Article

Analyzing Free-Hand Sound-Tracings of Melodic Phrases

Tejaswinee Kelkar * [ID] **and Alexander Refsum Jensenius** [ID]

University of Oslo, Department of Musicology, RITMO Centre for Interdisciplinary Studies in Rhythm, Time and Motion, 0371 Oslo, Norway; a.r.jensenius@imv.uio.no

* Correspondence: tejaswinee.kelkar@imv.uio.no; Tel.: +47-4544-8254

Abstract: In this paper, we report on a free-hand motion capture study in which 32 participants 'traced' 16 melodic vocal phrases with their hands in the air in two experimental conditions. Melodic contours are often thought of as correlated with vertical movement (up and down) in time, and this was also our initial expectation. We did find an arch shape for most of the tracings, although this did not correspond directly to the melodic contours. Furthermore, representation of pitch in the vertical dimension was but one of a diverse range of movement strategies used to trace the melodies. Six different mapping strategies were observed, and these strategies have been quantified and statistically tested. The conclusion is that metaphorical representation is much more common than a 'graph-like' rendering for such a melodic sound-tracing task. Other findings include a clear gender difference for some of the tracing strategies and an unexpected representation of melodies in terms of a small object for some of the Hindustani music examples. The data also show a tendency of participants moving within a shared 'social box'.

---

1. Introduction

How do people think about melodies as shapes? This question comes out of the authors' general interest in understanding more about how spatiotemporal elements influence the cognition of music. When it comes to the topic of melody and shape, these terms often seem to be interwoven. In fact, the Concise Oxford Dictionary of Music defines melody as: "A succession of notes, varying in pitch, which has an organized and recognizable shape." [1]. Here, shape is embedded as a component in the very definition of melody. However, what is meant by the term 'melodic shape', and how can we study such melodic shapes and their typologies?

Some researchers have argued for thinking of free-hand movements to music (or 'air instrument performance' [2]) as visible utterances similar to co-speech gestures [3–5]. From the first author's experience as an improvisational singer, a critical part of learning a new singing culture was the physical representation of melodic content. This physical representation includes bodily posture, gestural vocabulary and the use of the body to communicate sung phrases. In improvised music, this also includes the way in which one uses the hands to guide the music and the expectation of a familiar audience from the performing body. These body movements may refer to spatiotemporal metaphors, quite like the ones used in co-speech gestures.

In their theory of cognitive metaphors, Lakoff and Johnson point out how the metaphors in everyday life represent the structure through which we conceptualize one domain with the representation of another [6]. Zbikowski uses this theory to elaborate how words used to describe pitches in different languages are mapped onto the metaphorical space of the 'vertical dimension' [7]. Descriptions of melodies often use words related to height, for example: a 'high'- or 'low'-pitched

voice, melodies going 'up' and 'down'. Shayan et al. suggest that this mapping might be more strongly present in Western cultures, while the use of other metaphors in other languages, such as thick and thin pitch, might explain pitch using other non-vertical one-dimensional mapping schemata [8]. The vertical metaphor, when tested with longer melodic lines, shows that we respond non-linearly to the vertical metaphors of static and dynamic pitch stimuli [9]. Research in music psychology has investigated both the richness of this vocabulary and its perceptual and metaphorical allusions [10]. However, the idea that the vertical dimension is the most important schema of melodic motion is very persistent [11]. Experimentally, pitch-height correspondences are often elicited by comparing two or three notes at a time. However, when stimuli become more complex, resembling real melodies, the persistence of pitch verticality is less clear. For 'real' melodies, shape descriptions are often used, such as arches, curves and slides [12].

In this paper, we investigate shape descriptions through a sound-tracing approach. This was done by asking people to listen to melodic excerpts and then move their body as if their movement was creating the sound. The aim is to answer the following research questions:

1. How do people present melodic contour as shape?
2. How important is vertical movement in the representation of melodic contour in sound-tracing?
3. Are there any differences in sound-tracings between genders, age groups and levels of musical experience?
4. How can we understand the metaphors in sound-tracings and quantify them from the data obtained?

The paper starts with an overview of related research, before the experiment and various analyses are presented and discussed.

2. Background

Drawing melodies as lines has seemed intuitive across different geographies and time periods, from the Medieval neumes to contemporary graphical scores. Even the description of melodies as lines enumerates some of their key properties: continuity, connectedness and appearance as a shape. Most musical cultures in the world are predominantly melodic, which means that the central melodic line is important for memorability. Melodies display several integral patterns of organization and time-scales, including melodic ornaments, motifs, repeating patterns, themes and variations. These are all devices for organizing melodic patterns and can be found in most musical cultures.

2.1. Melody, Prosody and Memory

Musical melodies may be thought of as closely related to language. For example, prosody, which can also be described as 'speech melody,' is essential for understanding affect in spoken language. Musical and linguistic experiences with melody can often influence one another [13]. Speech melodies and musical melodies are differentiated on the basis of variance of intervals, delineation and discrete pitches as scales [14]. While speech melodies show more diversity in intonation, there is lesser diversity in prosodic contours internally within a language. Analysis of these contours is used for recognition of languages, speakers and dialects in computation [15].

It has been argued that tonality makes musical melodies more memorable than speech melodies [16], while contour makes them more recognizable, especially in unfamiliar musical styles [17]. Dowling et al. suggest that adults use contour to recognize unfamiliar melodies, even when they have been transposed or when intervals are changed [16]. There is also neurological evidence supporting the idea that contour memory is independent of absolute pitch location [18]. Early research in contour memory and recognition demonstrated that acquisition of memory for melodic contour in infants and children precedes memory for intonation [17,19–22]. Melodic contour is also described as a 'coarse-grained' schema that lacks the detail from musical intervals [14].

2.2. Melodic Contour

Contour is often used to refer to sequences of up-down movement in melodies, but there are also several other terms that in different ways touch upon the same idea. Shape, for example, is more generally used for referring to overall melodic phrases. Adams uses the terms contour and shape interchangeably [23] and also adds melodic configuration and outline to the mix of descriptors. Tierney et al. discuss the biological origins and commonalities of melodic shape [24]. They also note the predominance of arch-shaped melodies in music, the long duration of the final notes of phrases, and the biases towards smooth pitch contours. The idea of shapes has also been used to analyze melodies of infants crying [25]. Motif, on the other hand, is often used to refer to a unit of melody that can be operated upon to create melodic variation, such as transposition, inversion, etc. Yet another term is that of melodic chunk, which is sometimes used to refer to the mnemonic properties of melodic units, while museme is used to indicate instantaneous perception. Of all these terms, we will stick with contour for the remainder of this paper.

2.3. Analyzing Melodic Contour

There are numerous analysis methods that can be used to study melodic contour, and they may be briefly divided into two main categories: signal-based or symbolic representations. When the contour analysis uses a signal-based representation, a recording of the audio is analyzed with computational methods to extract the melodic line, as well as other temporal and spectral features of the sound. The symbolic representations may start from notated or transcribed music scores and use the symbolic note representations for analysis. Similar to how we might whistle a short melodic excerpt to refer to a larger musical piece, melodic dictionaries have been created to index musical pieces [26]. Such indexes merit a thorough analysis of contour typologies, and several contour typologies were created to this end [23,27,28]. Contour typology methods are often developed from symbolic representations and notated as discrete pitch items. Adam's method for contour typology analysis, for example, codes the initial and final pitches of the melodies as numbers [23]. Parson's typology, on the other hand, uses note directions and their intervals as the units of analysis [26]. There are also examples of matrix comparison methods that code pitch patterns [27]. A comparison of these methods to perception and memory is carried out in [29,30], suggesting that the information-rich models do better than more simplistic ones. Perceptual responses to melodic contour changes have also been studied systematically [30–32], revealing differences between typologies and which ones come closest to resembling models of human contour perception.

The use of symbolic representations makes it easier to perform systematic analysis and modification of melodic music. While such systematic analysis works well for some types of pre-notated music, it is more challenging for non-notated or non-Western music. For such non-notated musics, the signal-based representations may be a better solution, particularly when it comes to providing representations that more accurately describe continuous contour features. Such continuous representations (as opposed to more discrete, note-based representations) allow the extraction of information from the actual sound signal, giving a generally richer dataset from which to work. The downside, of course, is that signal-based representations tend to be much more complex, and hence more difficult to generalize from.

