

A Statistical Approach to Analyzing Sound Tracings

Kristian Nymoen¹, Jim Torresen¹,
Rolf Inge Godøy², and Alexander Refsum Jensenius²

¹ University of Oslo, Department of Informatics, Oslo, Norway
{krisny,jimtoer}@ifi.uio.no,

² University of Oslo, Department of Musicology, Oslo, Norway
{r.i.godoy,a.r.jensenius}@imv.uio.no

Abstract. This paper presents an experiment on *sound tracing*, meaning an experiment on how people relate motion to sound. 38 participants were presented with 18 short sounds, and instructed to move their hands in the air while acting as though the sound was created by their hand motion. The hand motion of the participants was recorded, and has been analyzed using statistical tests, comparing results between different sounds, between different subjects, and between different sound classes. We have identified several relationships between sound and motion which are present in the majority of the subjects. A clear distinction was found in onset acceleration for motion to sounds with an impulsive dynamic envelope compared to non-impulsive sounds. Furthermore, vertical movement has been shown to be related to sound frequency, both in terms of spectral centroid and pitch. Moreover, a significantly higher amount of overall acceleration was observed for non-pitched sounds as compared to pitched sounds.

1 Introduction

Research on music and motion is an interdisciplinary area with links to a number of other fields of research. In addition to traditional research on music and on kinematics, this area relates to neuropsychology, cognition, linguistics, robotics, computer science, and more [20]. To be able to grasp the complexity of the relationship between music and motion, we need knowledge about how different factors influence the motion and how the musical sound is perceived and processed in the brain. In addition, a certain understanding of experimental and mathematical methods is necessary for analyzing this relationship.

In several fields dealing with sound and motion, it is essential to identify how features of sound relate to events in other modalities. This includes disciplines like auditory display, interaction design, and development of multi-modal interfaces [23, 27]. Furthermore, it has been suggested that this aspect could be utilized in music information retrieval research, for instance by querying sound data bases with body motion [3, 13].

In this paper we present an experiment where motion capture technology was used to measure people’s motions to sound. This data has been analyzed in view of perceptual correspondences between lower-level features of sound and motion. The participant’s motion was evaluated statistically to find cross-modal relationships that have significance within our data set. There is a need for systematic experiments on sound-action relationships to build a larger corpus of examples on the links between perception of music and motion. The findings in this paper serve as a contribution to this corpus.

We shall begin by introducing the background and motivation for this particular research, including elements from music cognition, accounts of previous experiments on music and motion, as well as our own reflections on the implications of these works. In Section 3 we introduce our experimental setup, followed by a description of the recorded data set, with necessary preprocessing and feature extraction in Section 4. Section 5 presents analysis of the data set, and the results are discussed in Section 6. Conclusions are provided in Section 7.

2 Background

Presently, we have observed an increased popularity of a so-called theory of *embodied cognition*, meaning that bodily sensorimotor processing is understood as an important factor in our cognitive system [28]. Leman [19] put this theory into a musical context in his introduction of *embodied music cognition*. This theory describes how people who interact with music try to understand musical intentions and forms by imitation through corporeal articulations like body motion (e.g. tapping the beat, attuning to a melody or harmony, etc.) and empathy (e.g. attuning to certain feelings or a mood conveyed by the music).

Godøy [9] posits that our understanding of discrete events in music can be explained through *gestural-sonic objects*. These objects are mental constructs that combine the auditory input with gestural parameters, enabling an understanding of the sonic object through its causality (e.g. a perceived sound producing action). The idea of a gestural-sonic object as a discrete perceptual unit, or *chunk*, is based upon Pierre Schaeffer’s *sonic object* [26], on Miller’s theory of recoding complex sensory information into perceptual chunks [22], and also on the phenomenological understanding of perception as a sequence of now-points introduced by Husserl [14]. According to Godøy, these objects take form at the *meso level* of a musical timescale [10]. In contrast, the *macro level* of a musical timescale could be a whole musical piece, and the *micro level* of the timescale takes place within the sonic object. We believe that *action-sound relationships* [15] are found at all timescale levels, which coexist when a person is involved in a musical experience. Certain musical features like rhythmic complexity or emotional content require a larger timescale perspective than for instance musical features like pitch and timbre which operate in the millisecond range [7].

In a similar manner to the listening experiments Schaeffer performed on sonic objects, we can learn more about gestural-sonic objects by studying lower-level features of sound-related motion. In other words, one can look at the meso

level object from a micro-level perspective. Godøy et al. explored gestural-sonic objects in an experiment they referred to as *sound tracing* [12]. Nine subjects were given the task of making gestures they believed corresponded well with the sounds they heard, by using a pen on a digital tablet. By qualitative comparisons of the sound tracings, the authors found a fair amount of consistency between subjects, and argued that this type of experiment should be done in a larger scale, and include more complex sound objects, to learn more about sound-gesture relationships. The same material was later also analyzed quantitatively by extracting features and classifying the sound tracings using a support vector machine classifier [8]. We shall inherit the term *sound tracing* in the experiment presented in this paper. To be more precise, a *sound tracing* in this sense describes a bodily gesture that has been performed in free air to imitate the perceptual features of a sound object.

Other researchers have also studied how lower-level features of sound objects are related to motion or motion descriptors. Merer et al. [21] asked people to put their own motion-labels on sounds with different sound features. This way they determined which sound parameters were most pertinent in describing motion-labels such as “rotate” and “pass by”. Eitan et al. found that for sounds with changing pitch, people imagined the movement of an animated character to follow the pitch up or down, however the authors also argued that changing pitch is related to other dimensions than simply vertical position [5, 6]. This corresponds well with previous research on metaphors and auditory display where increasing pitch has been related to an increase in other dimensions in other modalities, such as temperature [27]. The relationship between pitch and verticality was also found by Nymoen et al. [25] in a sound tracing experiment where participants used a rod to trace the perceptual features of a selection of sounds. In an experiment on synchronization with music, Kozak et al. [17] observed differences for quantity of motion between different lower-level features of sound like pitch, spectral centroid and loudness. Caramiaux et al. [2] applied Canonical Correlation Analysis to a set of sound and motion features derived from sound tracings.³ This method gave promising results in identifying correlations between features of sound and of motion, and was later applied by Nymoen et al. [24].

The present paper is intended to follow up on the sound tracing experiments presented above. The main idea in this research was to study sound tracings from a more systematic perspective, in particular by using systematically varied sound parameters. This entailed using a number of short sounds, some where only a single sound parameter was varied, and some where multiple sound parameters were varied. In this manner, it was possible to understand the influence of different sound parameters on the sound tracings, which provided knowledge about how these features are reflected in other modalities. Our analytical approach operates at the meso and micro levels of the musical timescale, combining features that describe chunks of sound and motion with continuously varying sound and motion features.

³ Caramiaux et al. do not refer to them as *sound tracings*, but following the definition presented above, their experiment falls into this category.

3 Experiment

A sound-tracing experiment was designed to be able to systematically distinguish between how people’s motion changes and varies in relation to changes in sound features. The data presented in this paper was recorded in Fall 2010.

3.1 Aim

The aim of the experiment was to identify how lower-level features of motion corresponded with features of sound across different participants. By using systematically designed sounds, we can isolate a single sound feature and compare how it relates to motion by itself, or in combination with other sound features.

3.2 Participants

38 people (29 male and 9 female) volunteered to participate in the experiment. They were recruited through mailing lists for students and staff at the University of Oslo and by posting an advertisement on the project website. After participating in the experiment, the participants filled out a questionnaire concerning their level of musical training. 12 people rated their level of musical training as extensive, 11 as medium, and 15 as having little or no musical training. The level of musical training was used in the analysis process to distinguish between experts and non-experts (cf. Section 5). They were also given the opportunity to comment on the experiment. The subjects were not asked for their age, but we estimate the age distribution to be 20–60 years, with most participants aged somewhere between 25 and 35.

3.3 Task

The participants were presented with a selection of short sounds (the sounds will be discussed in Section 3.5). They were instructed to imagine that they could create sound by moving their hands in the air, and move along with the sounds as if their hand motion created the sound. First, each participant was given a pre-listening of all 18 sounds. Following this, the sounds were played one by one in random order. Each sound was played twice: the first time, the participant would only listen, and the second time the participant’s hand motion was recorded. A three second countdown was given before each sound, so the participant would know exactly when the sound began.

