

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.*

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

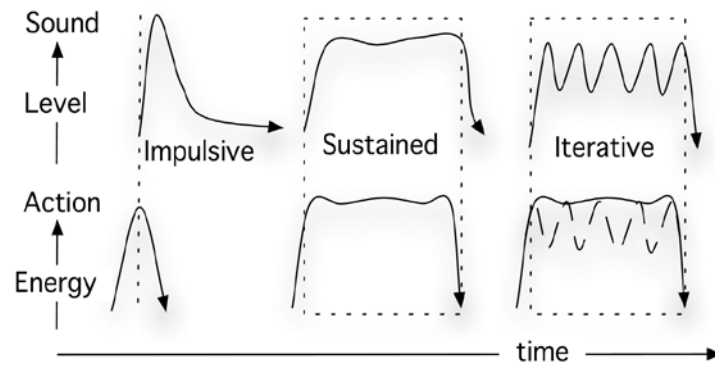


Figure 1. Illustration of the three, basic action–sound types: impulsive, sustained, and iterative (Jensenius, 2007; Used with permission).

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



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.

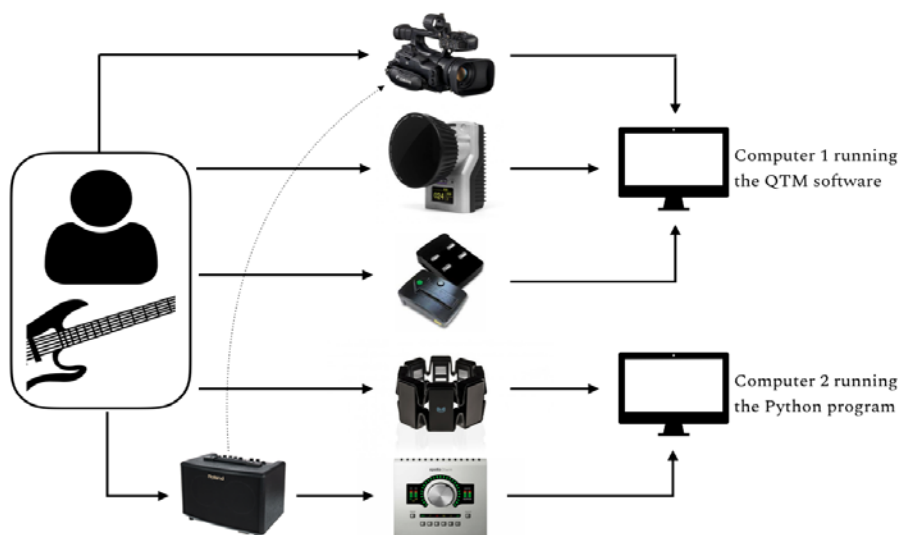


Figure 3. A simplified signal flow diagram of the experimental setup. Representative pictures of the equipment used, from top to bottom: Canon video camera, Qualisys Oqus infrared camera, Delsys Trigno electrodes, Myo armband, and Roland guitar amplifier, and Universal Audio Apollo Twin sound card.

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,

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

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,

$$EMG_m = \text{mean}EMG_m + PC1 \times EMG1_m + \dots + PCn \times EMGn_m$$

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, $t(246) = 16.98, p < .001$