In the field of music perception and cognition, the use of symbolic music representations, and computer-synthesized melodic stimuli, has been most common. This is the case even though the ecological validity of such stimuli may be questioned. Much of the previous research into the perception of melodic contour also suffers from a lack of representation of non-Western and also non-classical musical styles, with some notable exceptions such as [33–35].

Much of the recent research into melodic representations is found within the music information retrieval community. Here, the extraction of melodic contour and contour patterns directly from the signal is an active research topic, and efficient algorithms for extraction of the primary melody have been tested and compared in the MIREX (Music Information Retrieval Evaluation Exchange) competitions

for several years. Melodic contour is also used to describe the instrumentation of music from audio signals, for example in [36,37]. It is also interesting to note that melodic contour is used as the first step to identify musical structure in styles such as in Indian Classical music [38], and Flamenco [39].

2.4. Pitch and the Vertical Dimension

As described in Section 1, the vertical dimension (up-down) is a common way to describe pitch contours. This cross-modal correspondence has been demonstrated in infants [40], showing preferences for concurrence of auditory pitch 'rising,' visuospatial height, as well as visual 'sharpness'. The association with visuospatial height is elaborated further with the SMARC (Spatial-Musical Association of Response Codes) effect [11]. Here, participants show a shorter response time for lower pitches co-occurring with left or bottom response codes, while higher pitches strongly correlated with response codes for right or top. A large body of work tries to tease apart the nuances of the suggested effect. Some of the suggestions include the general setting of the instruments and the bias of reading and writing being from left to right in most of the participants [41], as contributing factors to the manifestation of this effect.

The concepts of contour rely upon pitch height being a key feature of our melodic multimodal experience. Even the enumeration of pitch in graphical formats plays on this persistent metaphor. Eitan brings out the variety of metaphors for pitch quality descriptions, suggesting that up and down might only be one of the ways in which cross-modal correspondence with pitch appears in different languages [9,10]. Many of the tendencies suggested in the SMARC effect are less pronounced when more, and more complex, stimuli appear. These have been tested in memory or matching tasks, rather than asking people to elicit the perceived contours. The SMARC effect may here be seen in combination with the STEARC (Spatial-Temporal Association of Response Codes) effect, stating that timelines are more often perceived as horizontally-moving objects. In combination, these two effects may explain the overwhelming representation of vertical pitch on timelines. The end result is that we now tend to think of melodic representation mostly in line-graph-based terms, along the lines of what Adams ([23], p. 183) suggested:

> There is a problem of the musical referents of the terms. As metaphoric depictions, most of these terms are more closely related to the visual and graphic representations of music than to its acoustical and auditory characteristics. Indeed, word-list typologies of melodic contour are frequently accompanied by 'explanatory' graphics.

This 'problem' of visual metaphors, however, may actually be seen as an opportunity to explore multimodal perception that was not possible to understand at the time.

2.5. Embodiment and Music

The accompaniment of movement to music is understood now as an important phenomenon in music perception and cognition [42]. Research studying the close relationship between sound and movement has shed light on the mechanism to understand action as sound [43] and sound as action [44,45]. Cross-modal correspondence is a phenomenon with a tight interactive loop with the body as a mediator for perceptual, as well as performative roles [46,47]. Some of these interactions show themselves in the form of motor cortex activation when only listening to music [48]. This has further led to empirical studies of how music and body movement share a common structure that affords equivalent and universal emotional expression [49]. Mazzola et al. have also worked on a topological understanding of musical space and the topological dynamics of musical gesture [50].

Studies of Hindustani music show that singers use a wide variety of movements and gestures that accompany spontaneous improvisation [4,51,52]. These movements are culturally codified; they appear in the performance space to aid improvisation and musical thought, and they also convey this information to the listener. The performers also demonstrate a variety of imaginary 'objects' with various physical properties to illustrate their musical thought.

Some other examples of research on body movement and melody include Huron's studies of how eyebrow height accompany singing as a cue response to melodic height [53], and studies suggesting that especially arch-shaped melodies have common biological origins that are related to motor constraints [24,54].

2.6. Sound-Tracings

Sound-tracing studies aim at analyzing spontaneous rendering of melodies to movement, capturing instantaneous multimodal associations of the participants. Typically, subjects are asked to draw (or trace) a sound example or short musical excerpt as they are listening. Several of these studies have been carried out with digital tablets as the transducer or the medium [2,44,55–59]. One restriction of using tablets is that the canvas of the rendering space is very restrictive. Furthermore, the dimensionality does not evolve over time and represents a narrow bandwidth of possible movements.

An alternative to tablet-based sound-tracing setups is that of using full-body motion capture. This may be seen as a variation of 'air performance' studies, in which participants try to imitate the sound-producing actions of the music to which they listen [2]. Nymoen et al. carried out a series of sound-tracing studies focusing on movements of the hands [60,61], elaborating on several feature extraction methods to be used for sound-tracing as a methodology. The current paper is inspired by these studies, but extending the methodology to full-body motion capture.

3. Sound-Tracing of Melodic Phrases

3.1. Stimuli

Based on the above considerations and motivations, we designed a sound-tracing study of melodic phrases. We decided to use melodic phrases from vocal genres that have a tradition of singing without words. Vocal phrases without words were chosen so as to not introduce lexical meaning as a confounding variable. Leaving out instruments also avoids the problem of subjects having to choose between different musical layers in their sound-tracing.

The final stimulus set consists of four different musical genres and four stimuli for each genre. The musical genres selected are: (1) Hindustani (North Indian) music, (2) Sami joik, (3) scat singing, (4) Western classical vocalize. The melodic fragments are phrases taken from real recordings, to retain melodies within their original musical context. As can be seen in the pitch plots in Figure 1, the melodies are of varying durations with an average of 4.5 s (SD = 1.5 s). The Hindustani and joik phrases are sung by male vocalists, whereas the scat and vocalize phrases are sung by female vocalists. This is represented in the pitch range of each phrase as seen in Figure 1. The Hindustani and joik melodies are mainly sung in a strong chest voice in this stimulus set. Scat vocals are sung with a transition voice from chest to head. The Vocalizes in this set are sung by a soprano, predominantly in the head register. Hindustani and vocalize samples have one dominant vowel that is used throughout the phrase. The Scat singing examples use traditional 'shoobi-doo-wop' syllables, and joik examples in this set predominantly contain syllables such as 'la-la-lo'.

To investigate the effects of timbre, we decided to create a 'clean' melody representation of each fragment. This was done by running the sound files through an autocorrelation algorithm to create phrases that accurately resemble the pitch content, but without the vocal, timbral and vowel content of the melodic stimulus. These 16 re-synthesized sounds were added to the stimulus set, thus obtaining a total of 32 sound stimuli (summarized in Table 1).
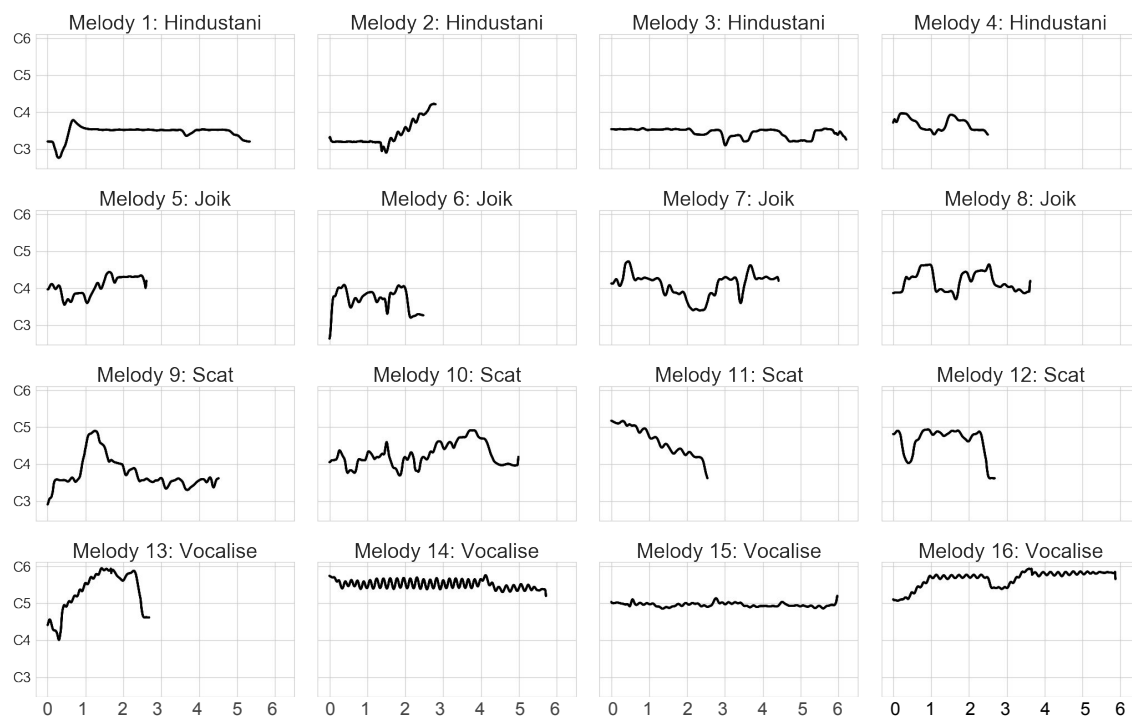
Figure 1. Pitch plots of all the 16 melodic phrases used as experiment stimuli, from each genre. The x axis represents time in seconds, and the y axis represents notes. The extracted pitches were re-synthesized to create a total of 32 melodic phrases used in the experiment.