3.4 Motion Capture

A Qualisys optical infrared marker-based motion capture system was used to record the motion of the people that participated in the experiment. The participants grasped two handles (Figure 1), each one equipped with 5 markers, and the center position of each handle was recorded. There are several advantages to

using this technology for recording motion. The system is very accurate, with a high resolution in both time and space. In our recordings, we used a sampling frequency of 100 Hz. Using several markers on each handle made it possible to uniquely identify the left and right handle, respectively, and enabled tracking of the position and the orientation of each handle.



Fig. 1. One of the two handles that was used for recording the participant’s motion.

The main limitation we have experienced with the technology is so-called *marker-dropouts*. This happens when a marker is occluded (e.g. by the body limbs of the participant) or moved out of the calibrated capture space. Marker-dropouts caused a loss of a number of data-frames in several recordings, and it became necessary to perform so-called gap-filling. We will return to how this was done in Section 4. The marker dropouts made it necessary to disregard the orientation data from the handles, although this was initially recorded. This is because gap-filling of the orientation data was more difficult than gap-filling of the position data (interpolation even over small gaps introduces large errors).

3.5 Sounds

A total of 18 short sound objects, each 3 seconds in length, were designed in Max5 using *frequency modulation* (FM) synthesis and digital filters. The design process was to a large extent based on trial and error, to find sounds where the

envelopes of *pitch* (perceived tone height) and *spectral centroid* (here interpreted as perceived brightness) were distinct. *Envelope*, in this sense, is a generic term for a curve describing the development of a sound feature in the time domain. An example of the sound feature envelopes is given in Figure 2. The sound files are available for download at the project website.⁴

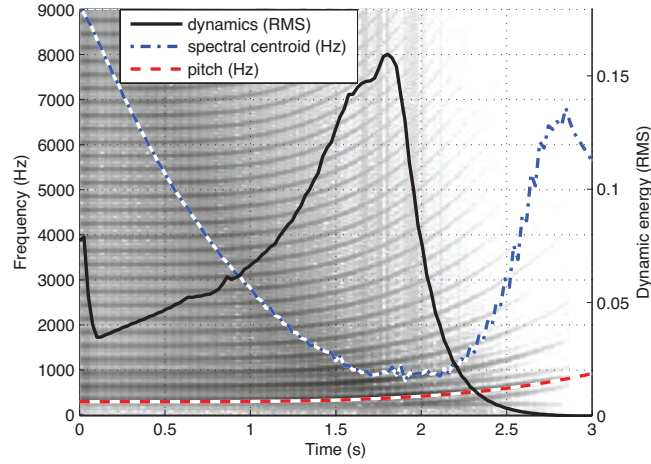


Fig. 2. Spectrogram and corresponding sound features for Sound 15. Pitch and spectral centroid (in Hz) on the left axis. The dynamic envelope scale is on the right axis.

Table 1. Simple description of the sounds used in the experiment. The columns display the pitch envelope, spectral centroid envelope and the dynamic envelope of each sound.

Sound	Pitch	Sp.Centroid	Dyn.Env.	Sound	Pitch	Sp.Centroid	Dyn.Env.
1	Rising	Falling	Bell-shape	10	Noise	Falling	Bell-shape
2	Falling	Rising	Bell-shape	11	Noise	Rising	Increasing
3	Falling	Falling	Bell-shape	12	Noise	Steady	Increasing
4	Rising	Rising	Bell-shape	13	Steady	Rising	Increasing
5	Rising	Steady	Increasing	14	Steady	Falling	Increasing
6	Falling	Steady	Increasing	15	Rising	Falling	Impulsive
7	Steady	Falling	Bell-shape	16	Steady	Steady	Impulsive
8	Steady	Rising	Bell-shape	17	Noise	Steady	Impulsive
9	Steady	Steady	Increasing	18	Noise	Falling	Impulsive

An overview of all of the sounds is presented in Table 1. In the first nine sounds, pitch and spectral centroid were manipulated by controlling the fundamental frequency of the FM sound, and the center frequency of a parametric

⁴ <http://folk.uio.no/krisny/cmmr2011>

equalizer which boosted certain parts of the sound spectrum. These sounds were generated by changing the envelopes of pitch between 300 and 1000 Hz (*rising*, *falling* and *steady*) and equalizer center frequency between 50 and 13000 Hz (*rising* and *falling* as well as filter bypass, here interpreted as *steady* spectral centroid). This allowed for an appropriate discrimination between the individual sound parameter changes taking place within the sound. Sounds 10–12 were based on noise rather than a pitched FM sound, and only the filter was adjusted for these sounds. In Sounds 13 and 14, a second parametric equalizer was added. In Sound 13, the center frequencies of the equalizers started at 1000 and 5000 Hz and approached each other towards 3000 Hz, and in Sound 14, the center frequencies started at 3000 Hz, and moved apart to 1000 and 5000 Hz.

The synthesized sounds mentioned in the previous paragraph were multiplied by a window function to control the overall dynamic envelope. Here, we wanted to keep a main focus on the pitch and spectral properties of the whole sound, while influence from onset characteristics of the sounds (changes in sound features during the first part of the sound) was not desired. Therefore, Sounds 1–14 were made with a slow attack and increasing amplitude by applying the amplitude envelope displayed in Figure 3(a).

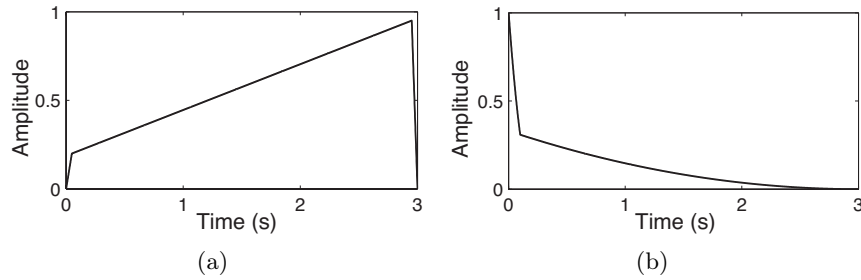


Fig. 3. The envelopes that were used for the amplitude control: (a) Non-impulsive sounds, and (b) Impulsive sounds.

The characteristics of the pitch envelope and the filter frequency also influenced the final dynamic envelope of the sounds. Some of which had a bell-shaped dynamic envelope, displayed in Figure 4(a), while others had a continuously increasing one, displayed in Figure 4(b).

We also wanted to investigate how the onset characteristics of a sound influenced the sound tracings that were performed to it. Therefore, impulsive versions of four of the sounds were made. Sounds 15–18 were versions of Sounds 1, 9, 10 and 12, the only difference was that instead of the slowly increasing dynamic envelope, we applied the impulsive envelope shown in Figure 3(b). It should be noted that the dynamic envelope of Sound 15 was different compared to the other impulsive sounds, because the varying pitch and filter frequency influenced the

dynamics. This resulted in a dynamic envelope which was a combination of the impulsive and bell-shaped envelopes, as shown in Figure 4(c).

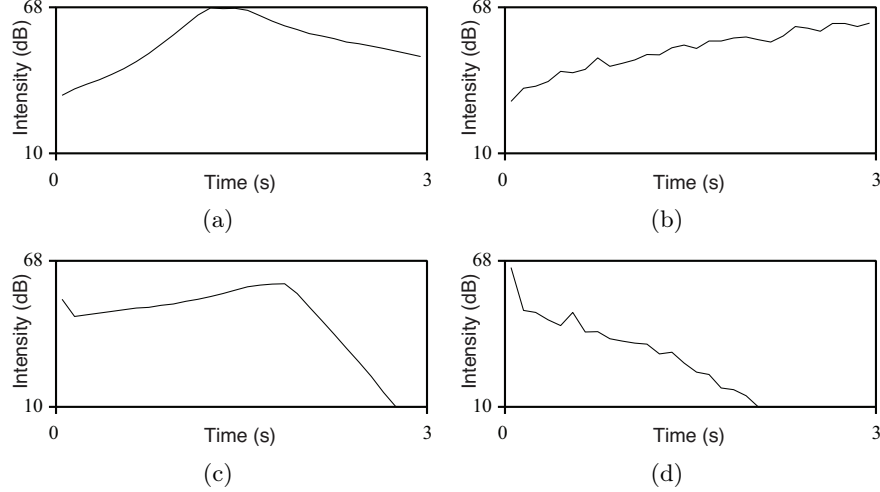


Fig. 4. The figure displays the dynamic envelopes of 4 sounds, analyzed with a perceptual model in the sound analysis software Praat. (a) Bell-shaped envelope (Sound 2), (b) Increasing envelope (Sound 9), (c) Impulsive and Bell-shaped envelope (Sound 15), and (d) Impulsive envelope (Sound 16).

