Representation Strategies in Two-handed Melodic Sound-Tracing

Tejaswinee Kelkar
University of Oslo
Department of Musicology, fourMs Lab
Oslo, Norway
tejaswinee.kelkar@imv.uio.no

Alexander Refsum Jensenius
University of Oslo
Department of Musicology, fourMs Lab
Oslo, Norway
a.r.jensenius@imv.uio.no

ABSTRACT

This paper describes an experiment in which the subjects performed a sound-tracing task to vocal melodies. They could move freely in the air with two hands, and their motion was captured using an infrared, marker-based system. We present a typology of distinct strategies used by the recruited participants to represent their perception of the melodies. These strategies appear as ways to represent time and space through the finite motion possibilities of two hands moving freely in space. We observe these strategies and present their typology through qualitative analysis. Then we numerically verify the consistency of these strategies by conducting tests of significance between labeled and random samples.

KEYWORDS

Sound-tracing, music-related body motion, melody perception, motion capture, musical gesture

ACM Reference format:
Tejaswinee Kelkar and Alexander Refsum Jensenius. 2017. Representation Strategies in Two-handed Melodic Sound-Tracing . In *Proceedings of Movement and Computing Conference, London, UK, July 2017 (MOCO'17),* 4 pages.
DOI: 10.475/123_4

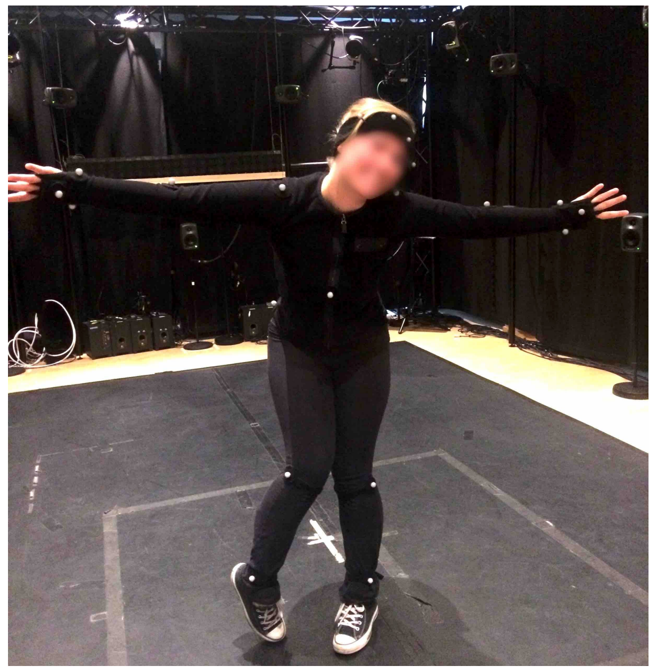


Figure 1: The experimental setup for the sound-tracing study: a subject wearing a motion capture suit with 21 markers, with eight wall-mounted cameras in the lab.

1 INTRODUCTION

Moving the hands in the air when listening to music, what we will refer to as *sound-tracing*, may be seen as a representation of our experience of the music in question [8]. Spontaneous sound-tracing can be found in many different musical contexts, both in public (such as at concerts) or in private settings. There are also examples of more formalized sound-tracing, for example in various types of dance and theatre performances, but also in music performance, such as the music-accompanying hand gesturing found in North Indian music. The common denominator seems to be that the hand actions represent changes in musical expression over time. They may also be seen to have both representative and communicative functions, not unlike co-speech hand gestures [10] [5]. The cognitive aspects of such motion have been explained through experimental listener studies [2], and by comparing the relationship of sound and music features in various contexts [6].

Sound-tracing has also been used as a research method to understand more about our cognition of music by looking at how people move when listening to music [7, 12, 13, 17]. The starting point here is the belief that our cognition of music is inherently *embodied* in nature [9], and that this may be connected to the way in which the auditory and motor regions of the brain are connected to each other at a neural level. Auditory activity automatically triggers motor regions in the brain, regardless of any physical motion being carried out [3]. The theory of bimanual motor control, symmetrical patterns of hand motion, and the influence of perception for the production of hand motion on each other is also pertinent to the discussion of this study. [11]

Several of the previous sound-tracing studies have found that there seems to be a correlation between how diverse people interpret the same musical stimuli. This paper is concerned with *melody*,