Table 1. An overview of the 32 different stimuli: four phrases from each musical genre, all of which were presented in both normal and re-synthesized versions.

Type	Hindustani	Joik	Scat	Vocalize
Normal	4	4	4	4
Re-synthesized	4	4	4	4

3.2. Subjects

A total of 32 subjects (17 female, 15 male) was recruited, with a mean age of 31 years (SD = 9 years). The participants were mainly university students and employees, both with and without musical training. Their musical experience was quantized using the OMSI (Ollen Musical Sophistication Index) questionnaire [62], and they were also asked about the familiarity with the musical genres, and their experience with dancing. The mean of the OMSI score was 694 (SD = 292), indicating that the general musical proficiency in this dataset was on the higher side. The average familiarity with Western classical music was 4.03 out of a possible 5 points, 3.25 for jazz music, 1.87 with joik, and 1.71 with Indian classical music. Thus, two genres (vocalize and scat) were more familiar than the two others (Hindustani and joik). All participants provided their written consent for inclusion before they participated in the study, and they were free to withdraw during the experiment. The study obtained ethical approval from the Norwegian Centre for Research Data (NSD), with the project code 49258 (approved on 22 August 2016).

3.3. Procedure

Each subject performed the experiment alone, and the total duration was around 10 min. They were instructed to move their hands as if their movement was creating the melody. The use of the term creating, instead of representing, is purposeful, as in earlier studies [60,63], to avoid the act of

playing or singing. The subjects could freely stand anywhere in the room and face whichever direction they liked, although nearly all of them faced the speakers and chose to stand in the center of the lab. The room lighting was dimmed to help the subjects feel more comfortable to move as they pleased.

The sounds were played at a comfortable listening level through a Genelec 8020 speaker, placed 3 m in front of the subjects. Each session consisted of an introduction, two example sequences, 32 trials and a conclusion, as sketched in Figure 2. Each melody was played twice with a 2-s pause in between. During the first presentation, the participants were asked to listen to the stimuli, while during the second presentation, they were asked to trace the melody. A long beep indicated the first presentation of a stimulus, while a short beep indicated the repetition of a stimuli. All the instructions and required guidelines were recorded and played back through the speaker to not interrupt the flow of the experiment.
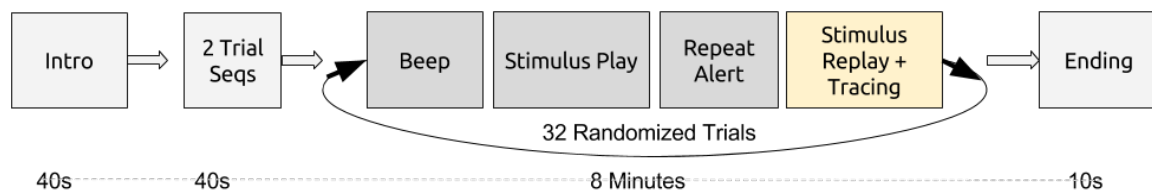

.

Figure 2. The experiment flow, with an approximate total duration of 10 min

The subjects' motion was recorded using an infrared marker-based motion capture system from Qualisys AB (Gothenburg, Sweden), with 8 Oqus 300 cameras surrounding the space (Figure 3a) and one regular video camera (Canon XF105 (manufactured in Tokyo, Japan)), for reference. Each subject wore a motion capture suit with 21 reflective markers on each joint (Figure 3b). The system captured at a rate of 200 Hz. The data were post-processed in the Qualisys Track Manager software (QTM, v2.16, Qualisys AB, Gothenburg, Sweden), which included labeling of markers and removal of ghost-markers (Figure 3c). We used polynomial interpolation to gap-fill the marker data, where needed. The post-processed data was exported to Python (v2.7.12 and MATLAB (R2013b, MathWorks, Natick, MA., USA) for further analysis. Here, all of the 10-min recordings were segmented using automatic windowing, and each of the segments were manually annotated for further analysis in Section 6.
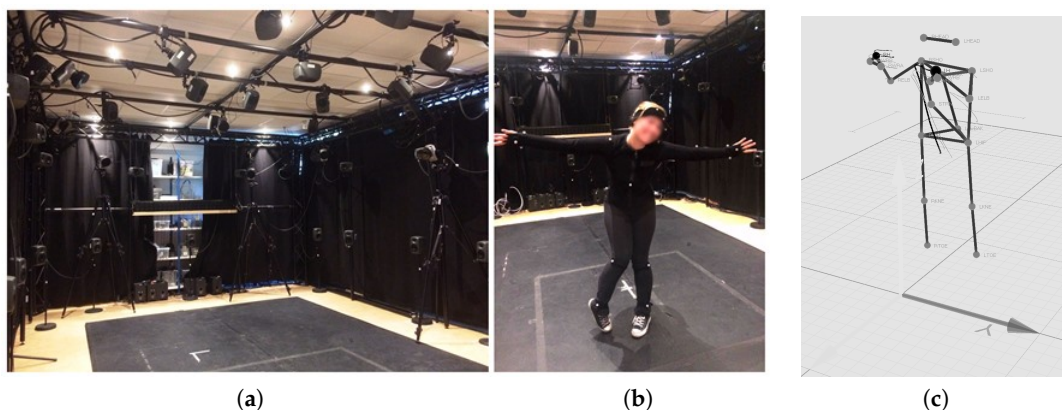

Figure 3. (**a**) The motion capture lab used for the experiments. (**b**) The subjects wore a motion capture suit with 21 reflective markers. (**c**) Screenshot of a stick-figure after post-processing in Qualisys Track Manager software (QTM).

4. Analysis

Even though full-body motion capture was performed, we will in the following analysis only consider data from the right and left hand markers. Marker occlusions from six of the subjects were difficult to account for in the manual annotation process, so only data from 26 participants were used in the analysis that will be presented in Section 5. This analysis is done using comparisons of means and distribution patterns. The occlusion problems were easier to tackle with the automatic analysis, so the analysis that will be presented in Section 6 was performed on data from all 32 participants.

4.1. Feature Selection from Motion Capture Data

Table 2 shows a summary of the features extracted from the motion capture data. Vertical velocity is calculated as the first derivative of the z-axis (vertical motion) for each tracing over time. 'Quantity of motion' is a dimensionless quantity representing the overall amount of motion in any direction from frame to frame. Hand distance is calculated as the euclidean distance between the x,y,z coordinates for each marker for each hand. We also calculate the sample-wise distance traveled for each hand marker.

Table 2. The features extracted from the motion capture data.

	Motion Features	Description
1	VerticalMotion	z-axis coordinates at each instant of each hand
2	Range	(Min, Max) tuple for each hand
3	HandDistance	The euclidean distance between the 2d coordinates of each hand
4	QuantityofMotion	The sum of absolute velocities of all the markers
5	DistanceTraveled	Cumulative euclidean distance traveled by each hand per sample
6	AbsoluteVelocity	Uniform linear velocity of all dimensions
7	AbsoluteAcceleration	The derivative of the absolute velocity
8	Smoothness	The number of knots of a quadratic spline interpolation fitted to each tracing
9	VerticalVelocity	The first derivative of the z-axis in each participant's tracing
10	CubicSpline10Knots	10 knots fitted to a quadratic spline for each tracing

4.2. Feature Selection from Melodic Phrases.

Pitch curves from the melodic phrases were calculated using the autocorrelation algorithm in Praat (v6.0.30, University of Amsterdam, The Netherlands), eliminating octave jumps. These pitch curves were then exported for further analysis together with the motion capture features in Python. Based on contour analysis from the literature [23,30,64], we extracted three different melodic features, as summarized in Table 3.