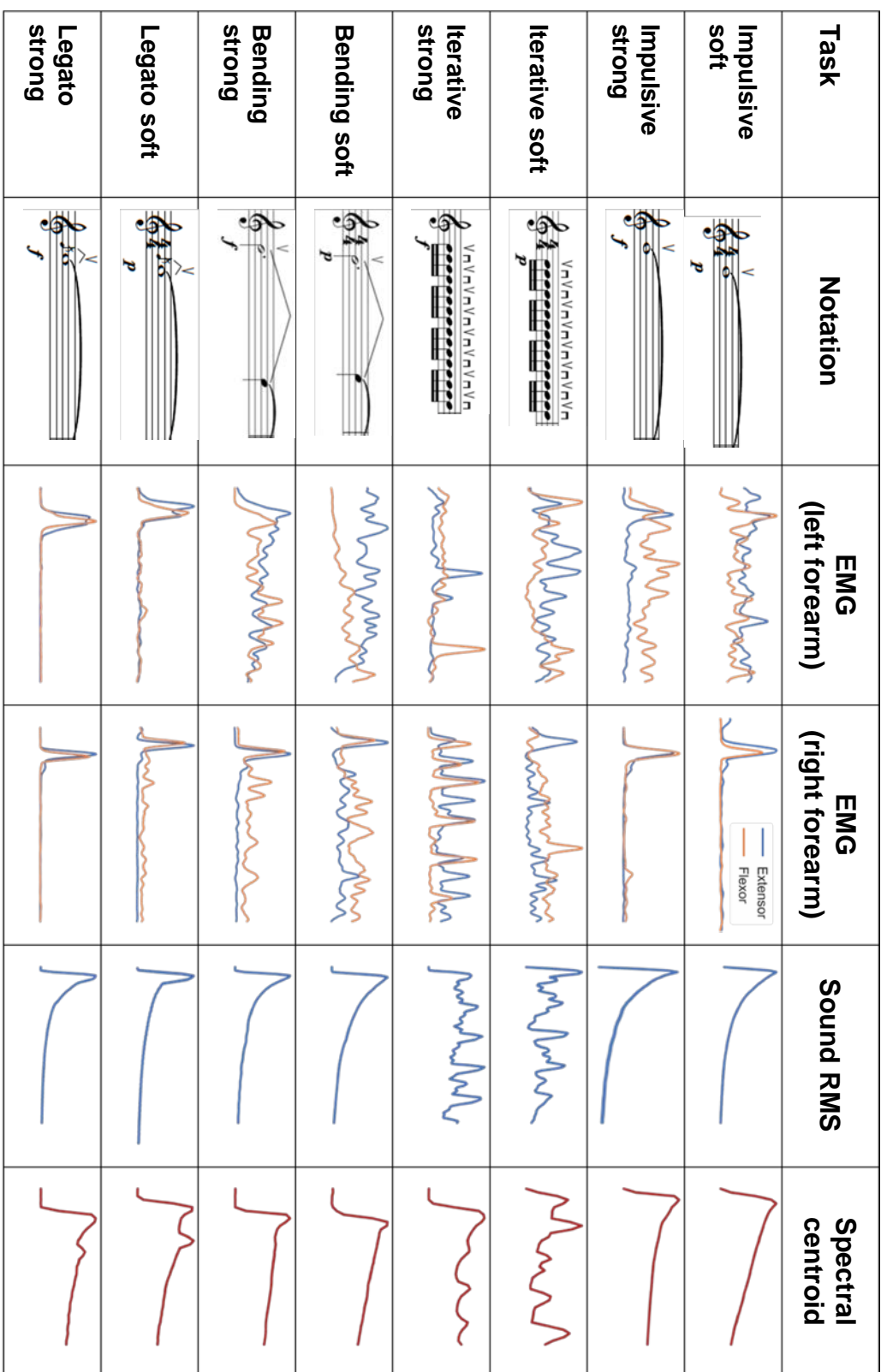


Figure 4. An overview of how notated music transforms into an audio waveform when playing the electric guitar. Trends of signals were extracted using Singular Spectrum Analysis (SSA) with a window length $L = 10$.

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

LCC	r	Impulsive		Iterative		Bending		Legato	
		soft	strong	soft	strong	soft	strong	soft	strong
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
$-r$	Flexor (left)	0.55	0.55	0.65	0.65	0.48	0.63	0.51	0.48
	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
μ	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
	Extensor (right)	0.17	0.24	0.28	0.33	0.26	0.28	0.00	0.09
σ	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)

SCC		Impulsive		Iterative		Bending		Bending		Legato	
		soft	strong	soft	strong	soft	strong	soft	strong		
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.

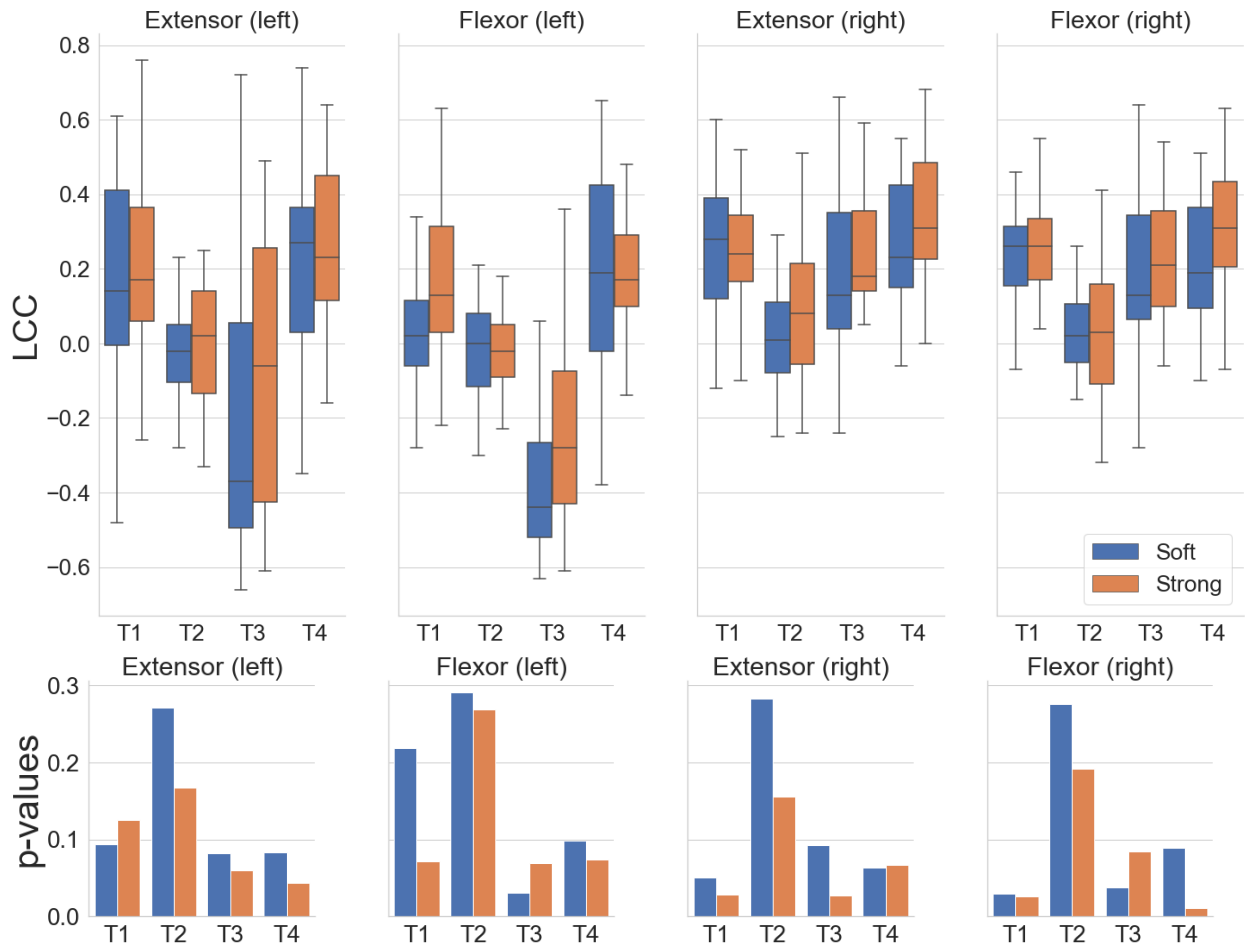


Figure 5. Pearson’s product–moment correlations between EMG and Sound RMS envelopes. $LCC > 0$ denotes a positive correlation while $LCC < 0$ refers to the negative. The box plots show the interquartile ranges of correlation distribution per task, separately for soft and strong dynamics. The bar plots 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	$M = 0.05, SD = 0.34$	$M = 0.20, SD = 0.24$	$t(495) = 9.04, p < .001$

