference* (pp. 1805–1809). Retrieved April 28, 2020, from https://hal.archives-ouvertes.fr/hal-01073237/document

Criton, P., Genevois, H., Falcucci, E., & Patiño-Lakatos, G. (2014). Histoires sensibles. Une expérience de création Vibsonore à l'Institut National de Jeunes Sourds [Sensitive stories: A vibrasonorous creation experience at the National Institute for Deaf Youth]. *Journal de Saint Jacques, 43*, 18–20.

Didier-Weill, A. (1995). *Les trois temps de la loi* [The three times of the law]. Paris, France: Seuil.

Didier-Weill, A. (1998). *Invocations: Dionysos, Saint-Paul et Freud* [Invocations: Dionysus, Saint-Paul and Freud]. Paris: Calmann-Lévy.

Ferrari, L. (2009a). Presque Rien N°1 Ou Le Lever du Jour au Bord de la Mer. In *Luc Ferrari –L'Œuvre Électronique*, 10 CD Box-set. Paris, France: INA-GRM. (Original work created 1970)

Ferrari, L. (2009b). Presque Rien N°4, La Remontée du Village. In *Luc Ferrari –L'Œuvre Électronique*, 10 CD Box-set. Paris, France: INA-GRM. (Original work created 1990-1998)

Freud, S. (1950). Project for a scientific psychology. In *The standard edition of the complete psychological works of Sigmund Freud*, Vol. 1 (J. Strachey, Trans.; pp. 281–391). London: Hogarth Press and the Institute of Psycho-Analysis. (Original work published 1895)

Freud, S. (1957). Instincts and their vicissitudes. In *The standard edition of the complete psychological works of Sigmund Freud*, Vol. 14 (J. Strachey, Trans.; pp. 109–140). London: Hogarth Press and the Institute of Psycho-Analysis. (Original work published 1915)

Fröhlich, A. (2000). *La stimulation basale—Le Concept* [Basal stimulation—The concept]. Bern, Switzerland: SZH /SPC.

Gandhi, M. S., Sesek, R., Tuckett, R., & Bamberg, S. J. (2011). Progress in vibrotactile threshold evaluation techniques: A review. *Journal of Hand Therapy, 24*(3), 240–256. https://doi.org/10.1016/j.jht.2011.01.001

Genevois, H., & de Vivo, R. (Eds.). (1999). *Les nouveaux gestes de la musique* [The new gestures of music]. Marseille, France: Parenthèses.

Giordano M., & Wanderley, M. (2015, July). Follow the tactile metronome: Vibrotactile stimulation for tempo synchronisation in music performance. Retrieved April 28, 2020, from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.702.3381&rep=rep1&type=pdf

Golse, B. (2010). *Les destins du développement chez l'enfant* [The vicissitudes of child's development]. Toulouse, France: Erès.

Hopkins, C., Maté-Cid, S., Fulford, R., Seiffert, G., & Ginsborg, J. (2016). Vibrotactile presentation of musical notes to the glabrous skin for adults with normal hearing or a hearing impairment: Thresholds, dynamic range and high-frequency perception. *PLoS One, 11*(5), e0155807. https://doi.org/10.1371/journal.pone.0155807

Jacobsen S. L., Pedersen I. N., & Bonde, L. O. (Eds.). (2019). *A comprehensive guide to music therapy*. London, UK: Jessica Kingsley Publishers.

Jeammet, P., & Corcos, M. (2010). *Évolution des problématiques à l'adolescence: L'émergence de la dépendance et ses aménagements* [Evolution of issues in adolescence: The emergence of addiction and its arrangements]. Rueil-Malmaison, France: Doin.

Kaplan, L. (1995). *Adolescence: The farewell to childhood.* New York, NY, USA: Pocket Books. (Original work published 1984)

Kestemberg, É., Kestemberg J., & Decobert, S. (2005). *La faim et le corps. Une étude psychanalytique de l'anorexie mentale* [Hunger and the body: A psychoanalytic study of anorexia nervosa]. Paris, France: PUF.

Lauras-Petit, A. (2009). Rythmes et contenants psychiques [Rythms and psychic containers]. *Champ psychosomatique, 54*(2), 105–126.

Lindenmeyer, C. (2019). *Les embarras du féminin* [The embarrassments of the feminine]. Paris, France: PUF.

Logak, Z., & Barbosa, I. (2016). Prise en charge psychanalytique des troubles des conduites alimentaires: au risque de se rencontrer soi-même [Psychoanalytic treatment of eating disorders at the risk of meeting oneself]. In S. Criquillion (Ed.), *Anorexie, boulimie: Nouveaux concepts, nouvelles approches* [Collection cahiers de Sainte-Anne]. Paris, France: Lavoisier.

McDougall, J. (1989). *Theatres of the body: A psychoanalytic approach to psychosomatic illness*. New York, NY, USA: W. W. Norton.