Table 3. The features extracted from the melodic phrases.

	Melodic Features	Description
1	SignedIntervalDirection	Interval directions (up/down) calculated for each note
2	InitialFinalHighestLowest	Four essential notes of a melody: initial, final, highest, lowest
3	SignedRelativeDistances	Feature 1 combined with relative distances of each successive note from the next, only considering the number of semitones for each successive change.

5. Analysis of Overall Trends

5.1. General Motion Contours

One global finding from the sound-tracing data is that of a clear arch shape when looking at the vertical motion component over time. Figure 4 shows the average contours calculated from the z-values of the motion capture data of all subjects for each of the melodic phrases. It is interesting to note a clear arch-like shape for all of the graphs. This fits with suggestions of a motor constraint

hypothesis suggesting that melodic phrases in general have arch-like shapes [24,54]. In our study, however, the melodies have several different types of shapes (Figure 1), so this may not be the best explanation for the arch-like motion shapes. A better explanation may be that the subjects would start and end their tracing from a resting position in which the hands would be lower than during the tracing, thus leading to the arch shapes seen in Figure 4.
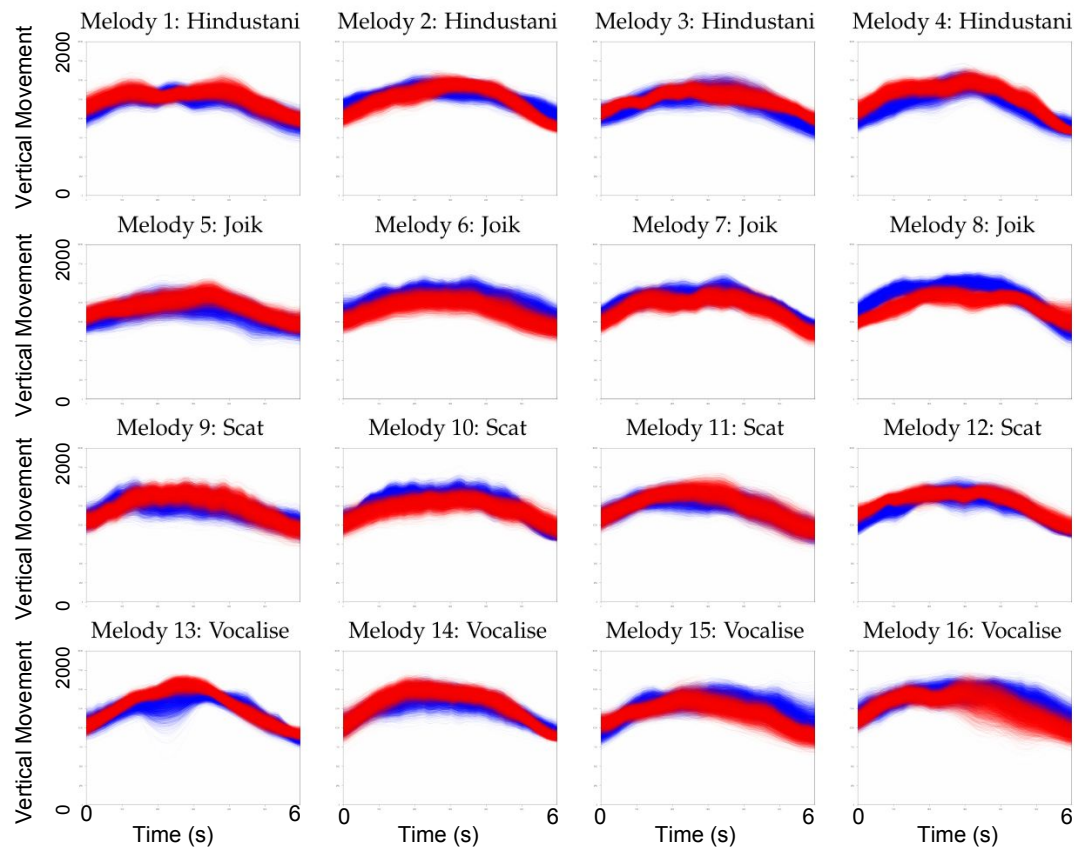

Figure 4. Average contours plotted from the vertical motion capture data in mm (z-axis) for each of the melodies (red for the original and blue for the re-synthesized versions of the melodies). The x-axis represents normalized time, and the y-axis represents aggregated tracing height for all participants.

5.2. Relationship between Vertical Movement and Melodic Pitch Distribution

To investigate more closely the relationship between vertical movement and the distribution of the pitches in the melodic fragments, we may start by considering the genre differences. Figure 5 presents the distribution of pitches in each genre in the stimulus set. These are plotted on a logarithmic frequency scale to represent the perceptual relationships between them. In the plot, each of the four genres are represented by their individual distributions. The color distinction is on the basis of whether the melodic phrase has one direction or many. Phrases closer to being ascending, descending, or stationary are coded as not changing direction. We see that in all of these conditions, the vocalize phrases in the dataset have the highest pitch profiles and the Hindustani phrases have the lowest.

If we then turn to look at the vertical dimension of the tracing data, we would expect to see a similar distribution between genres as that for the pitch heights of the melodies. Figure 6 shows the distribution of motion capture data for all tracings, sorted in the four genres. Here, the distribution of maximum z-values of the motion capture data shows a quite even distribution between genres.
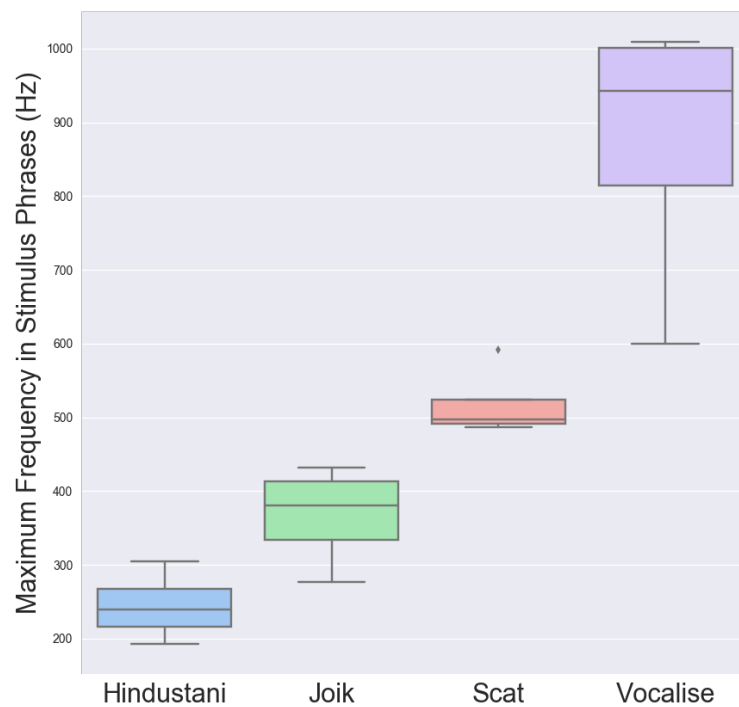
Figure 5. Pitch distribution for each genre based on mean pitches in each phrase. If movement tracings were an accurate representation of absolute pitch height, movement plots should resemble this plot.
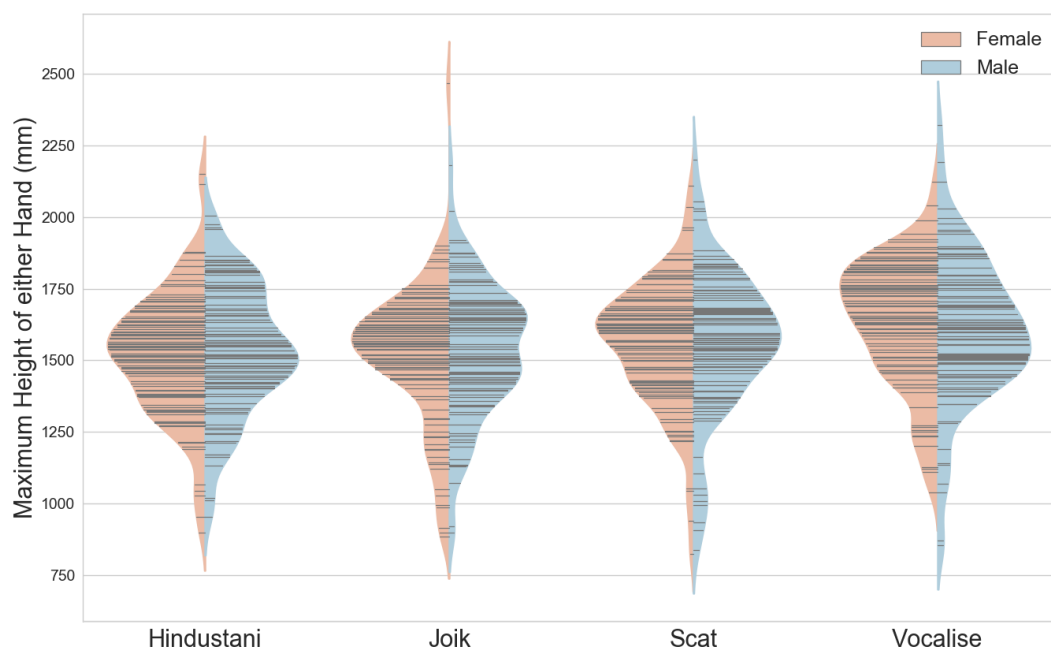