4 Data Processing

In this section, we will describe the processing that was performed on the motion data to prepare it for the analysis process. The position data from the two handles was used, but it was not sufficient for our purpose to use it directly, and hence a number of data processing steps were taken. These steps included gap-filling, smoothing, feature extraction, data reduction and normalization.

4.1 Preprocessing

As mentioned in Section 3.4, some recordings contained missing data frames, and therefore gap-filling was required. We applied gap-filling on small data gaps by interpolating between the first and last missing frame using a piecewise cubic Hermite spline function with the preceding and succeeding frames as reference. A number of gaps were too large for gap-filling to be possible. In these cases, the recordings were discarded.

Certain participants had a large number of discarded recordings, which was due to poor calibration of the system in some sessions, but also because some

participants repeatedly occluded the reflective markers or moved the handles out of the capture space. If a single participant had too many discarded recordings, the consequence would be that this person only influenced a small portion of the data set, and we could risk that one participant only influenced one side of the analysis when two subsets of the dataset were compared. For this reason, we decided to discard the remaining recordings for subjects that had more than 1/3 (i.e. more than six) of their recordings removed.

The datasets from seven subjects were discarded completely, in addition to 30 recordings distributed among the other participants. In total, 156 of the 684 recordings were discarded. After the gap-filling process, a sliding mean filter of 5 samples (i.e. 50 ms) was applied to the position data in all the recordings to reduce measurement noise.

4.2 Motion Features

From the left and right handle position data we calculated a number of features that were used for analysis and comparison of the different sound tracings. Based on the position data, we calculated velocity and acceleration, as these features are related to kinetic energy and change in kinetic energy of the handles. The three axes of the position data cannot all be used directly. Movement in the vertical direction has been used directly as a motion feature, however, as will be explained shortly, the two horizontal axes are conceptually different from the vertical one, and have not been used directly.

The position data describes the position of each handle in relation to the room, or more precisely, in relation to the position of the calibration frame that was used when calibrating the motion capture system. This calibration frame determined the origin of the coordinate system and the direction of the axes. The position of the handles in relation to the calibration frame is not really relevant in light of the task that was given to the participants. The participants could not relate to the position of the calibration frame since it was removed after calibration. Furthermore, the participants were not instructed to face in any particular direction during the experiment, or precisely where to stand in the room. For this reason, we find it misleading to base our analysis directly on the horizontal position data. In contrast, the vertical position of the handles is a reference that was the same for all participants. The floor level remained constant, and was independent of where an individual stood, regardless of the direction he or she faced.

The one thing that varied between the subjects was the height range, as one participant could reach his arms to 2.2 m, while another up to 2.5 m. This was adjusted for by normalization as we will return to in Section 4.4. Based on these considerations, and on experiences regarding which features have proven to be most pertinent in previous experiments [2, 24, 25] the following data series for motion features was calculated:

- Vertical position: The distance to the floor.
- Vertical velocity: The derivative of the vertical position feature.

- Absolute velocity: Euclidean distance between successive position samples.
- Absolute acceleration: Euclidean distance between the successive derivatives of the position data.
- Distance: Euclidean distance between the hands.
- Change in distance: The derivative of the distance feature.

The features mentioned above are all data series, which we shall refer to as *serial features*. From these data series we calculated *single-value features*, meaning features that are given by a single number. These features describe a general tendency for an entire sound tracing. Examples of such features are mean vertical velocity and mean acceleration.

4.3 Data Reduction

To be able to compare the sound tracings, the data representation of each recording should be equal. In our case, this is not the case with the raw data, since some participants varied between using both hands, and only the left or right hand. Out of the 528 recordings, 454 were performed with both hands, 15 with only the left hand, and 59 with only the right hand. The participants were not specifically instructed whether to use one or two hands. In order to achieve equal data representation for all sound tracings, we had to choose between using the separate data streams from both hands in all cases, or reducing the data to fewer data streams keeping only the pertinent information from each sound tracing. Basing the analysis on data from a hand that was clearly not meant to be part of the sound tracing appeared less accurate to us, than to base the analysis on the data streams from only the active hand(s). Therefore, we calculated one serial position feature from each sound tracing, as well as one velocity feature, acceleration feature, and so forth. For the one-handed sound tracings, we used feature vectors of the active hand directly, and for the two-handed sound tracings, we calculated the average of both hands on a sample-by-sample basis. We did not change the distance feature for the single-handed sound tracings.

Admittedly, this difference between single-handed and two-handed performances presents a weakness in our experiment design, and we could have chosen different approaches to dealing with this challenge. We will continue searching for more comprehensive analysis methods which take into account this extra degree of freedom. If new methods for analysis are not found, a solution could be to instruct the participants to always use both hands.

4.4 Normalization

All feature vectors have been normalized for each subject. This means that all the calculated features were scaled to a range between 0 and 1, where the value was determined by the particular subject's maximum value for that feature. For example, if Subject 14 had a maximum vertical position of 2 meters across all of their sound tracings, all of the vertical position data series related to Subject 14 were divided by 2 meters. This type of normalization reduced individual

differences that were due to height, arm length, and so forth. This means that the data displayed in the plots in the coming section will all be scaled between 0 and 1. A similar normalization was performed on all of the single-value features.

5 Analysis and Results

The following sections present comparisons between three different aspects of the sounds. We will start in Section 5.1 by introducing our analysis method. In Section 5.2, the effect of the onset properties of the sounds are presented. In Sections 5.3 and 5.4, we present how the envelopes of pitch and spectral centroid tend to influence the sound tracings. Finally, in Section 5.5, differences between pitched and non-pitched sounds are presented.

5.1 Analysis Method

Our analysis is based on statistical comparisons between the individual data series, both sample-by-sample in serial features, and also on a higher level, comparing single-value features for the whole data series. The analyses of serial features are presented in plots where the individual data series are displayed together with the average vector of the data series. To facilitate the reading of these plots, we include a small example plot in Figure 5. This particular plot displays five data series ranging between 0 (white) and 1 (black). The vertical dashed lines show the beginning and end of the sound file, the motion capture recording began 0.5 seconds before the start of the sound file, and also lasted beyond the entire duration of the sound file. The black solid and dashed lines show the mean and standard deviations across the five data series on a sample-by-sample basis. From this figure, it is difficult to get precise readings of the values of the individual sound tracings, but the horizontal grayscale plots still give some impression of the distribution of this data set. The 0–1 scale on the y-axis is for the mean and standard deviation curves.

When certain tendencies are observed for different groups, we evaluate the statistical significance of the tendencies by applying one-tailed t -tests.⁵ Results from the tests are presented in tables, where df denotes the degrees of freedom,⁶ and t is the t -value from the t -test. p is calculated based on df and t , and denotes the probability that the two data sets are equally distributed, a p -value of less than 0.05 denotes a statistically significant difference between the groups.

Two subgroups were selected from the data set for analyzing the impact of musical training on the results in the experiment. Because of the somewhat imprecise classification of subjects' level of musical training, we chose to look at only the subjects that labeled themselves as having either no musical training or extensive musical training. This was done to ensure that there was indeed a difference in musical experience between the two groups.

⁵ A t -test is a method to estimate the probability that a difference between two data sets is due to chance. See <http://en.wikipedia.org/wiki/T-test> for details.

⁶ df is a statistical variable related to the t -test, denoting the size of the data material. It is not to be confused with e.g. 6DOF position-and-orientation data.