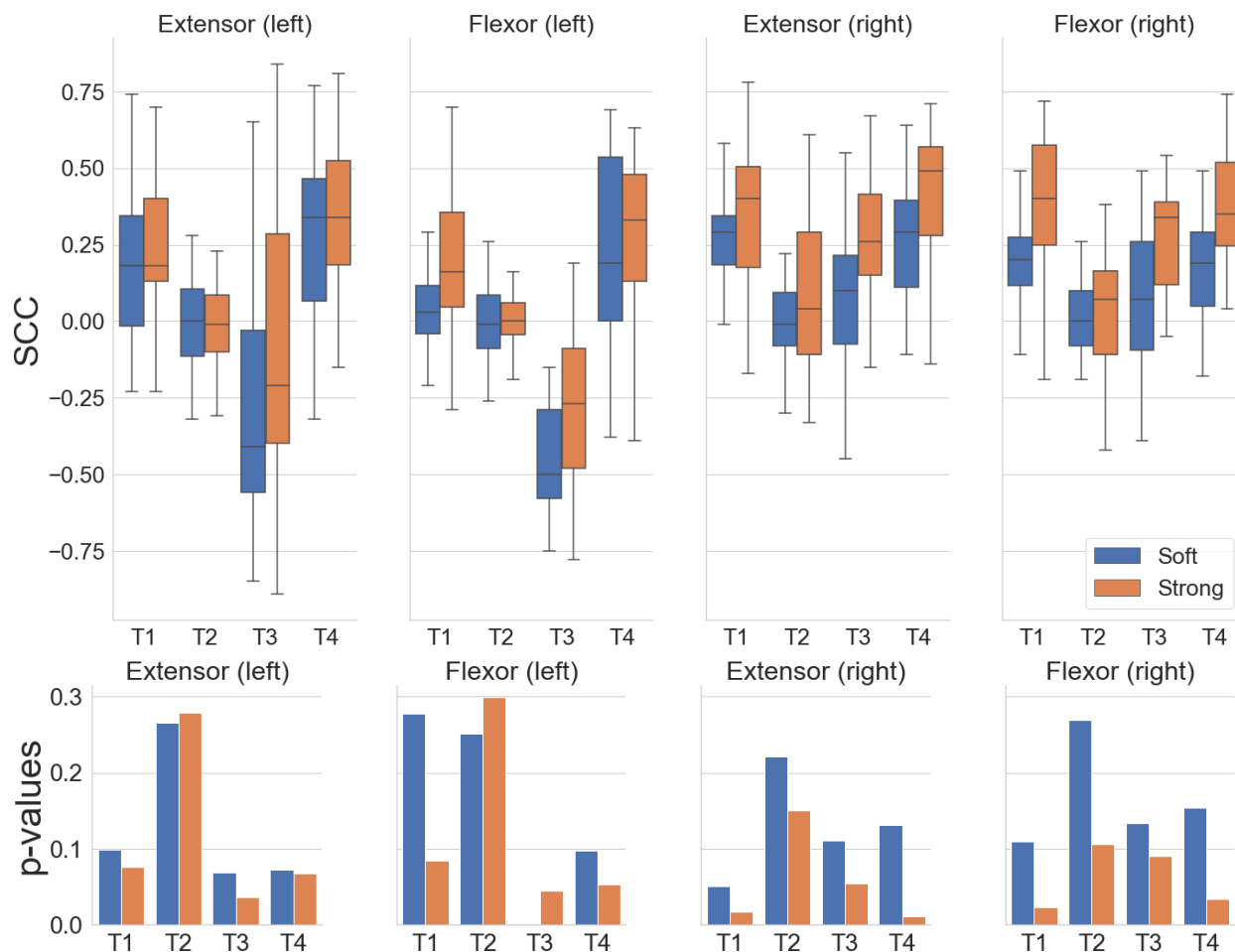


Figure 6. Spearman’s rank correlations between EMG and Sound RMS amplitude envelopes. $SCC > 0$ denotes a covariance in the same direction while $SCC < 0$ refers to the opposite direction. The box plots show the interquartile ranges of correlation distribution per task, separately for soft and strong 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.