Figure 6. Plots of the maximum height of either hand for each genre. The left/pink distributions are from female subjects, while the right/blue distributions are from the male subjects. Each half of each section of the violin plot represents the probability distribution of the samples. The black lines represent each individual data point.

5.3. Direction Differences

The direction differences in the tracings can be studied by calculating the coefficients of variation of movement in all three axes for both the left (LH) and right (RH) hands. These coefficients are

found to be LHvar (x,y,z) = (63.7,45.7,26.6); RHvar (x,y,z) = (56.6,87.8,23.1), suggesting that the amount of dispersion on the z-axis (the vertical) is the most consistent. This suggests that a wide array of representations in the x and y-axes are used.

The average standard deviations for the different dimensions were found to be LHstd = (99 mm, 89 mm, 185 mm); RHstd = (110 mm, 102 mm, 257 mm). This means that most variation is found in the vertical movement of the right hand, indicating an effect of right-handedness among the subjects.

5.4. Individual Subject Differences

Plots of the distributions of the quantitiy of motion (QoM) for each subject for all stimuli show a large degree of variation (Figure 7). Subjects 4 and 12, for example, have very small diversity in the average QoM for all of their tracings, whereas Subjects 2 and 5 show a large spread. There are also other participants, such as 22, who move very little on average for all their tracings. Out of the two types of representations (original and re-synthesized stimuli), we see that there is in general a larger diversity of movement for the regular melodies as opposed to the synthesized ones. However, the statistical difference between synthesized and original melodies was not significant.
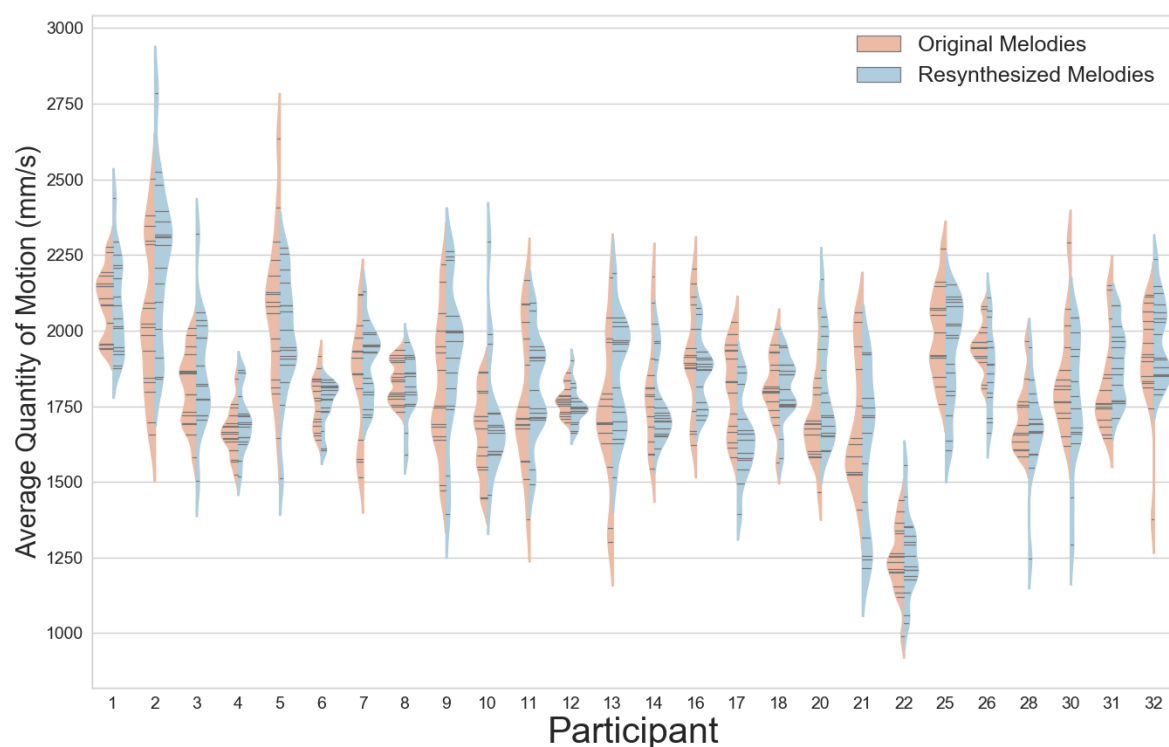

Figure 7. Distribution of the average quantity of motion for each participant. Left/red distributions are of the synthesized stimuli, while the right/green distributions are of the normal recordings.

5.5. Social Box

Another general finding from the data that is not directly related to the question at hand, but that is still relevant for understanding the distribution of data, is what we call a shared 'social box' among the subjects. Figure 6 shows that the maximum tracing height of the male subjects were higher than those of the female subjects. This is as expected, but a plot of the 'tracing volume' (the spatial distribution in all dimensions) reveals that a comparably small volume was used to represent most of the melodies (Figure 8). Qualitative observation of the recordings reveal that shorter subjects were more comfortable about stretching their hands out, while taller participants tended to use a more restrictive space relative to their height. This happened without there being any physical constraints of

their movements, and no instructions that had pointed in the direction of the volume to be covered by the tracings.
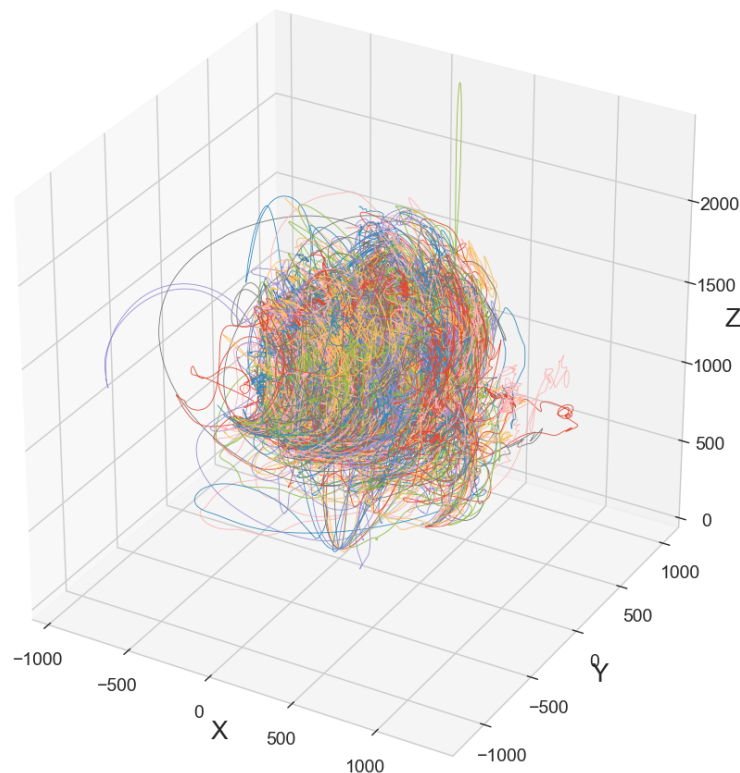

Figure 8. A three-dimensional plot of all sound-tracings for all participants reveal a fairly constrained tracing volume, a kind of 'social box' defined by the subjects. Each color represents the tracings of a single participant, and numbers along each axis are milimetres.