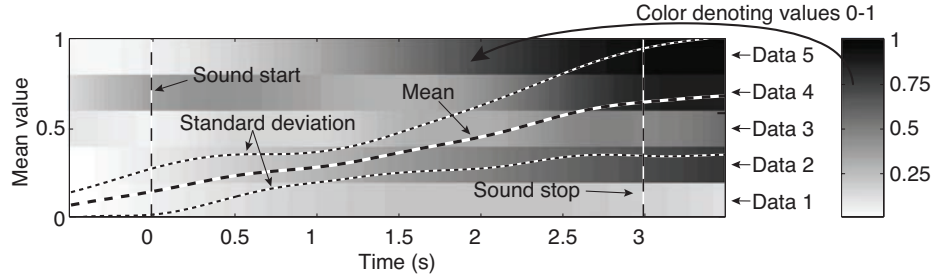


Fig. 5. The figure explains how to read the plots presented below. This is a reduced example with only 5 data series. The data series range between 0 (white) and 1 (black). Please refer to the text for explanation. A single subject is typically associated with multiple data-series, and tick marks on the right Y-axis denote different subjects. The ticks are not equidistant since some recordings were discarded (cf. Section 4.1). Furthermore, some plots display more data series per subject than others, thus the Y-axis resolution differs between plots.

5.2 Onset Acceleration for Impulsive and Non-impulsive Sounds

We evaluated how the onset characteristics of sound influence sound tracings by comparing the acceleration envelopes of impulsive sounds to non-impulsive sounds. We observed a distinct difference in acceleration envelope for sound tracings of the impulsive sounds compared to the rest of the data set, as displayed in Figure 6. To evaluate the significance of this difference, we compared the onset acceleration of the sound tracings. *Onset acceleration* is a single-value feature, which was calculated as the mean acceleration in the beginning of the sound tracing. Figure 6(b) shows that most subjects made an accentuated attack after the start of the sound file. Therefore we used a window from 0.2 seconds (20 samples) before the sound started to 0.5 seconds (50 samples) after the sound started to calculate the onset acceleration. The results of t -tests comparing the onset acceleration for impulsive and non-impulsive sounds are displayed in Table 2. The table shows that onset acceleration values for impulsive sounds are significantly higher than non-impulsive sounds, $t(526) = 13.65$, $p < 0.01$.⁷

Figure 7 displays separate acceleration curves of impulsive sounds for musical experts and non-experts. The figure shows that both groups have similar onset acceleration levels, and a t -test showed no statistical difference between the onset acceleration levels from the two groups, $t(84) = 0.55$, $p = 0.29$. However, the plots do show a difference in timing. By defining time of onset as the time of maximum acceleration within the previously defined onset interval, experts hit on average 163 ms after the start of the sound file, while non-experts hit 238 ms after the start of the sound file, a difference which was statistically significant, $t(63) = 2.51$, $p = 0.007$. This calculation was based only on Sounds 16–18, because several subjects did not perform an accentuated onset for Sound 15.

⁷ This is the American Psychological Association style for reporting statistical results. Please refer to [1] for details.

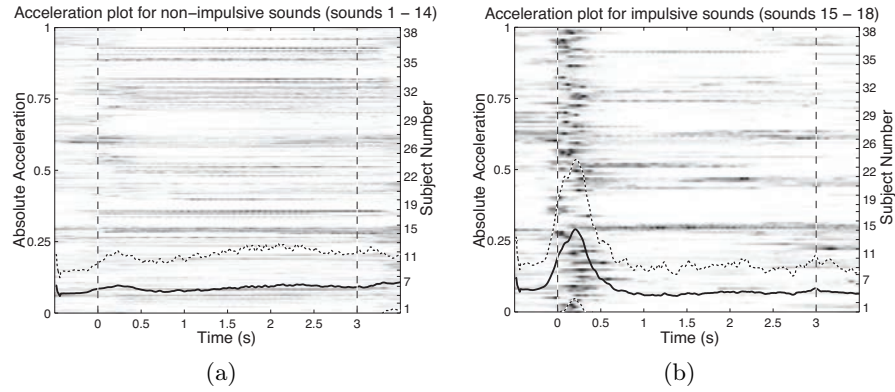


Fig. 6. Acceleration for (a) non-impulsive sounds (406 sound tracings) and (b) impulsive sounds (122 sound tracings). The black solid and dashed lines show the mean value and standard deviation across all sound tracings. Each horizontal line in the image displays the acceleration vector of a single sound tracing ranging between 0 (white) and 1 (black), normalized per subject. See Figure 5 for guidelines on how to read these plots.

Table 2. Results from t -tests comparing onset acceleration for impulsive sounds to non-impulsive sounds. There was a significant difference between the groups for both expert and non-expert subjects. See the text for explanation of the variables.

Onset acceleration, impulsive and non-impulsive sounds			
Test description	df	t	p
Impulsive vs non-impulsive, all subjects	526	13.65	< 0.01
Impulsive vs non-impulsive, expert subjects	182	8.65	< 0.01
Impulsive vs non-impulsive, non-expert subjects	183	7.86	< 0.01
Onset acceleration level, experts vs non-experts	84	0.55	0.29
Onset time, expert vs non-expert subjects	63	2.51	< 0.01

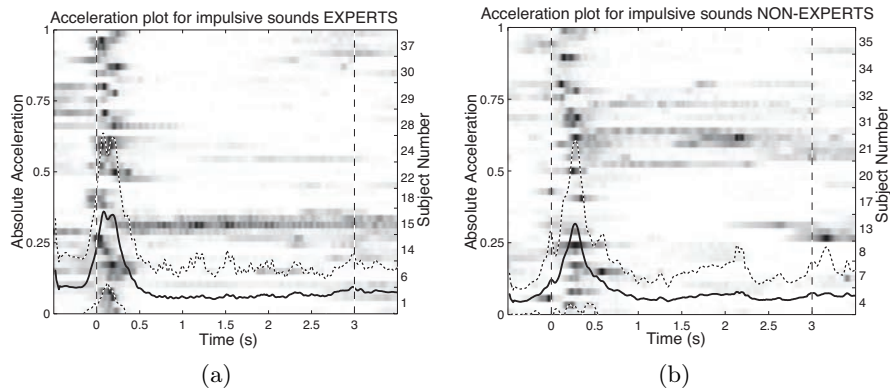


Fig. 7. The plot from Figure 6(b) separated into (a) experts and (b) non-experts.

5.3 Vertical Position and Sound Frequency Features

As mentioned in Section 2, other researchers have documented a relationship between vertical position and pitch. Not surprisingly, this relationship was also found in the data set presented in this paper. In addition to pitch, we observed that the frequency of the spectral centroid is relevant to the vertical position.

Sounds 1, 4 and 5 all had rising pitch envelopes, and Sounds 8 and 11 had rising spectral centroids combined with stable pitch and noise respectively. For the sound tracings of these sounds, there was a clear tendency of upward movement. Similarly, for the sounds with falling pitch, or with falling spectral centroid, there was a clear tendency of downward movement. *t*-tests comparing the average vertical velocity of the “rising” sounds to the “falling” sounds showed highly significant distinctions between the groups, as shown in Table 3. The mean normalized vertical velocity for the first group was 0.74, and for the second group 0.28 (a value of 0.5 indicates no vertical motion). This is shown in Figure 8.

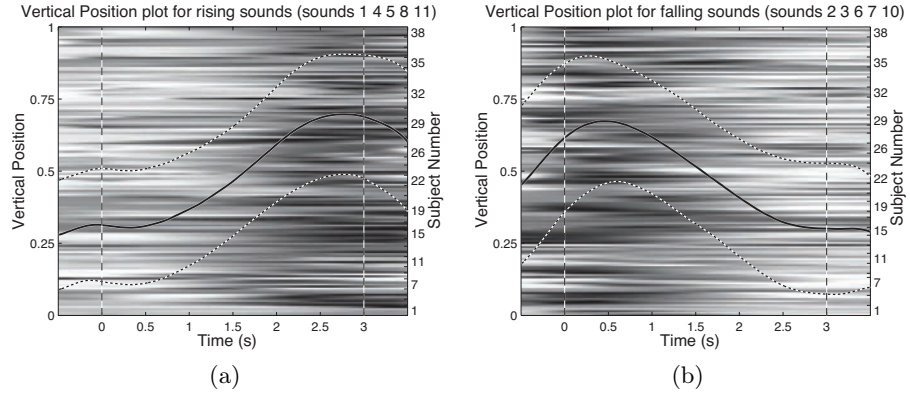


Fig. 8. Vertical position for (a) rising sounds (142 sound tracings) and (b) falling sounds (144 sound tracings). The black line shows the mean value across all the data series, each horizontal line in the image displays the vertical position of a single sound tracing normalized per subject between 0 (lower position, white) and 1 (higher position, black).

Table 3. T-tests comparing the average vertical velocity of rising and falling sounds.