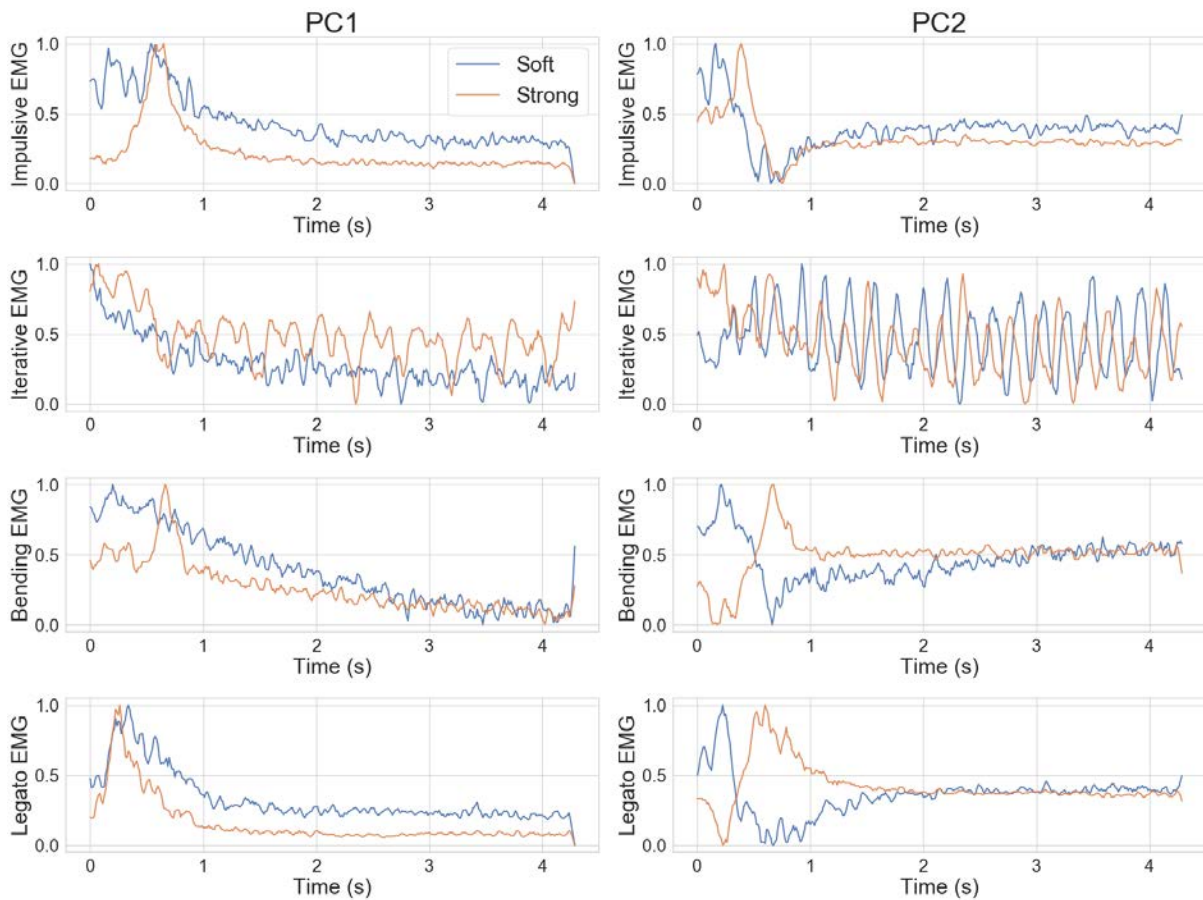


Figure 7. Two principal components (PC1 and PC2) of the combined left and right forearm EMG data of all participants rescaled to $(0, \dots, 1)$ (See the text for more information about the PCA analysis).

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.

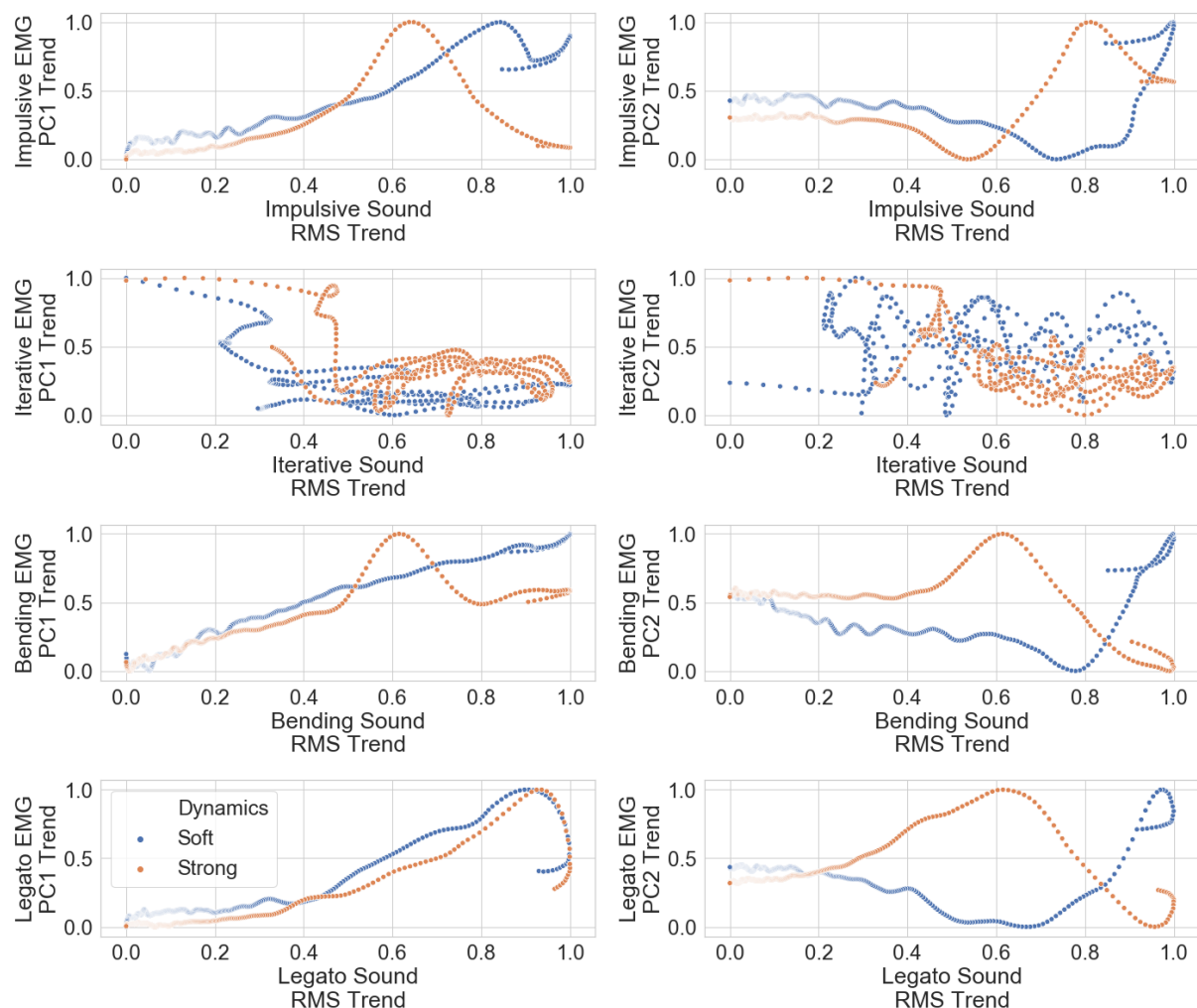


Figure 8. Decomposed principal components (PC1 and PC2) against resultant Sound RMS of all participants (SSA window length $L = 10$). The plots show to what extent the EMG and resultant sound RMS envelopes have a linear relationship at every time step.