It is almost as if the participants wanted to fill up an invisible 'social box,' serving as the collective canvas on which everything can be represented. This may be explained by the fact that we share a world together that has the same dimensions: doors, cars, ceilings, and walls are common to us all, making us understand our body as a part of the world in a particular way. In the data from this experiment, we explore this by analyzing the range of motion relative to the heights of the participants through linear regression. The scaled movement range in the horizontal plane is represented in Figure 9. and shows that the scaled range reduces steadily over time as the height of the participants increases. Shorter participants occupy a larger area in the horizontal plane, while taller participants occupy a relatively smaller area. The R^2 coefficient of regression is found to be 0.993, meaning that this effect is significant.
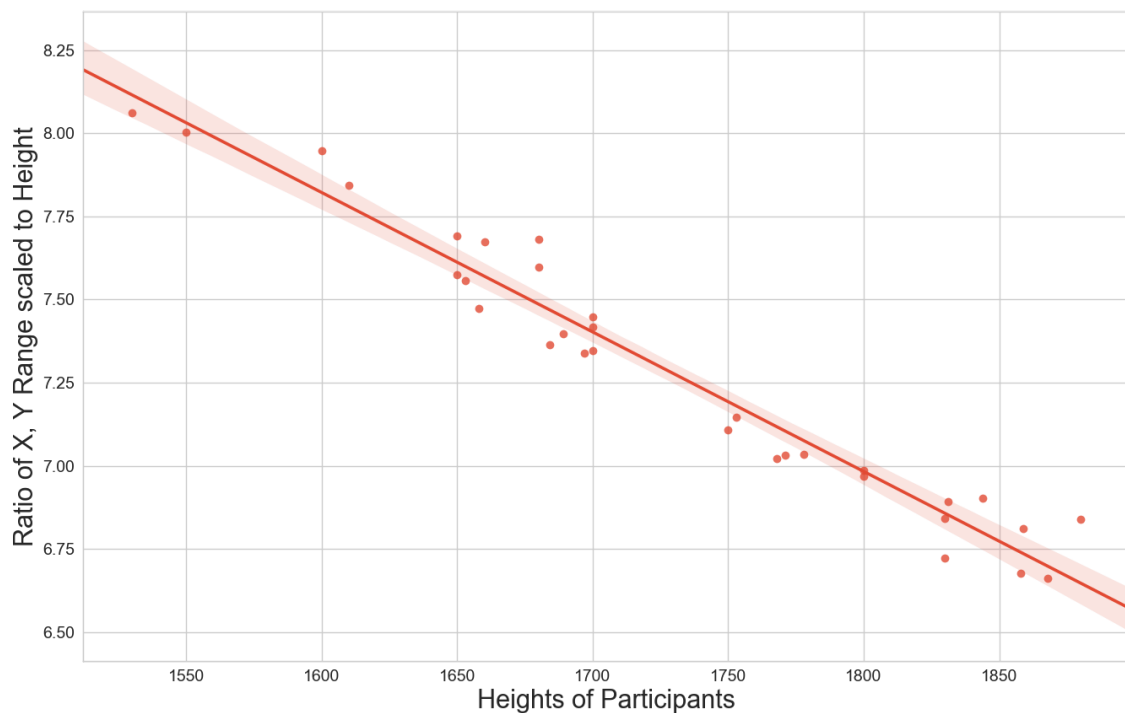
Figure 9. Regression plot of the heights of the participants against scaled (x, y) ranges. There is a clearly decreasing trend for the scaled range of movements in the horizontal plane. The taller a participant, the lower is their scaled range.

5.6. An Imagined Canvas

In a two-dimensional tracing task, such as with pen on paper, the 'canvas' of the tracing is both finite and visible all the time. Such a canvas is experienced also for tasks performed with a graphical tablet, even if the line of the tracing is not visible. In the current experiment, however, the canvas is three-dimensional and invisible, and it has to be imagined by the participant. Participants who trace by just moving one hand at a time seem to be using the metaphor of a needle sketching on a moving paper, much like an analogue ECG (Electro CardioGram) machine. Depending on the size of the tracing, the participants would have to rotate or translate their bodies to move within this imagined canvas. We observe different strategies when it comes to how they reach beyond the constraints of their kinesphere, the maximum volume you can reach without moving to a new location. They may step sideways, representing a flat canvas placed before them, or may rotate, representing a cylindrical canvas around their bodies.

6. Mapping Strategies

Through visual inspection of the recordings, we identify a limited number of strategies used in the sound-tracings. We therefore propose six schemes of representation that encompass most of the variation seen in the hands' movement, as illustrated in Figure 10 and summarized 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 center of the body in the horizontal plane
4. Two hands moving together to represent holding and manipulating an object
5. Two hands drawing arcs along an imaginary circle
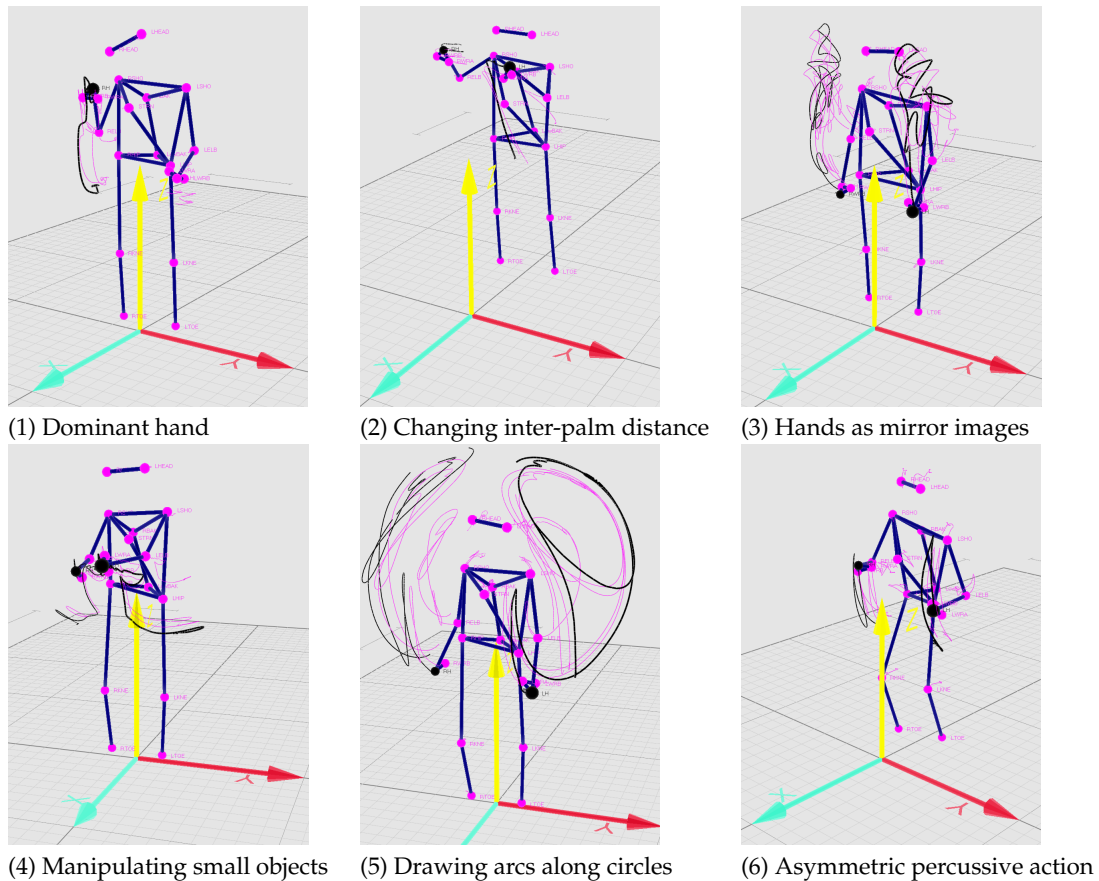6. Two hands following each other in a percussive pattern

(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 10. Motion history images exemplifying the six dominant sound-tracing strategies. The black lines from the hands of the stick figures indicate the motion traces of each tracing.