Average vertical velocity, rising and falling sounds			
Test description	<i>df</i>	<i>t</i>	<i>p</i>
Rising vs falling, all subjects	284	18.89	< 0.01
Rising vs falling, non-expert subjects	98	8.86	< 0.01
Rising vs falling, expert subjects	97	11.69	< 0.01
Rising, experts vs non-experts	98	0.58	0.28
Falling, experts vs non-experts	97	1.79	0.04

There was no significant difference between the average vertical velocity for experts and non-experts for the rising sounds, however, for the falling sounds there was some difference between the two groups. Experts had a higher extent of downward motion than non-experts, $t(97) = 1.7982$, $p = 0.04$.

It is worth noting that even though Sounds 1 and 2 had increasing and decreasing pitch envelopes, respectively, they had opposing spectral centroid envelopes. When the spectral centroid envelope and the pitch envelope moved in opposite directions, most subjects in our data set chose to let the vertical motion follow the direction of the pitch envelope. The direction of vertical motion seems to be more strongly related to pitch than to spectral centroid.

The observed difference between sounds with varying pitch and sounds with only varying spectral centroid makes it interesting to take a more in depth look at the individual sounds in the *rising* and *falling* classes. Since subjects tended to follow pitch more than spectral centroid in the sounds where the two feature envelopes moved in opposite directions, it is natural to assume that subjects would move more to sounds where the pitch was varied, than to sounds where only the spectral centroid was varied. Figure 9 displays box plots of the average vertical velocities for rising and falling sounds. In Figures 9(a) and 9(b), we observed that the difference between the sounds is larger for falling than for rising sounds. We can also see that Sounds 7 and 8, which are sounds where the pitch is constant but spectral centroid is moving, show less extreme values than the rest of the sounds. Figures 9(c) and 9(d) suggest that the difference between the sounds is larger for expert subjects than for non-expert subjects. There also seems to be more inter-subjective similarities among experts than non-experts, as the variances among experts are lower.

Table 4 shows the results of one-way analyses of variance (ANOVAs) applied to the sound tracings in the rising and falling class, respectively. The table shows that on the one hand, the difference in vertical velocity between the falling sounds was statistically significant $F(4, 139) = 7.76$, $p < 0.01$. On the other hand, the corresponding difference between the rising sounds was not statistically significant $F(4, 137) = 1.53$, $p = 0.20$. The table also reveals that the significant difference between the groups was only present for expert subjects, $F(4, 44) = 4.92$, $p < 0.01$, and not for non-experts, $F(4, 45) = 1.52$, $p = 0.21$.

Table 4. Results from one-way ANOVAs of the vertical velocity for sound tracings within the *rising* and *falling* classes. There is a significant difference between the five falling sounds for expert subjects. *df* are the degrees of freedom (between groups, within groups), *F* is the F-value with the associated f-test, *p* is the probability that the null-hypothesis is true.

Subjects	Rising sounds			Falling sounds		
	<i>df</i>	<i>F</i>	<i>p</i>	<i>df</i>	<i>F</i>	<i>p</i>
All subjects	(4, 137)	1.53	0.20	(4, 139)	7.76	< 0.01
Expert subjects	(4, 45)	0.39	0.81	(4, 44)	4.92	< 0.01
Non-expert subjects	(4, 45)	0.25	0.90	(4, 45)	1.52	0.21

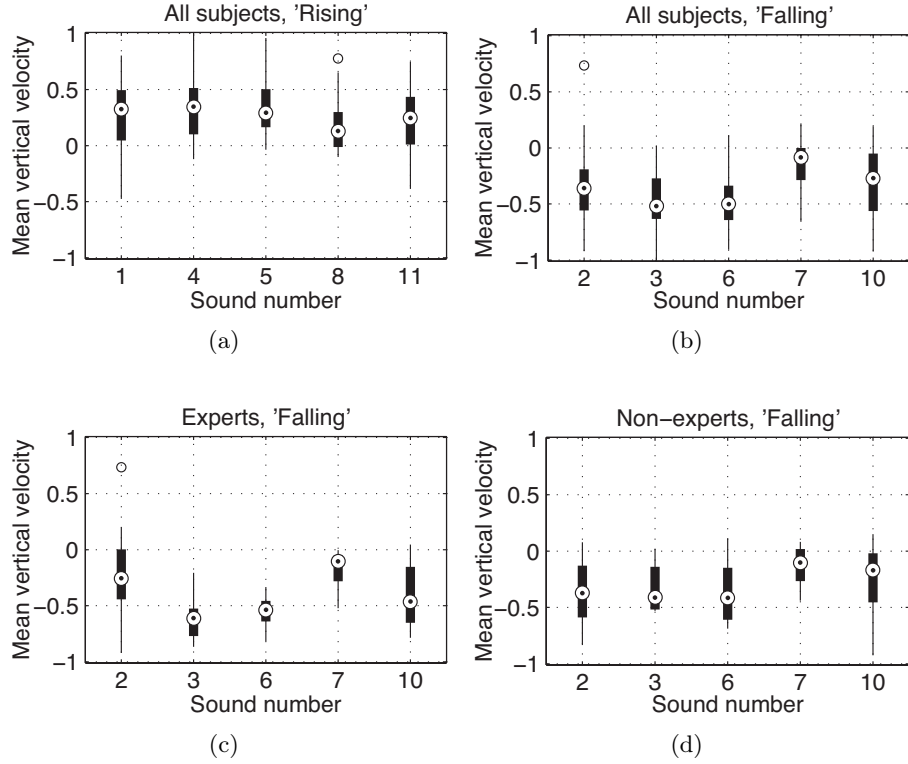


Fig. 9. Box plots of average vertical velocities for (a) rising sounds and (b) falling sounds. The difference between the sounds is greater for falling sounds than for rising sounds, and greater for (c) experts than for (d) non-experts. Note also that the sounds where the pitch is constant and only the spectral centroid is manipulated (Sounds 7 and 8) have the least extreme values in all the plots.

5.4 Pitch and Distance Between Hands

Eitan and Timmers pointed out that the relationship between pitch and motion features may be more complex than mapping pitch to vertical position [6]. For this reason, we have also analyzed how the distance between the hands corresponds to pitch frequency.

Figures 10(a) and 10(b) display the distance between the hands for sound tracings to rising and falling sounds, respectively. On the one hand, the black lines displaying the average distance features do not show very clear overall tendencies towards increasing or decreasing distance, but on the other hand, the underlying images, displaying the individual sound tracings, show that there is a substantial amount of change in the distance between the hands, both for the rising and the falling sounds.

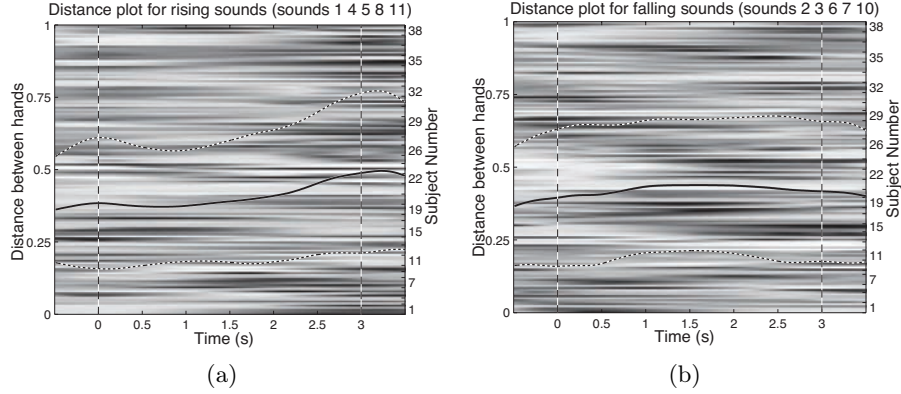


Fig. 10. Distance between hands for (a) rising sounds (142 sound tracings) and (b) falling sounds (144 sound tracings). The black line shows the mean value across all the data series, each horizontal line in the image displays the vertical position of a single sound tracing normalized per subject between 0 (hands close together, white) and 1 (hands far apart, black).

Figure 10 seems to vaguely suggest that participants let the hand distance increase for sounds with increasing pitch. Table 5 compares the change in distance for rising sounds versus falling sounds. For the two sounds where only the pitch is varied, there is a significant difference between the rising and falling sound tracings. The same is true when all of the rising sounds are compared to the falling sounds. On the contrary, we do not observe significant differences in the instances where the spectral centroid is varied.