DOI: 10.475/123_4

	Joik	Scat	Vocalises	NICM
Processed	4 Stimuli	4 Stimuli	4 Stimuli	4 Stimuli
Unprocessed	4 Stimuli	4 Stimuli	4 Stimuli	4 Stimuli

Table 1: The 32 stimuli types used in the experiment.

that is, sequences of pitches, and how people trace melodies with their hands. Pitch appears to be one important musical feature that people easily relate to when tracing sounds, even with changing timbres [13]. Melodic contour has been studied in terms of symbolic pitch [15, 18]. There have also been proposed analytical methods to grouping melodic contours and for computing similarity measures for two distinct melodies [1, 16]. The perception of melodic contour, and our subjective experience of such contours, have also been investigated through isochronous symbolic melodies [4].

The following research questions were the most important for the scope of this paper:

(1) Are melodic sound-tracings by several people systematic?
(2) Which strategies are used for melodic sound-tracings?

Although the question of sound–motion correspondences is of high interest to the authors, in this paper we primarily focus on motion *properties* and motion *strategies*. We start by describing the experimental setup (Figure 1 before describing the results of both qualitative and quantitative analyses of the gathered data.

2 EXPERIMENT DESCRIPTION

The aim of the experiment was to analyse the complex relationships between music-induced body motion and the perception of melodic contour in the musical examples. The participants were instructed to move their hands as if they were creating the melodic fragments that they heard. The idea was that they would "shape" the sound with their hands in physical space. As such, this type of free-hand sound-tracing task is quite different from some sound-tracing experiments using pen on paper or on a digital tablet. Participants in a free-hand tracing situation would be less fixated upon the precise location of their actions, thus revealing perceptually salient properties of the melodies that they choose to represent.

2.1 Stimuli

We selected 16 melodic fragments from four genres of music that use vocalisations without words: (1) scat singing, (2) western classical vocalise, (3) Sami joik, (4) North Indian music. The melodic fragments complete phrases taken from real recordings containing, to retain melodies within their original musical context. Vocal melodies were chosen to eliminate the perceptual effect of lyrics, but also to eliminate the possibility of imitating the sound-producing actions on instruments ("air-instrument" performance).

The average duration of the phrases was 4.5s (SD=1.5s), with a 2s pause between phrases. The samples were presented in two conditions: (1) the real recording, and (2) a re-synthesised version using a sawtooth waveform to play the extracted pitch profile. There was thus a total of 32 stimuli per participant, as sketched in Table 1, played in random order.

The sounds were played at comfortable listening level through a Genelec 8020 speaker, placed 3 metres ahead of the participants. The experiment lasted for a total duration of 10 minutes.

2.2 Participants

A total of 32 participants (17 female, 15 male) were recruited, with a mean age of 31 years (SD=9y). The participants were mainly students and employees from the University of Oslo, both with and without musical training. Their musical experience was quantified using the OMSI questionnaire [14], with an average score of 694 (SD=292). The participants also reported on their familiarity with the musical genres used and signed consent forms. They were free to withdraw during the experiment, if they wished. The study obtained ethical approval from the Norwegian Centre for Research Data (NSD).

2.3 Lab setup

The experiment was run in our motion capture lab, using a Qualisys motion capture system with eight wall-mounted Oqus 300 cameras (Figure 1), capturing at 200 Hz. The experiment was conducted in dim light, with no observers, to make sure that participants felt free to move as they liked. A total of 21 markers were placed on the body of the participants: the head, shoulders, elbows, wrists, knees, ankles, the torso, and the back of the body. The recordings were post-processed in Qualisys Track Manager (QTM) and analysed further in Matlab.

3 QUALITATIVE OBSERVATIONS

Based on a qualitative observation of the data, we propose six schemes of representation that encompass most of the variation in the hands' motion, as illustrated in Figure 2 and summarised as:

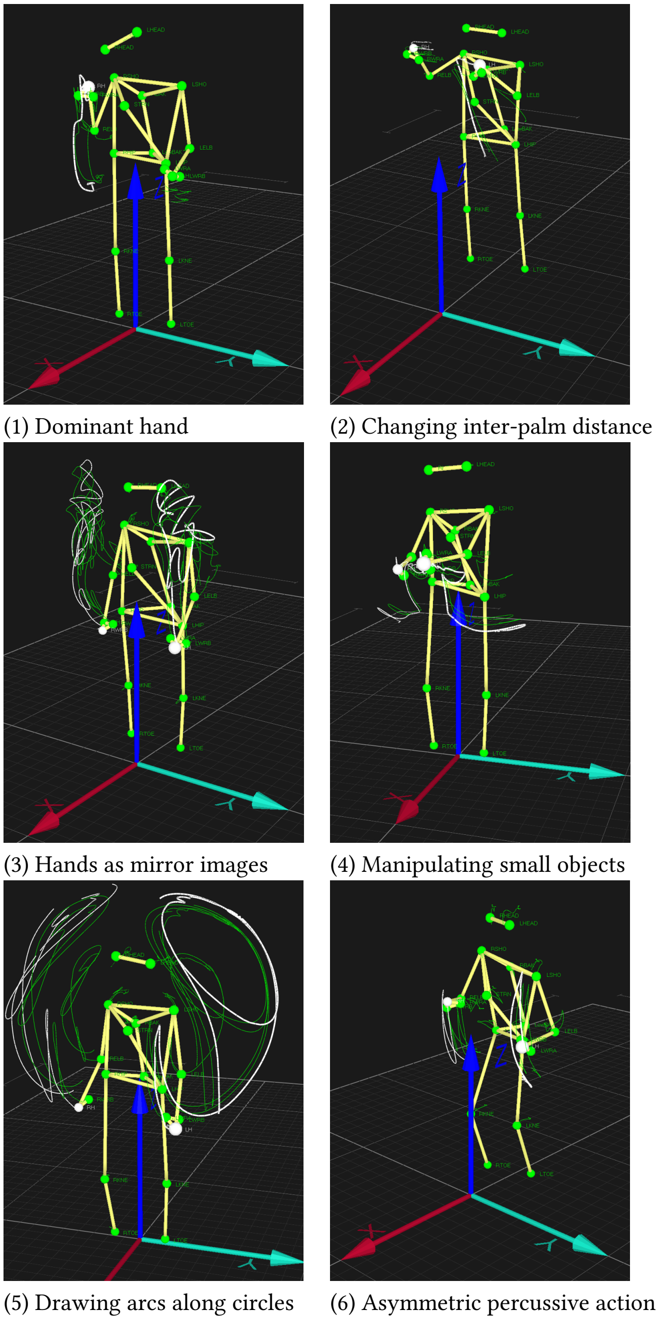
(1) One outstretched hand, changing the height of the palm
(2) Two hands stretching or compressing an "object"
(3) Two hands symmetrically moving away from the centre of the body
(4) Two hands moving together to represent the motion of a smaller body through space
(5) Two hands drawing arcs along an imaginary circle
(6) Two hands moving opposite to each other in a percussive pattern

A general observation is that the end point of the palm seems to be the most significant "object" representing the melody. It is also the palm that is modified according to changes in timbre and vowel shapes. Another qualitative finding is that people appear to switch between the different tracing strategies depending upon the musical stimulus.

4 QUANTITATIVE ANALYSIS

Even though we recorded 21 markers per subject, we will only consider the left and right hand markers here. Thus the starting point of the quantitative analysis is a six-column dataset as the input vector, sampled at 200 Hz. The pipeline for the analysis is sketched in Figure 3, and consists of the following steps:

- Feature Selection: In this stage, we segment the motion capture data into the relevant six-column feature vector



(1) Dominant hand (2) Changing inter-palm distance

(3) Hands as mirror images (4) Manipulating small objects

(5) Drawing arcs along circles (6) Asymmetric percussive action

Figure 2: Motion history images exemplifying the six dominant sound-tracing strategies.

containing the (x,y,z) coordinates of the right palm and the left palm respectively.