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,

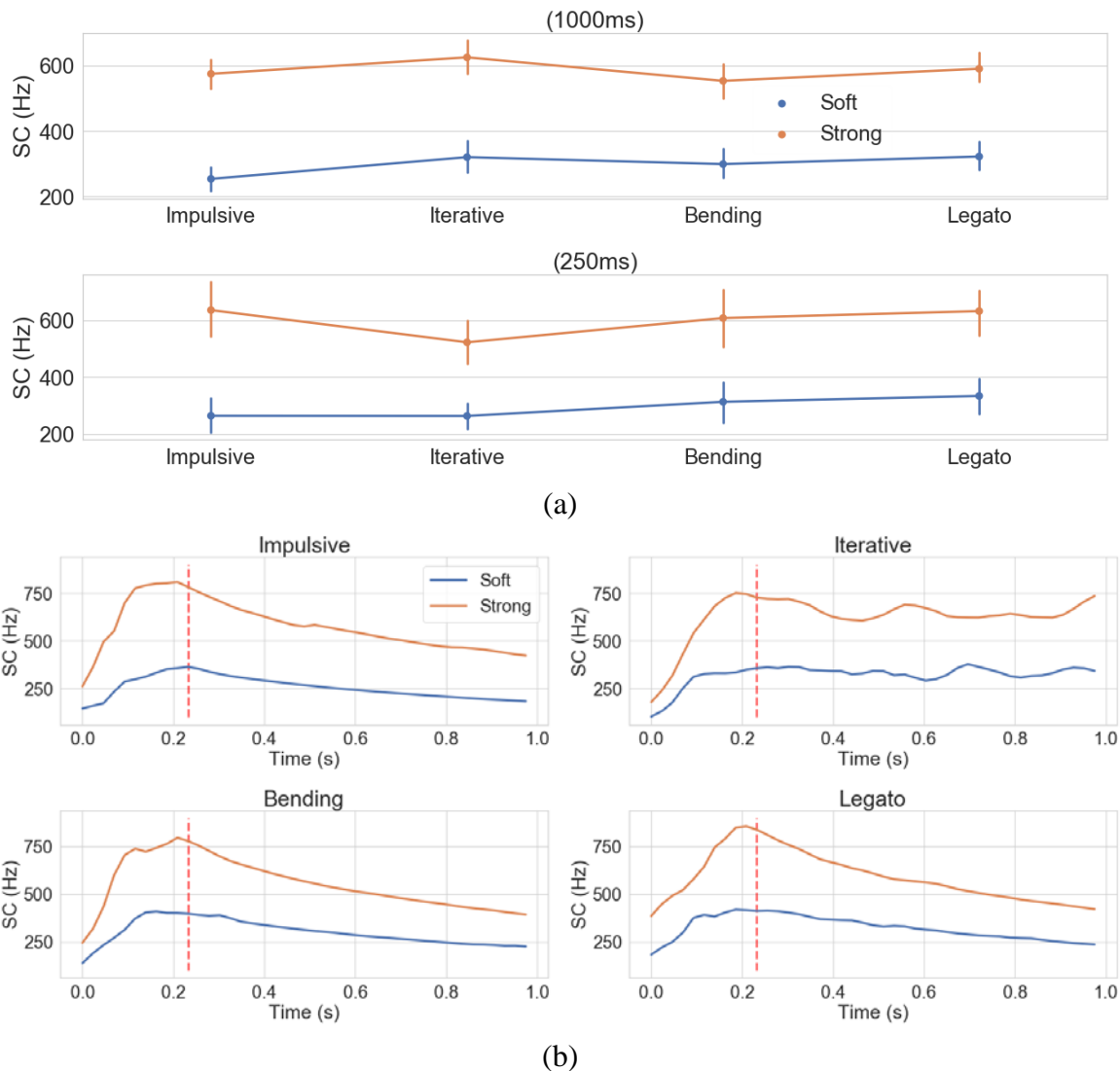


Figure 9. Spectral centroid (SC) of the resultant sound (a) SC distribution between soft and strong dynamics in chunks of 1000 ms and 250 ms duration. (b) SC envelopes averaged across all participants. The red vertical lines on the left sides of the plots show the cut point of 250 ms. Note that the segments 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.

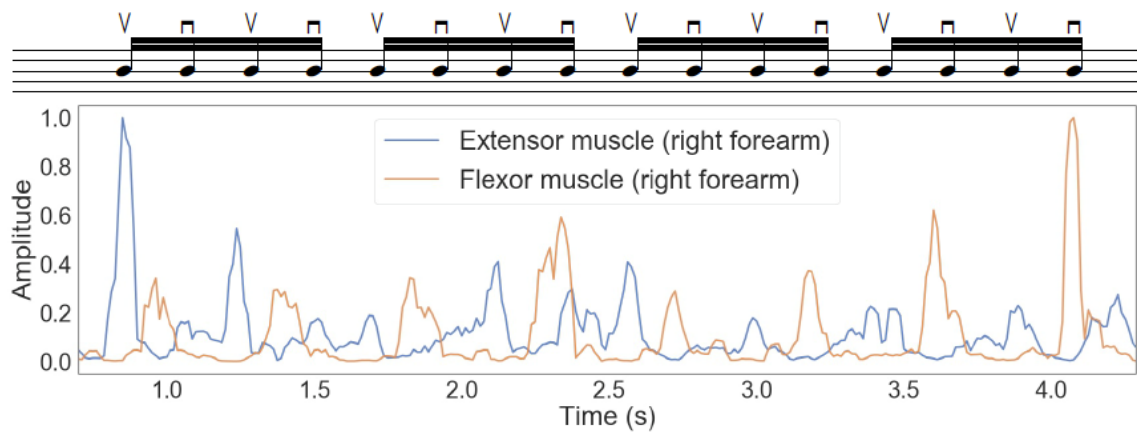


Figure 10. EMG amplitude of the excitation motion during iterative task demonstrating distinct activation of extensor and flexor muscles for down and up strokes, respectively, during a series of 16th notes.