Table 5. Results from t -tests comparing the average change in distance between hands of rising and falling sounds.

Average change in hand distance, rising versus falling sounds				
Sounds	Varying features	df	t	p
5 vs. 6	Pitch	52	3.24	< 0.01
4 vs. 3	Pitch and spectral centroid	54	1.50	0.07
1 vs. 2	Pitch, opposing spectral centroid	53	1.09	0.14
8 vs. 7	Spectral centroid, pitched	60	0.48	0.32
11 vs. 10	Spectral centroid, non-pitched	57	-0.44	0.67
All rising sounds vs. all falling sounds		284	2.60	< 0.01

5.5 Acceleration Envelope for Pitched and Non-pitched Sounds

We have evaluated how the presence of a distinct pitch influences the sound tracing by comparing acceleration envelopes of sound tracings performed to pitched sounds and non-pitched sounds.

Three of the sounds used in the experiment were based on noise, and three were based on a stable tone with a fundamental frequency of 342 Hz. Within each of these categories, one sound had a falling spectral centroid, one had a rising spectral centroid and one had a stable spectral centroid.⁸ Figure 11 shows the acceleration curves from the sound tracings to non-pitched and pitched sounds respectively. The mean acceleration was significantly higher for non-pitched sounds than pitched sounds, $t(179) = 5.53$, $p < 0.01$. For non-pitched sounds the mean normalized acceleration was 0.52, and for pitched sounds it was 0.28.

This significant distinction between acceleration values for pitched and non-pitched sounds was also found when the data from experts and non-experts was analyzed individually. Furthermore, no significant difference was found between the acceleration levels of experts and non-experts, $p = 0.46$ for both pitched and non-pitched sounds, respectively. See Table 6 for statistical results.

Table 6. Results from t -tests comparing acceleration of pitched to non-pitched sounds.

Acceleration, non-pitched and pitched sounds			
Test description	df	t	p
Non-pitched vs pitched, all subjects	179	5.53	< 0.01
Non-pitched vs pitched, expert subjects	62	3.31	< 0.01
Non-pitched vs pitched, non-expert subjects	62	3.68	< 0.01
Noise, experts vs non-experts	61	0.10	0.46
Stable tone, experts vs non-experts	63	0.11	0.46

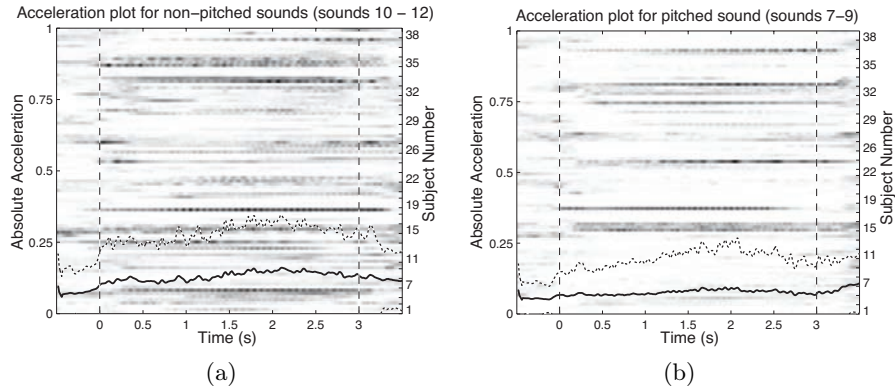


Fig. 11. Acceleration for (a) non-pitched sounds and (b) pitched sounds. The black line shows the mean value per sample, each horizontal line displays the acceleration of a single sound tracing ranging from 0 (white) to 1 (black). The mean acceleration levels for the non-pitched sounds are generally higher than for pitched sounds.

⁸ Sounds based on noise: 10, 11, and 12. Sounds based on a stable tone: 7, 8, and 9.

5.6 Summary of the Results

The results in this section have shown that the participants produced sound tracings with higher onset acceleration for impulsive sounds than for non-impulsive sounds. This was true for experts and non-experts. The onset time of musical experts was on average 75 ms ahead of non-experts. Furthermore, sounds without a distinct pitch seem to induce higher acceleration than pitched sounds.

Vertical displacement of the hands was found to be related to pitch and to spectral centroid. When pitch and spectral centroid moved in opposite directions, most subjects let the vertical position of the hands follow the perceived pitch. When only the spectral centroid was varied, there was less vertical motion than for sounds with varying pitch. This was particularly true for sounds with a stable pitch, as opposed to no perceivable pitch. Overall, falling sounds induced more vertical motion than rising sounds. For the falling sounds, the variance between the vertical velocity of the subjects was low, suggesting that there is more consistency within the expert group than in the non-expert group. Finally, there was significant difference between the change in hand distance for some of the sounds with falling and rising envelopes. We will discuss these findings in the next section.

6 Discussion

The following discussion will address the analysis method, the results from the previous section, as well as how the experiment setup and task may have influenced the results. We will put the results into context in relation to previous research, and in this way try to assess what can be learned from our findings. For certain sound features and specific motion features, we have observed a quite high consistency across the subjects. This supports the claim that there is a relationship between auditory and motor modalities.

The discussion is structured as follows: The statistical approach is discussed in Section 6.1. In Section 6.2, we evaluate the results from Section 5.2. Section 6.3 discusses the results from Sections 5.3 and 5.4, and results from Section 5.5 are discussed in Section 6.4. In Section 6.5, we provide a more general evaluation of the results.

6.1 Statistical Method

Using statistics to evaluate the differences between the groups does provide some indication of the tendencies in our data set. However, it should be noted that the *t*-test and ANOVA methods assume that the data is normally distributed. The subsets of data in our statistical analyses were tested for normality using a Jaque-Bera test⁹ with significance level 0.05. This test revealed that 13 out of the 52 data sets in our experiments do not follow a normal distribution, and thus the statistical results can not alone be used to make strong conclusions.

⁹ http://en.wikipedia.org/wiki/Jarque-Bera_test

Nevertheless, the results from the statistical tests support the results that are shown in the corresponding feature plots. This gives us reason to believe that the statistical results are trustworthy.

It should be noted that for the sample-by-sample based plots, the standard deviations are quite high, particularly for the acceleration curves shown in Figure 11. Since these plots were derived on a sample-by-sample basis, the high standard deviations are not very surprising. In Figure 11, the high standard deviation reflects that even though several subjects had a high overall acceleration, they vary between high and low acceleration throughout the sound tracing. This demonstrates the importance of looking at the individual sound tracings in the plot, not only the mean and standard deviation curves.

6.2 Impulsive Sound Onset

Let us have a look at the results presented in Section 5.2, where the onset acceleration of sound tracings to impulsive sounds was shown to be much higher than for non-impulsive sounds. In our opinion, these results can best be explained from a *causation* perspective. In other words: people link the impulsive characteristics of the sound to some sort of impulsive action that could have generated it. An example of an impulsive action is displayed in Figure 12. The figure shows how the subject performs an accentuated attack, with high acceleration, followed by a slower falling slope down to a resting position. The sound tracing resembles that of crashing two cymbals together.

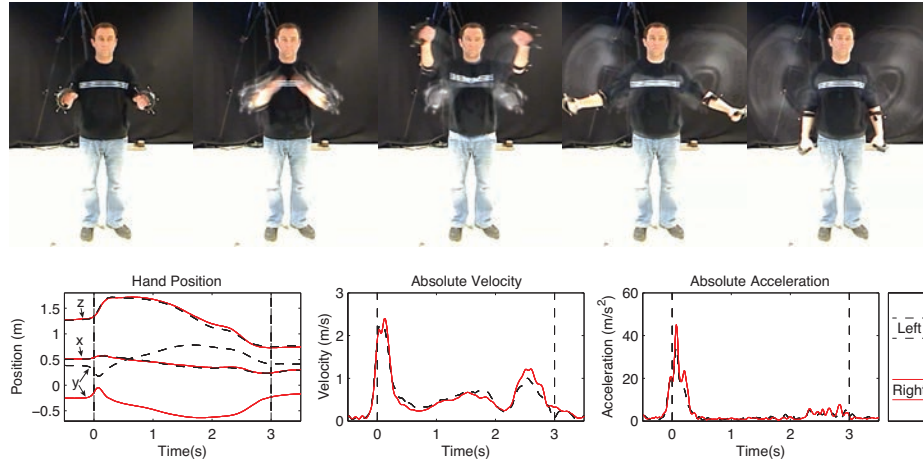


Fig. 12. The figure displays an impulsive action, with motion history images and the calculated motion features for the left (black) and right (red) hand. Motion history images show the current frame and the average frame difference for past video frames [16]. x (back/forth), y (sideways) and z (up/down) are position coordinates from the motion capture system.