Merchel, S., & Altinsoy, E. (2013). Auditory-tactile music perception. In *Proceedings of Meetings on Acoustics, 19*(1), 1–7. https://doi.org/10.1121/1.4799137

Merleau-Ponty, M. (2002). *Phenomenology of perception* (C. Smith, Trans.). London, UK: Routledge. (Original work published 1945)

Merleau-Ponty, M. (1968). *The visible and the invisible* (A. Lingis, Trans.). Evanston, IL, USA: Northwestern University Press. (Original work published 1964)

Pankow, G. (1977). *Structure familiale et psychose* [Family structure and psychosis]. Paris, France: Flammarion.

Parmegiani, B. (2008). De Natura Sonorum (deuxième série) : Matières induites. In *Bernard Parmegiani – L'Œuvre Musicale*, 12 CD Box-set. Paris, France: INA-GRM. (Original work created 1974)

Patiño-Lakatos, G. (2015, October). Toucher et voir le son: Approches sensorielles du son dans une pédagogie multimodale [Touching and seeing the sound: Sensory approaches to sound in a multimodal pedagogy]. Expérience à l'Institut National de Jeunes Sourds de Paris (2012–2013): Rapport de recherche pour l'Agence Nationale de la Recherche. Retrieved November 25, 2019 from http://inshea.fr/fr/content/p%C3%A9dagogie-artistique-num%C3%A9rique-accessible-et-multimodale

Patiño-Lakatos, G., Navarret, B., & Genevois, H. (2019). Paradigmes et expériences pour une sémiotisation des sensations vibrotactiles [Paradigms and experiences for a semiotization of vibrotactile sensations]. *Alter, Revue européenne de recherche sur le handicap, 13,* 155–167.

Patiño-Lakatos, G., Navarret B., & Genevois H. (2020). De la sensation à la sémiotique vibrotactile. Médiations pour l'expérience musicale [From sensation to vibrotactile semiotics. Mediations for the musical experience]. *Hybrid, 6*, unpaginated. Available at https://hybrid.univ-paris8.fr/lodel/index.php?id=1286&file=1

Redolfi, M. (1988). Pacific tubular waves: Inner tube. In *Desert track*s [CD]. Paris, France: INA-GRM. (Original work created 1979)

Rouget, G. (1990). Double xylophone sur fosse (recording; unknown performers). In *Le Chant du monde*. Paris, France: CNRS-Musée de l'Homme. https://archives.crem-cnrs.fr/archives/items/CNRSMH_E_1990_014_001_001_32/

Rüütel, E., Ratnik, M., Tamm, E., & Zilensk H. (2004). The experience of vibroacoustic therapy in the therapeutic intervention of adolescent girls. *Nordic Journal of Music Therapy*, *13*(1), 33–46.

Sacks, O. (2007). *Musicophilia: Tales of music and the brain*. New York, NY, USA: Knopf.

Schaeffer, P. (2010a). Cinq études de bruits: Étude aux chemins de fer. In *Pierre Schaeffer – L'Œuvre Musicale,* 3 CD Box-set. Paris, France: INA-GRM. (Original work created 1948)

Schaeffer, P. (2010b). Cinq études de bruits: Étude aux Tourniquets. In *Pierre Schaeffer – L'Œuvre Musicale*, 3 CD Box-set. Paris, France: INA-GRM. (Original work created 1948)

Schaeffer, P. (2010c). Bilude. In *Pierre Schaeffer – L'Œuvre Musicale*, 3 CD Box-set. Paris, France: INA-GRM. (Original work created 1979)

Sentana, W. (performer) (1991). Tabuh Endung Geriya. In *Jegog, the bamboo gamelan of Bali*. Walter Quintus (recording). Kreuzau, Germany: Creative Music Productions Records. Catalog number: CMP CD 3011.

Skille, O. (1989). VibroAcoustic therapy. *Music Therapy, 8*(1), 61–67.

Winnicott, D. W. (1971a). *Playing and reality.* London, UK: Tavistock.

Winnicott, D. W. (1971b). *Therapeutic consultations in child psychiatry*. London, UK: Chatto & Windus.

Wollman, I., Fritz, C., & Frelat, J. (2015). On the characterization of vibrotactile feedback in violinists' left hand: A case study. *Acta Acustica United with Acustica, 101,* 360–368. https://doi.org/10.3813/AAA.918832

Authors' Note

This project was coordinated by Cristina Lindenmeyer and funded by the Fondation de France as part of the call for projects *Soutenir les jeunes en souffrance psychique* [Supporting young people in mental distress], 2019. The Ethics Committee of the University of Paris, France, approved this project: CER-PD: 2019-39-BARBOSA.

This original article was written in French and translated to English for publication in *Human Technology* by Kristina Valendinova.

All correspondence should be addressed to
Gabriela Patiño-Lakatos
Sorbonne Université
Institut Jean Le Rond d'Alembert
UMR 7190
Cases 161 et 162
4, place Jussieu
75252 Paris CEDEX 05
gabriela.patino-lakatos@sorbonne-universite.fr

Human Technology
ISSN 1795-6889
www.humantechnology.jyu.fi

Human Technology

www.humantechnology.jyu.fi

humantechnology@jyu.fi

Submissions: humantechnologypublishing.jyu.fi