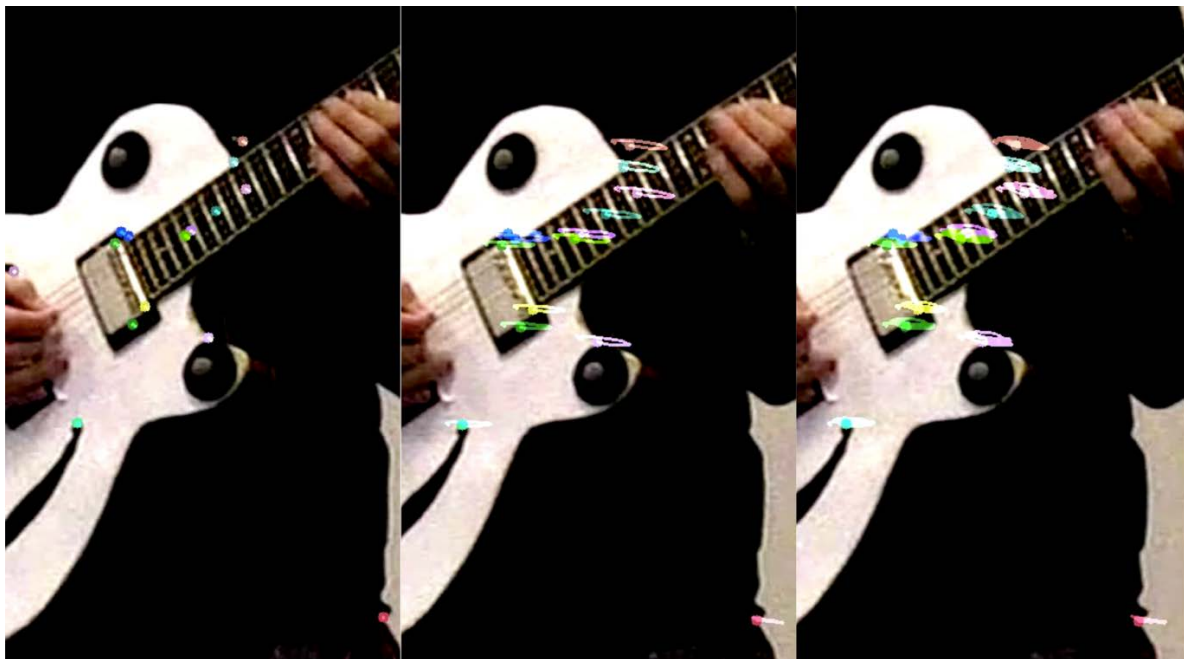


Figure 11. The sparse optical flow shows the trajectory of multiple points on a close-up video segment while a participant is performing an impulsive task. Three subsequent screenshots demonstrate the ancillary motion reflected on the guitar over the period of 1 bar (~3.43 s). The multicolored points 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-Yáñez, 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

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

where x_{RMS} are the recorded values, \hat{x}_{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

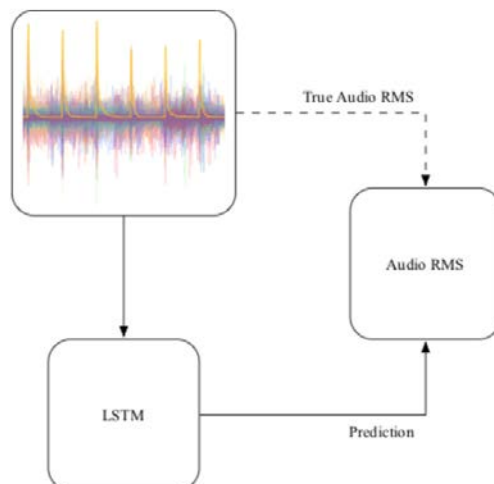


Figure 12. Sketch of the training model: A 16-channel Raw EMG as the source and sound RMS as the target data are passed into an LSTM cell, which then outputs a prediction.

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

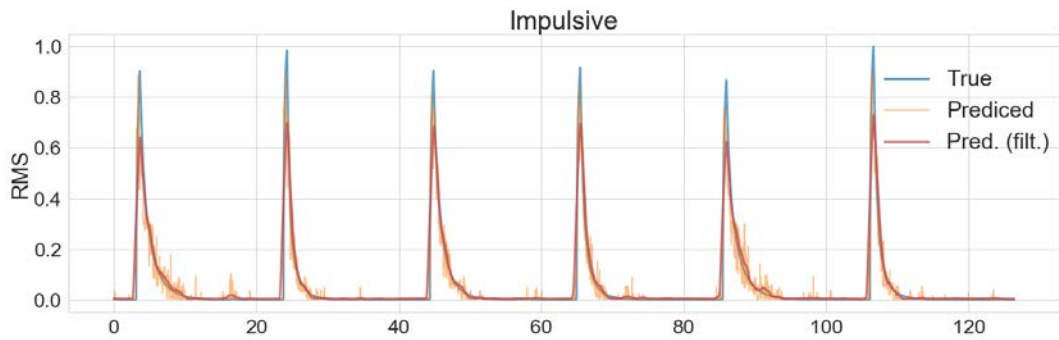
All model configurations were generally capable of predicting the sound RMS (see Figure 13). The model with two hidden layers and 64 units had the best results, which can be seen in the figures of recorded versus predicted RMS of the impulsive (Figure 13a) and iterative tasks (Figure 13b). For the latter,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