Lakoff and Johnson [18] talk about causation as an important element in understanding objects and events in the world. As an example, they mention a paper airplane which we understand primarily as paper, and secondarily as airplane. The making, or cause, of the paper airplane is an essential element in our understanding of this object. It is interesting to compare the *causation* term to Schaeffer’s theory of the sonic object [26]. According to Schaeffer, sonic objects are defined by their *causal coherence*, something which Godøy linked to gesture-sensations and chunking in sound [9]. According to Lakoff and Johnson [18], the causation of an object is partly emergent, or something that is present in the object itself, which makes it possible to understand an object as a holistic and metaphorical unit. Again, compared with the sonic object, this emergent property works well with Schaeffer’s principle of *stress-articulation*, or natural discontinuities in the continuous sound signal [26].

Following these thoughts, it seems apparent that people link the impulsive onset of sounds to some sort of impulsive or ballistic action or event. Given the constraints of two handles to imitate the sound, some participants imitated the action of crashing two cymbals together, while others imitated a single-handed or two-handed striking action. The discontinuity of stress-articulation in sound has its motor counterpart in the higher derivatives of position data, here shown by a high onset acceleration.

6.3 Sound Frequency Features

In addition to the causation perspective, Lakoff and Johnson also introduced the *metaphor* perspective. Metaphors are crucial to our understanding of events and objects. We understand some event or object by using a metaphor to describe it. According to Eitan and Timmers [6], Cox [4] has linked the metaphor “more is up” to a perceptual relationship between vertical position and pitch. In our experiment, the results show clearly that there is a relationship between these features, and that most subjects follow rising pitch with upward motion, and falling pitch with downward motion.

However, Eitan and Timmers have shown that for pitch, up is not always the best metaphor [6]. In their experiments, low pitch has also been associated with metaphors like “heavy” or “big”. Also, Walker [27] described rising pitch to be a good descriptor for increasing temperature. For this exact reason, we also investigated if the motion feature *hand distance* was related to the rising and falling envelopes. Our results show that when all the rising sounds were compared to the falling ones, there was a significant difference in the average change in hand distance. However, a closer look at the results revealed that on average, for sounds with a rising envelope and with a falling envelope alike, the distance between the hands increased. The significant difference was therefore only due to a faster increase in distance for rising sounds than for falling sounds. In addition, a significant difference between rising and falling envelopes occurred when only the parameter *pitch* was varied. In this case, the average hand distance decreased for falling pitch and increased for rising pitch, and thus to some extent, defined a relationship between these features. Nevertheless, even though several

subjects did change the distance between their hands for these sounds, there was much less similarity among the subjects compared to the vertical position feature. Some subjects moved their hands apart while other moved them towards each other. So to conclude, the “more-is-up” metaphor for pitch seems to be the best metaphor to describe the results in our data set.

An example of a sound tracing performed to a sound with falling pitch and rising spectral centroid is shown in Figure 13. The motion history images show how the subject prepares for the sound tracing by moving his hands up, then moving them down and out in such a way that the vertical position follows the pitch envelope. At the end of the sound tracing, the subjects increasingly vibrates the right hand, as shown in the acceleration plot. This might be a gesture performed to imitate the increased spectral centroid which is increasingly prominent towards the end of the sound file. As the motion history images show, the hand distance first increases and then decreases in a sound where the pitch is constantly falling and the spectral centroid is constantly rising.

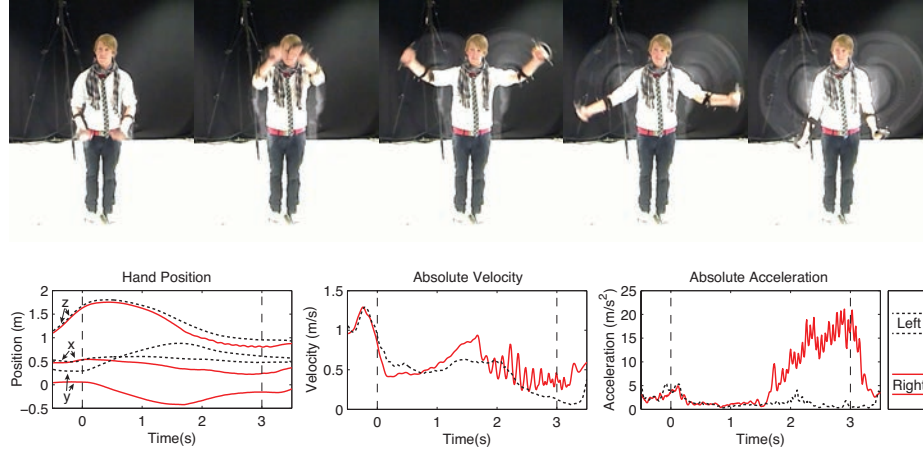


Fig. 13. The figure displays motion history images and feature plots for the left and right hand for a sound tracing performed to a sound with falling pitch, rising spectral centroid, and a bell-shaped dynamic envelope.

An interesting feature regarding the rising and falling sounds, and the correlation to vertical position, is that sound seems to be more descriptive than motion. Our results show that even though the sounds were easy to tell apart, the sound tracings that were performed to the different sounds were similar. This implies that although you can describe certain perceptual features of the sound through an action, it is not necessarily clear which perceptual feature(s) the action imitates. Elevating a hand might refer to increasing pitch or to increasing spectral centroid, or to something else.

6.4 Pitched versus Non-pitched Sounds

We observed a significantly higher amount of acceleration in sound tracings performed to non-pitched sounds than to pitched sounds. This may be explained by participant’s associations with this sound property. Sounds based on noise have wind-like properties, which might cause people to move a lot, as if they were blowing with the wind or creating the wind themselves. Pitched sounds, on the other hand, seem to provide something stable for the participants to “hold on to”, that is not provided by the non-pitched sounds.

For Sounds 9 and 12, which both had stable spectral centroids, we observed that some participants started shaking or rotating their hands, gradually increasing the frequency or amplitude of the shaking. One example of this is shown in Figure 14. As these sounds had no change in pitch or spectral centroid, the loudness envelope of the sounds seem to have been the main influence in these instances. The increased shaking or rotation intensity may be explained by some sort of engine metaphor: we believe participants wanted to follow the gradually increasing loudness envelope by supplying more and more energy to the sound through their motion.

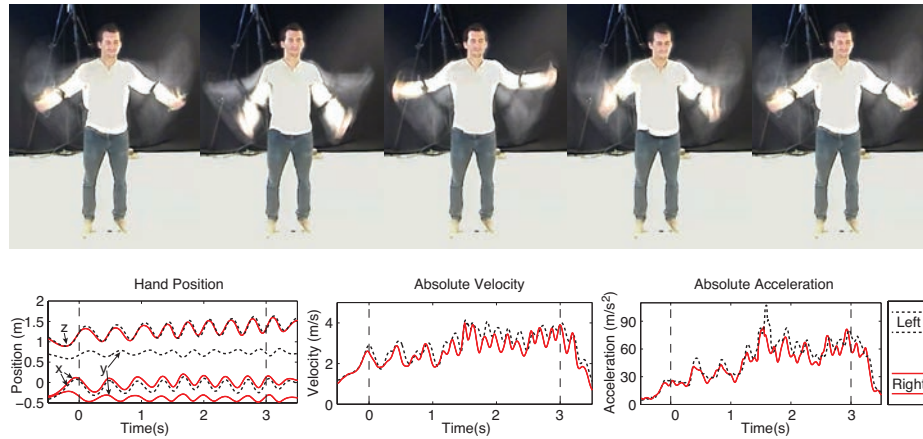


Fig. 14. The figure shows motion history images of a subject moving to a noise-based sound. Plots of xyz-position, velocity, and acceleration are shown below.