These qualitatively derived strategies were the starting point for setting up an automatic extraction of features from the motion capture data. The pipeline for this quantitative analysis consists of the following steps:

1. Feature selection: Segment the motion capture data into a six-column feature vector containing the (x,y,z) coordinates of the right palm and the left palm, respectively.

2. Calculate quantity of motion (QoM): Calculate the average of the vector magnitude for each sample.

3. Segmentation: Trim data using a sliding window of 1100 samples in size. This corresponds to 5.5 s, to accommodate the average duration of 4.5 s of the melodic phrases. The hop size for the windows is 10 samples, to obtain a large set of windowed segments. The segments that have the maximum mean values are then separated out to get a set of sound-tracings.

4. Feature analysis: Calculate features from Table 4 for each segment.

5. Thresholding: Minimize the six numerical criteria by thresholding the segments based on two-times the standard deviation for each of the computed features.

6. Labeling and separation: Obtain tracings that can be classified as dominantly belonging to one of the six strategy types.

Table 4. Quantitative motion capture features that match the qualitatively selected strategies. QoM, quantities of motion.

#	Strategy	Distinguishing Features	Description	Mean	SD
1	Dominant hand as needle	Right hand QoM much greater than left QoM	$QoM(LHY) \gg \vee \ll QoM(RHY)$	0.50	0.06
2	Changing inter-palm distance	Root mean squared difference of left, right hands in x	RMS(LHX) $-$ RMS(RHX)	0.64	0.12
3	Lateral symmetry between hands	Nearly constant difference between left and right hands	RHX $-$ LHX = C	0.34	0.11
4	Manipulating a small object	Right and left hands follow similar trajectories in x	RH(x,y,z) = LH(x,y,z) + C	0.72	0.07
5	Drawing arcs along circles	Fit of (x,y,z) for left and right hands to a sphere	$x^2 + y^2 + z^2$	0.17	0.04
6	Percussive asymmetry	Dynamic time warp of (x,y,z) of Left, Right Hands	dtw(RH(xyz), LH(xyz))	0.56	0.07

After running the segmentation and labeling on the complete data set, we performed a *t*-test to determine whether there is a significant difference between the labeled samples and the other samples. The results, summarized in Table 5 show that the selected features demonstrate the dominance of one particular strategy for many tracings. All except Feature 4 (manipulating a small 'object') show significant results compared to all other tracing samples for automatic annotation of hand strategies. While this feature cannot be extracted from the aforementioned heuristic, the simple feature for euclidean distance between two hands proves effective to be able to explain this strategy.

Table 5. Significance testing for each feature against the rest of the features.

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

In Figure 11, we see that hand distance might be an effective way to compare different hand strategies. Strategy 2 performs the best on testing for separability. The hand distance for Strategy 4, for example is significantly lower than the rest. This is because this tracing style represents people who use the metaphor of an imaginary object to represent music. This imaginary object seldom changes its physical properties—its length and breadth and general size is usually maintained.

Taking demographics into account, we see that the distribution of the female subjects' tracings for vocalizes have a much wider peak than the rest of the genres. In the use of hand strategies, we observe that women use a wider range of hand strategies as compared to men (Figure 11). Furthermore, Strategy 5 (drawing arcs) is done entirely by women. The representation of music as objects is also seen to be more prominent in women, as is the use of asymmetrical percussive motion. Comparing the same distribution of genders for genres, we do not find a difference in overall movement or general body use between the genders. If anything, the 'social box effect' makes the height differences of genres smaller than they are.

In Figure 12, we visualize the use of these hand strategies for every melody by all the participants. Strategy 2 is used in 206 tracings, whereas Strategy 5 is used for only 8 tracings. Strategies 1, 3, 4 and 5 are used 182, 180, 161 and 57 times,respectively. Through this heat map in 12, we also can find some outliers for the strategies that are more infrequently used. For example, we see that Melodies 4, 13 and 16 show specially dominant use of some hand strategies.
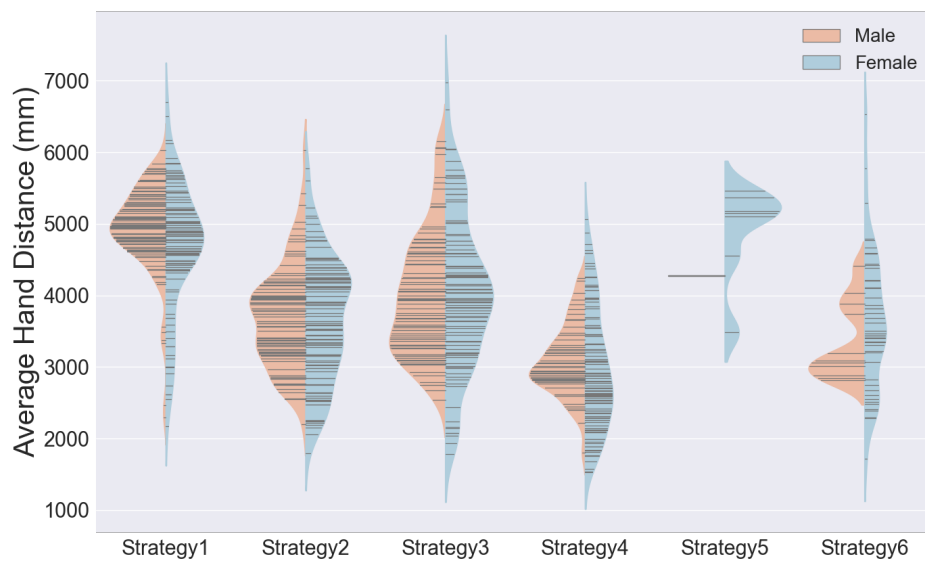
Figure 11. Hand distance as a feature to discriminate between tracing strategies.

Figure 12. Heat map of representation of hand strategy per melody.

7. Discussion

In this study, we have analyzed people's tracings to melodies from four musical genres. Although much of the literature points to correlations between melodic pitch and vertical movement, our findings show a much more complex picture. For example, relative pitch height appears to be much more important than absolute pitch height. People seem to think about vocal melodies as actions, rather than interpreting the pitches purely in one dimension (vertical) over time. The analysis of contour features from the literature shows that while tracing melodies through an allocentric representation of the

listening body, the notions of pitch height representations matter much less than previously thought. Therefore contour features cannot be extracted merely by cross-modal comparisons of two data sets. We propose that other strategies can be used for contour representations, but this is something that will have to be developed more in future research.

According to the gestural affordances of musical sound theory [65], several gestural representations can exist for the same sound, but there is a limit to how much they can be manipulated. Our data support this idea of a number of possible and overlapping action strategies. Several spatial and visual metaphors are used by a wide range of people. One interesting finding is that there are gender differences between the representations of the different sound-tracing strategies. Women seem to show a greater diversity of strategies in general, and they also use object-like representations more often than men.

We expected musical genre to have some impact on the results. For example, given that Western vocalizes are sung with a pitch range higher than the rest of the genres in this dataset (Figure 5), it is interesting to note that, on average, people do represent vocalize tracings spatially higher than the rest of the genres. We also found that the melodies with the maximum amount of vibrato (melodies 14 and 16 in Figure 5) are represented with the largest changes of acceleration in the motion capture data. This implies that although the pitch deviation in this case is not so significant, the perception of a moving melody is much stronger by comparison to other melodies that have larger changes in pitch. It could be argued that both melody 4 and 16 contain motivic repetition that cause this pattern. However, repeating motifs are as much parts of melodies 6 and 8 (joik). The values represented by these melodies are applicable to their tracings as original as well as synthesized phrases. The effect of the vowels used in these melodies can also thus be negated. As seen in Figure 12, there are some melodies that stand out for some hand strategies. Melody 4 (Hindustani) is curiously highly represented as a small object. Melody 12 is overwhelmingly represented by symmetrical movements of both hands, while Melodies 8 and 9 are overwhelmingly represented by using 1 hand as the tracing needle.

We find it particularly interesting that subjects picked up on the idea of using small objects as a representation technique in their tracings. The use of small objects to represent melodies is well documented in Hindustani music [52,66–68]. However, the subjects' familiarity score with Hindustani music was quite low, so familiarity can not explain this interesting choice of representation in our study. Looking at the melodic features of melody 4, for example, it is steadily descending in intervals until it ascends again and comes down the same intervals. This may be argued to resemble an object that smoothly slips on a slope, and could be a probable reason for the overwhelming object representation of this particular melody. In future studies, it would be interesting to see whether we can recreate this mapping in other melodies, or model melodies in terms of naturally occurring melodic shapes born out of physical forces interacting with each other.