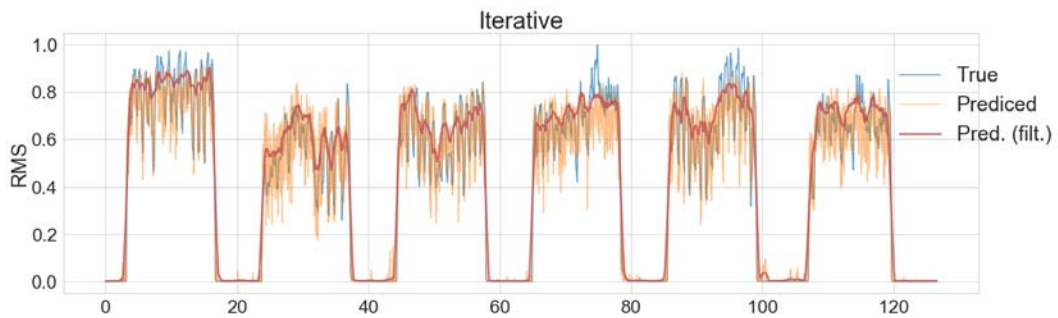
The prediction of the bending task brought an interesting result (Figure 13c). Normal guitar performance does not afford sustained excitation action, although it can be accomplished with a bow on the strings, as Led Zeppelin’s guitarist, Jimmy Page, popularized in the late 1960s. However, apart from 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.

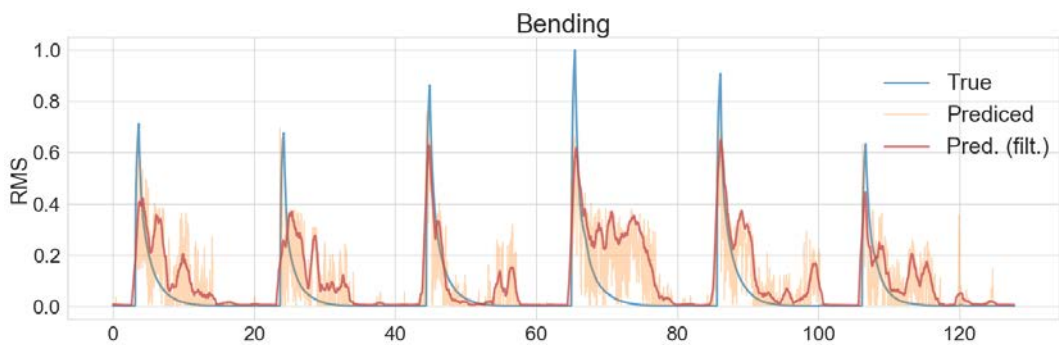
We also tested various model sizes using Euclidean distance measure (EDM), which is a common method for measuring the distance between objects. EDM is calculated as the root of square



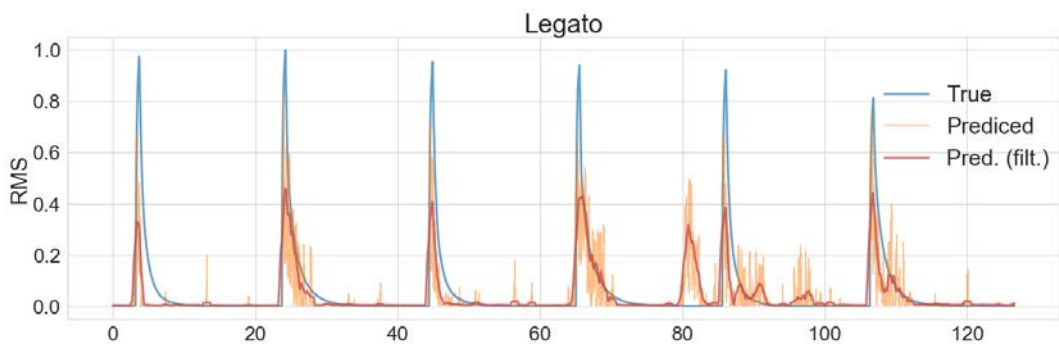
(a) The RMS of the recorded sound and the model prediction for the impulsive task.



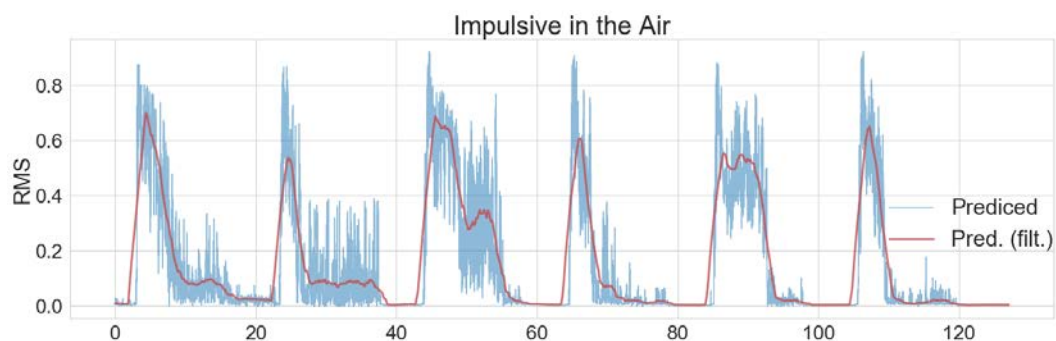
(b) The RMS of the recorded sound and the model prediction for the iterative task.