- Calculate Quantity of Motion (QoM): Calculated as the average of the vector magnitude for each sample.
- Segmentation: The data is trimmed using a sliding window of 1100 samples in size. This corresponds to 5.5s, to accommodate the average duration of 4.5s of the melodic phrases. The windows are segmented at every 10 samples to obtain a large set of sliding windows from the data. The windows that have the maximum mean values are then separated out to get a set of sound tracings. Examples of some tracing segments can be seen in Figure 4.
- Feature Analysis: The features mentioned in Table 2 are analysed for each of the segmented tracings. Statistics for these are calculated.
- Thresholding: If a sample tracing segment truly belongs exclusively to any of these strategies, the six numerical criteria should be minimized. The segmented tracings are therefore thresholded based on 2 times the standard deviation for each of the computed features.
- Labelling and Separation: On the basis of these thresholds, we obtain tracings that can be classified as dominantly belonging to one of the six strategy types.
- *t*-Test of significance: A *t*-test determines whether there is a significant difference between the labelled samples and the other samples, as summarized in Table 3

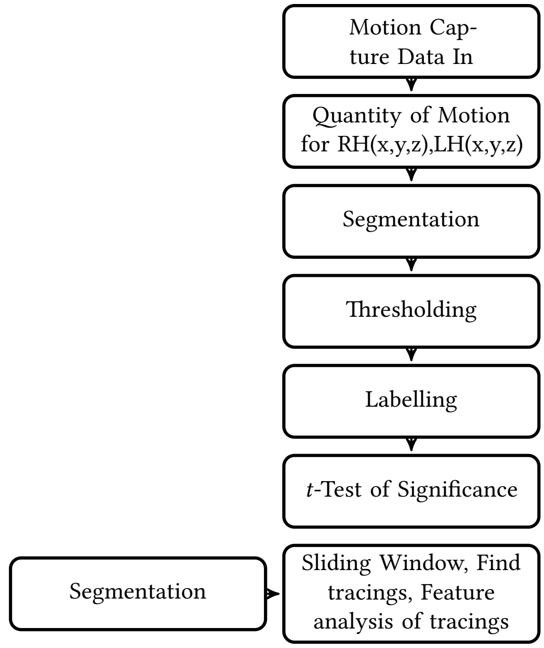


Figure 3: Illustration of the analysis pipeline.

5 DISCUSSION

The quantified features presented here should be seen as a starting point for using numerical methods to investigate motion capture data. We have found a feature set that represents the qualitatively selected strategies, as illustrated in Table 2. The distribution of means and standard deviations for the normalized QoMs for each of the target tracings indicate that not every tracing lies only in one particular strategy. In future research it will be particularly interesting to investigate how people switch between the strategies with regards to the musical content.

#	Strategy	Distinguishing features	Description	Mean	SD
1	Dominant hand as needle	QoM(LHY) != QoM(RHY)	Left hand Y position greater than right Y position	0.500	0.063
2	Changing inter-palm distance	RMS(LHX) - RMS(RHX)	Root mean squared difference of left, right hands in x	0.647	0.120
3	Both hands as mirror images	RHX - LHX = C	Threshold of difference between left and right hands	0.346	0.111
4	Manipulating a small object	RH(x,y,z) = LH(x,y,z) + C	Right and left hands follow similar trajectories in x	0.725	0.072
5	Drawing arcs along circles	x^2+y^2+z^2	Fit of (x,y,z) for left and right hands to a sphere	0.171	0.042
6	Percussive asymmetry	dtw(RH(xyz),LH(xyz))	Dynamic Time Warp of (x,y,z) of Left, Right Hands	0.561	0.075

Table 2: Quantitative motion capture features that match the qualitatively selected strategies.

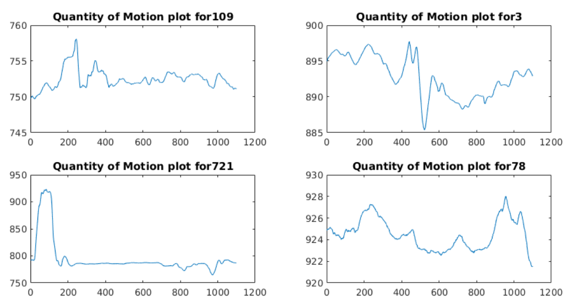


Figure 4: Examples of sound-tracing segments representing Quantity of Motion.

Despite the variations, the selected features demonstrate the dominance of one particular strategy for many tracings. There also seems to be a significant difference between strategies as illustrated by the distribution of means and standard deviations of QoM as described in table 3. All except Feature 4 (manipulating a small "object") show significant results compared to all other tracing samples.

One main challenge when comparing motion capture data with (musical) sound, is the large difference between the feature space of body motion and that of musical (sound) features. Relevant literature, as well as the current experimentation, points to a conceptual mapping between sound and motion. However, this mapping is subject to a lot of variation pertaining to the individual participants and the nature of the musical examples. Previous studies have shown relationships between, for example, peaks of rhythmic motion and beats in the music; or correlations between quantity of motion and musical intensity. For motion to be compared to melodic sound, however, we believe it is more relevant to analyse motion trajectories in relation to melodic contours.