6.5 Final Remarks

Admittedly, the results presented in this paper are related to the context of the experiment, and cannot necessarily be claimed to be valid outside this setting. The way in which subjects solved the task may have been influenced by the instructions, which were to imagine that moving the hands in the air created the sound. Even though we did not provide the subjects with an a priori metaphor

connecting upward motion to increasing pitch, the options for movement were limited. Godøy has postulated that our musical motions are goal-directed, and that these motions are guided by goal-postures describing the shape and position of our end-effectors [11]. These positional goal-points may have been more consciously exposed than features describing the relationship between the hands or details of the trajectory between, for instance, a high and low goal point. In light of the experiment task, other pitch relationships like the one between low pitch and “heavy” or “big” [6], may have been less accessible than drawing a trajectory between two positional goal points. Some subjects may have consciously or unconsciously used a feature like hand distance to describe pitch, but as our results show, this was more inconsistent between subjects.

Even though the experimental setup may have prevented the subjects from using other descriptors than spatio-temporal ones, we are confident that the results show some indication of how people relate spatio-temporal features of motion to features of sound. The participants were given an imagined instrument, and they made their own mental model of how this instrument worked. Some aspects of these mental models were similar for the majority of the subjects.

7 Conclusions and Future Work

We have presented an experiment on sound tracing, where motions performed to sound have been analyzed from micro and meso timescale perspectives. Plotting of serial motion features at the micro timescale was used to obtain impressions of general tendencies in the data set, and statistical evaluations of single-value features at the meso timescale indicated the significance of these tendencies.

Rising pitch, and rising spectral centroid correlated strongly with upward motion, and similarly, falling pitch and spectral centroid, correlated strongly with downward motion. When pitch and spectral centroid moved in opposite directions, participants followed the pitch feature. Furthermore, sounds based on noise induced higher overall acceleration than sounds based on a steady pitch, and sounds with an impulsive onset caused a high acceleration peak in the beginning of the sound tracing.

To follow up on this experiment, we are currently starting to evaluate advantages and disadvantages of different methods for analyzing sound tracings. We believe that the different approaches that have been taken to analyze such data provide different types of knowledge, and that the choice of analysis method is important. For instance, some methods may be good at revealing action-sound relationships at a low timescale level, while others may work better at chunk-level or higher. We are also moving towards applying the results from our studies in development of new interfaces for musical expression.

8 Acknowledgements

The research leading to these results has received funding from the European Union Seventh Framework Programme under grant agreement n° 257906, Engineering Proprioception in Computer Systems (EPiCS).

References

1. Publication Manual of the American Psychological Association. American Psychological Association, Washington, DC, 6th edn. (2010)
2. Caramiaux, B., Bevilacqua, F., Schnell, N.: Towards a gesture-sound cross-modal analysis. In: Kopp, S., Wachsmuth, I. (eds.) *Gesture in Embodied Communication and Human-Computer Interaction*, LNCS, vol. 5934, pp. 158–170. Springer, Heidelberg (2010)
3. Caramiaux, B., Bevilacqua, F., Schnell, N.: Sound selection by gestures. In: *Proceedings of the International Conference on New Interfaces for Musical Expression*. pp. 329–330. Oslo, Norway (2011)
4. Cox, A.W.: *The Metaphoric Logic of Musical Motion and Space*. Ph.D. thesis, University of Oregon (1999)
5. Eitan, Z., Granot, R.Y.: How music moves: Musical parameters and listeners' images of motion. *Music Perception* 23(3), 221–248 (2006)
6. Eitan, Z., Timmers, R.: Beethoven's last piano sonata and those who follow crocodiles: Cross-domain mappings of auditory pitch in a musical context. *Cognition* 114(3), 405–422 (2010)
7. Gjerdingen, R., Perrott, D.: Scanning the dial: The rapid recognition of music genres. *Journal of New Music Research* 37(2), 93–100 (2008)
8. Glette, K., Jensenius, A.R., Godøy, R.I.: Extracting action-sound features from a sound-tracing study. In: *Proceedings of the Second Norwegian Artificial Intelligence Symposium*. pp. 63–66 (2010)
9. Godøy, R.I.: Gestural-sonorous objects: Embodied extensions of Schaeffer's conceptual apparatus. *Organised Sound* 11(2), 149–157 (2006)
10. Godøy, R.I.: Chunking sound for musical analysis. In: Ystad, S., Kronland-Martinet, R., Jensen, K. (eds.) *Computer Music Modeling and Retrieval. Genesis of Meaning in Sound and Music*, LNCS, vol. 5493, pp. 67–80. Springer, Heidelberg (2009)
11. Godøy, R.I.: Gestural affordances of musical sound. In: Godøy, R.I., Leman, M. (eds.) *Musical Gestures: Sound, Movement, and Meaning*, chap. 5, pp. 103–125. Routledge (2010)
12. Godøy, R.I., Haga, E., Jensenius, A.R.: Exploring music-related gestures by sound-tracing. A preliminary study. In: Ng, K. (ed.) *COST287-ConGAS 2nd Int. Symposium on Gesture Interfaces for Multimedia Systems*. pp. 27–33. Leeds (2006)
13. Godøy, R.I., Jensenius, A.R.: Body movement in music information retrieval. In: *Proceedings of the 10th International Society for Music Information Retrieval Conference*. Kobe, Japan (2009)
14. Husserl, E.: *The Phenomenology of Internal Time Consciousness*. Indiana University Press, Bloomington (1964, trans JS Churchill)
15. Jensenius, A.R.: *Action-Sound : Developing Methods and Tools for Studying Music-Related Bodily Movement*. Ph.D. thesis, University of Oslo (2007)

16. Jensenius, A.R.: Some video abstraction techniques for displaying body movement in analysis and performance. *Leonardo Music Journal* (2012, to appear)
17. Kozak, M., Nymoen, K., Godøy, R.I.: The effects of spectral features of sound on gesture type and timing. In: Efthimiou, E., Kouroupetroglou, G., Fotinea, S.E. (eds.) "Gesture and Sign Language in Human-Computer Interaction and Embodied Communication". 9th International Gesture Workshop - GW 2011, May 2011, Athens, Greece. Revised selected papers., *Lecture Notes in Computer Science/LNAI*, vol. 7206. Springer, Berlin Heidelberg (2012, to appear)
18. Lakoff, G., Johnson, M.: *Metaphors We Live By*. University of Chicago Press, Chicago, IL. (1980)
19. Leman, M.: *Embodied Music Cognition and Mediation Technology*. MIT Press, Cambridge, Mass. (2007)
20. Leman, M.: Systematic musicology at the crossroads of modern music research. In: Schneider, A. (ed.) *Systematic and Comparative Musicology: Concepts, Methods, Findings*, *Hamburger Jahrbuch für Musikwissenschaft*, vol. 24, pp. 89–115. Peter Lang, Frankfurt (2008)
21. Merer, A., Ystad, S., Kronland-Martinet, R., Aramaki, M.: Semiotics of sounds evoking motions: Categorization and acoustic features. In: Kronland-Martinet, R., Ystad, S., Jensen, K. (eds.) *Computer Music Modeling and Retrieval. Sense of Sounds*, pp. 139–158. No. 4969 in *Lecture Notes in Computer Science*, Springer-Verlag, Heidelberg (2008)
22. Miller, G.A.: The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review* 63, 81–97 (1956)
23. van Nort, D.: Instrumental listening: sonic gesture as design principle. *Organised Sound* 14(02), 177–187 (2009)
24. Nymoen, K., Caramiaux, B., Kozak, M., Torresen, J.: Analyzing sound tracings: a multimodal approach to music information retrieval. In: *Proceedings of the 1st international ACM workshop on Music information retrieval with user-centered and multimodal strategies*. pp. 39–44. ACM, New York, NY, USA (2011)
25. Nymoen, K., Glette, K., Skogstad, S.A., Torresen, J., Jensenius, A.R.: Searching for cross-individual relationships between sound and movement features using an SVM classifier. In: *Proceedings of the International Conference on New Interfaces for Musical Expression*. pp. 259–262. Sydney (2010)
26. Schaeffer, P., Reibel, G.: *Solfège de l'objet sonore*. ORTF, Paris, France (1967, Reedited by INA-GRM 1998)
27. Walker, B.N., Kramer, G.: Mappings and metaphors in auditory displays: An experimental assessment. *ACM Trans. Appl. Percept.* 2(4), 407–412 (2005)
28. Wilson, M.: Six views of embodied cognition. *Psychonomic Bulletin & Review* 9(4), 625–636 (2002)