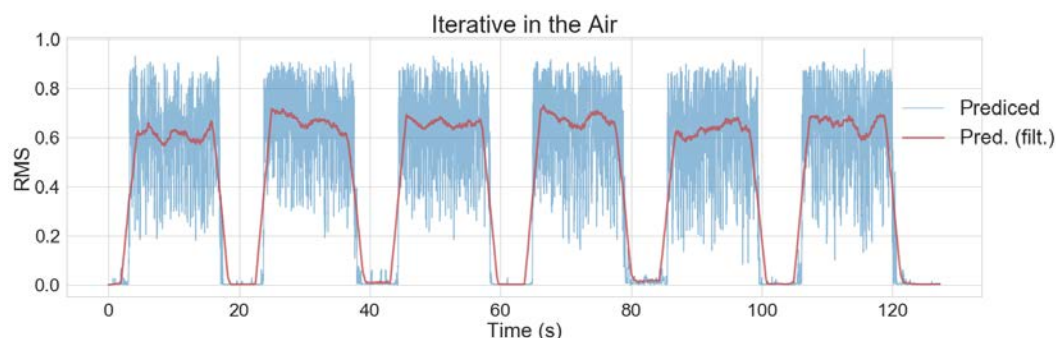
(c) RMS of the recorded sound and the model prediction for the bending task.



(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.

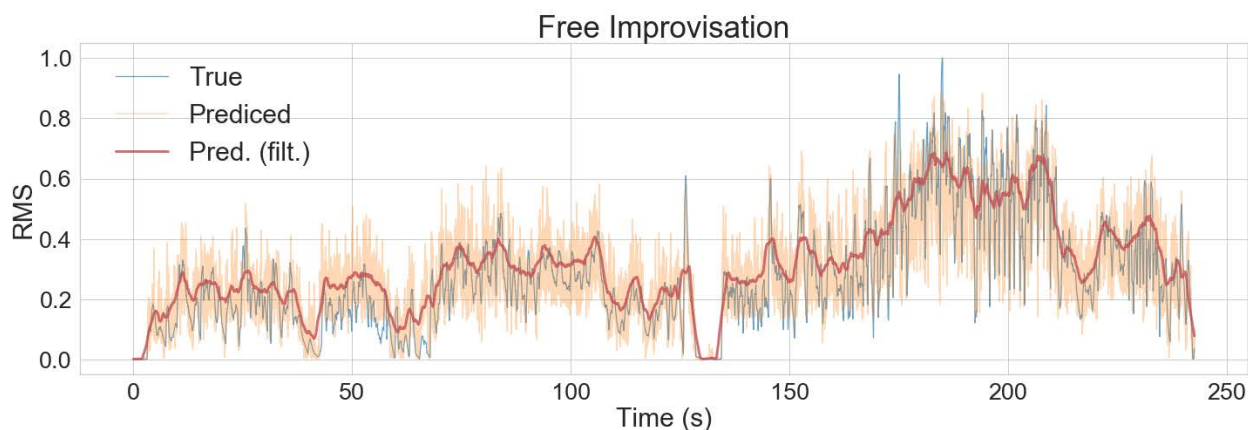


Figure 14. The RMS of the recorded sound and the model prediction of a free improvisation task. Predicted curves are filtered to reflect the general shape and facilitate the visual inspection.

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 $\sqrt{(p_1 - s_1)^2 + (p_2 - s_2)^2 \dots (p_n - s_n)^2}$. 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.

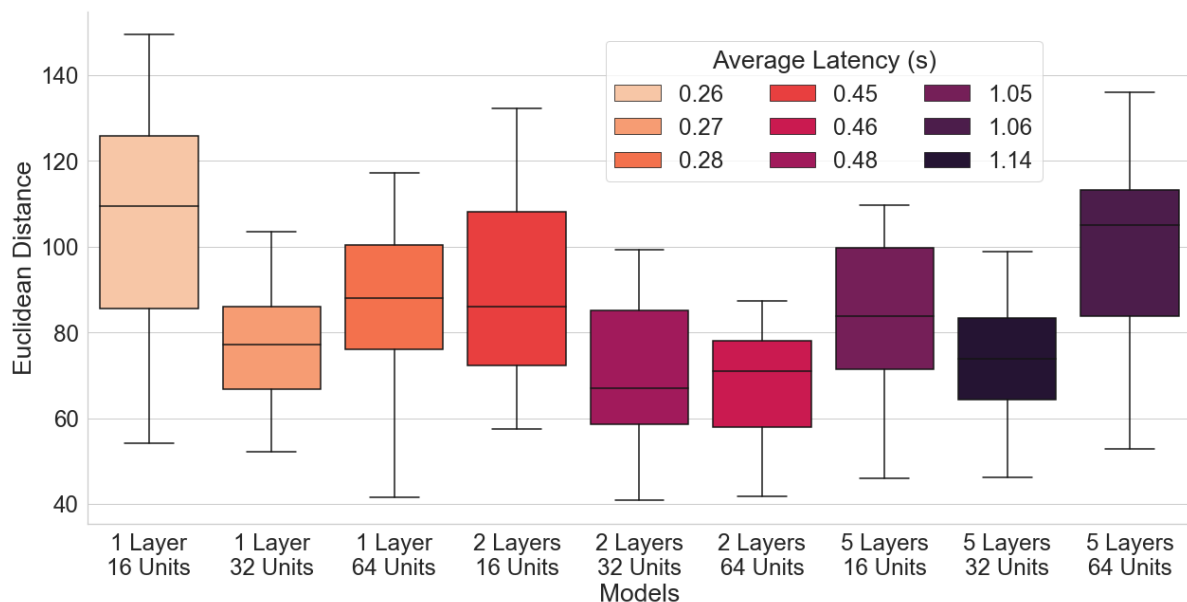


Figure 15. Euclidean distances between true RMS envelope of the free improvisation task and its corresponding prediction of RMS envelope based on nine model configurations. The boxes display the interquartile ranges while the central lines show the median. The whiskers show the minimum and maximum values 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-Yáñez, 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](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., Nymoén 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>
- Nymoën, 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/SchWoITarICMPC8.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). AI-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 the 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