Strategy #	*p*-Value
Feature 1 vs rest	0.003
Feature 2 vs rest	0.011
Feature 3 vs rest	0.005
Feature 4 vs rest	0.487
Feature 5 vs rest	0.003
Feature 6 vs rest	0.006

Table 3: Significance testing for each feature against the rest of the samples

6 FUTURE WORK

A core question for the further exploration of our current (and future) data sets, is: Why do people change representation strategies to adapt to new musical stimuli? We believe that the features developed here may provide some more insight, and be used in comparisons with musical contours (such as pitch and intensity). The mapping of features in melodic contours, and similarities in people's representations of melodic contours through motion, are the two questions we will focus on, in further analysis of this dataset.

REFERENCES

[1] Charles R Adams. 1976. Melodic contour typology. *Ethnomusicology* (1976), 179–215.
[2] Baptiste Caramiaux, Frédéric Bevilacqua, Tommaso Bianco, Norbert Schnell, Olivier Houix, and Patrick Susini. 2014. The role of sound source perception in gestural sound description. *ACM Transactions on Applied Perception (TAP)* 11, 1 (2014), 1.
[3] Joyce L Chen, Virginia B Penhune, and Robert J Zatorre. 2008. Moving on time: brain network for auditory-motor synchronization is modulated by rhythm complexity and musical training. *Journal of Cognitive Neuroscience* 20, 2 (2008), 226–239.
[4] Tuomas Eerola and Micah Bregman. 2007. Melodic and contextual similarity of folk song phrases. *Musicae Scientiae* 11, 1 suppl (2007), 211–233.
[5] Rolf Inge Godøy. 2015. Ubiquitous Motor Cognition in Musical Experience: Open Peer Review of Jacques Launay's" Musical Sounds, Motor Resonance, and Detectable Agency". *Empirical Musicology Review* 10, 1-2 (2015), 41–45.
[6] Rolf Inge Godøy, Minho Song, Kristian Nymoen, Mari Romarheim Haugen, and Alexander Refsum Jensenius. 2016. Exploring Sound-Motion Similarity in Musical Experience. *Journal of New Music Research* 45, 3 (2016), 210–222.
[7] Alexander Refsum Jensenius. 2007. Action-sound: Developing methods and tools to study music-related body movement. (2007).
[8] Alexander Refsum Jensenius. 2009. Musical Gestures. *Musical Gestures: Sound, Movement, and Meaning* 12 (2009).
[9] Marc Leman. 2008. *Embodied music cognition and mediation technology.* MIT Press, Cambridge, Mass.
[10] David McNeill. 1992. *Hand and mind: What gestures reveal about thought.* University of Chicago press.
[11] Franz Mechsner, Dirk Kerzel, GuÈnther Knoblich, and Wolfgang Prinz. 2001. Perceptual basis of bimanual coordination. *Nature* 414, 6859 (2001), 69–73.
[12] Genevieve Noyce. 2013. Quantifying Shapes: Mathematical Techniques for Analysing Visual Representations of Sound and Music. *Empirical Musicology Review* 8, 2 (2013), 128–154.
[13] Kristian Nymoen. 2011. 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.*
[14] Joy E Ollen. 2006. *A criterion-related validity test of selected indicators of musical sophistication using expert ratings.* Ph.D. Dissertation. The Ohio State University.
[15] Jon B Prince, Mark A Schmuckler, and William Forde Thompson. 2009. Cross-modal melodic contour similarity. *Canadian Acoustics* 37, 1 (2009), 35–49.
[16] Ian Quinn. 1999. The combinatorial model of pitch contour. *Music Perception: An Interdisciplinary Journal* 16, 4 (1999), 439–456.
[17] Udit Roy, Tejaswinee Kelkar, and Bipin Indurkhya. 2014. TrAP: An Interactive System to Generate Valid Raga Phrases from Sound-Tracings.. In *NIME.* 243–246.
[18] Mark A Schmuckler. 1999. Testing models of melodic contour similarity. *Music Perception: An Interdisciplinary Journal* 16, 3 (1999), 295–326.