It is worth noting that there are several limitations with the current experimental methodology and analysis. Any laboratory study of this kind would present subjects with an unfamiliar and non-ecological environment. The results would also to a large extent be influenced by the habitus of body use in general. Experience in dance, sign language, or musical traditions with simultaneous musical gestures (such as conducting), all play a part in the interpretation of music as motion. Despite these limitations, we do see a considerable amount of consistency between subjects.

8. Conclusions

The present study shows that there are consistencies in people's sound-tracing to the melodic phrases used in the experiment. Our main findings can be summarized as:

- There is a clear arch shape when looking at the averages of the motion capture data, regardless of the general shape of the melody itself. This may support the idea of a motor constraint hypothesis that has been used to explain the similar arch-like shape of sung melodies.
- The subjects chose between different strategies in their sound-tracings. We have qualitatively identified six such strategies and have created a set of heuristics to quantify and test their reliability.

- There is a clear gender difference for some of the strategies. This was most evident for Strategy 4 (representing small objects), which women performed more than men.
- The 'obscure' strategy of representing melodies in terms of a small object, as is typical in Hindustani music, was also found in participants who had no or little exposure to this musical genre.
- The data show a tendency of moving within a shared 'social box'. This may be thought of as an invisible space that people constrain their movements to, even without any exposure to the other participants' tracings. In future studies, it would be interesting to explore how constant such a space is, for example by comparing multiple recordings of the same participants over a period of time.

In future studies, we want to investigate all of these findings in greater detail. We are particularly interested in taking the rest of the body's motion into account. It would also be relevant to use the results from such studies in the creation of interactive systems, 'reversing' the process, that is, using tracing in the air as a method to retrieve melodies from a database. This could open up some exciting end-user applications and also be used as a tool for music performance.

Author Contributions: T.K. and A.R.J. both conceived and designed the experiments; T.K. performed the experiments and analyzed the data; both authors contributed analysis tools, and wrote the paper.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Kennedy, M. Rutherford-Johnson, T.; Kennedy, J. *The Oxford Dictionary of Music*, 6th ed.; Oxford University Press: Oxford, UK, 2013. Available online: http://www.oxfordreference.com/view/10.1093/acref/9780199578108.001.0001/acref-9780199578108 (accessed on 16 January 2018).
2. Godøy, R.I.; Haga, E.; Jensenius, A.R. *Playing "Air Instruments": Mimicry of Sound-Producing Gestures by Novices and Experts*; International Gesture Workshop; Springer: Berlin/Heidelberg, Germany, 2005; pp. 256–267.
3. Kendon, A. *Gesture: Visible Action as Utterance*; Cambridge University Press: Cambridge, UK, 2004.
4. Clayton, M.; Sager, R.; Will, U. In time with the music: The concept of entrainment and its significance for ethnomusicology. In *European Meetings in Ethnomusicology*; ESEM Counterpoint 1; Romanian Society for Ethnomusicology: Bucharest, Romania, 2005; Volume 11, pp. 1–82.
5. Zbikowski, L.M. Musical gesture and musical grammar: A cognitive approach. In *New Perspectives on Music and Gesture*; Ashgate Publishing Ltd.: Farnham, UK, 2011; pp. 83–98.
6. Lakoff, G.; Johnson, M. Conceptual metaphor in everyday language. *J. Philos.* **1980**, *77*, 453–486.
7. Zbikowski, L.M. Metaphor and music theory: Reflections from cognitive science. *Music Theory Online* **1998**, *4*, 1–8.
8. Shayan, S.; Ozturk, O.; Sicoli, M.A. The thickness of pitch: Crossmodal metaphors in Farsi, Turkish, and Zapotec. *Sens. Soc.* **2011**, *6*, 96–105.
9. Eitan, Z.; Schupak, A.; Gotler, A.; Marks, L.E. Lower pitch is larger, yet falling pitches shrink. *Exp. Psychol.* **2014**, *61*, 273–284.
10. 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* **2010**, *114*, 405–422.
11. Rusconi, E.; Kwan, B.; Giordano, B.L.; Umilta, C.; Butterworth, B. Spatial representation of pitch height: The SMARC effect. *Cognition* **2006**, *99*, 113–129.
12. Huron, D. The melodic arch in Western folksongs. *Comput. Musicol.* **1996**, *10*, 3–23.
13. Fedorenko, E.; Patel, A.; Casasanto, D.; Winawer, J.; Gibson, E. Structural integration in language and music: Evidence for a shared system. *Mem. Cognit.* **2009**, *37*, 1–9.
14. Patel, A.D. *Music, Language, and the Brain*; Oxford University Press: New York, NY, USA, 2010.

15. Adami, A.G.; Mihaescu, R.; Reynolds, D.A.; Godfrey, J.J. Modeling prosodic dynamics for speaker recognition. In Proceedings of the 2003 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP'03), Hong Kong, China, 6–10 April 2003; Volume 4.

16. Dowling, W.J. Scale and contour: Two components of a theory of memory for melodies. *Psychol. Rev.* **1978**, *85*, 341–354.

17. Dowling, W.J. Recognition of melodic transformations: Inversion, retrograde, and retrograde inversion. *Percept. Psychophys.* **1972**, *12*, 417–421.

18. Schindler, A.; Herdener, M.; Bartels, A. Coding of melodic gestalt in human auditory cortex. *Cereb. Cortex* **2012**, *23*, 2987–2993.

19. Trehub, S.E.; Becker, J.; Morley, I. Cross-cultural perspectives on music and musicality. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **2015**, *370*, 20140096.

20. Trehub, S.E.; Bull, D.; Thorpe, L.A. Infants' perception of melodies: The role of melodic contour. *Child Dev.* **1984**, *55*, 821–830.

21. Trehub, S.E.; Thorpe, L.A.; Morrongiello, B.A. Infants' perception of melodies: Changes in a single tone. *Infant Behav. Dev.* **1985**, *8*, 213–223.

22. Morrongiello, B.A.; Trehub, S.E.; Thorpe, L.A.; Capodilupo, S. Children's perception of melodies: The role of contour, frequency, and rate of presentation. *J. Exp. Child Psychol.* **1985**, *40*, 279–292.

23. Adams, C.R. Melodic contour typology. *Ethnomusicology* **1976**, *20*, 179–215.

24. Tierney, A.T.; Russo, F.A.; Patel, A.D. The motor origins of human and avian song structure. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 15510–15515.

25. Díaz, M.A.R.; García, C.A.R.; Robles, L.C.A.; Altamirano, J.E.X.; Mendoza, A.V. Automatic infant cry analysis for the identification of qualitative features to help opportune diagnosis. *Biomed. Signal Process. Control* **2012**, *7*, 43–49.

26. Parsons, D. *The Directory of Tunes and Musical Themes*; S. Brown: Cambridge, UK, 1975.

27. Quinn, I. The combinatorial model of pitch contour. *Music Percept. Interdiscip. J.* **1999**, *16*, 439–456.

28. Narmour, E. *The Analysis and Cognition of Melodic Complexity: The Implication-Realization Model*; University of Chicago Press: Chicago, IL, USA, 1992.

29. Marvin, E.W. A Generalized Theory of Musical Contour: Its Application to Melodic and Rhythmic Analysis of Non-Tonal Music and Its Perceptual and Pedagogical Implications. Ph.D. Thesis, University of Rochester, Rocheste, NY, USA, 1988.

30. Schmuckler, M.A. Testing models of melodic contour similarity. *Music Percept. Interdiscip. J.* **1999**, *16*, 295–326.

31. Schmuckler, M.A. Melodic contour similarity using folk melodies. *Music Percept. Interdiscip. J.* **2010**, *28*, 169–194.

32. Schmuckler, M.A. Expectation in music: Investigation of melodic and harmonic processes. *Music Percept. Interdiscip. J.* **1989**, *7*, 109–149.

33. Eerola, T.; Himberg, T.; Toiviainen, P.; Louhivuori, J. Perceived complexity of western and African folk melodies by western and African listeners. *Psychol. Music* **2006**, *34*, 337–371.

34. Eerola, T.; Bregman, M. Melodic and contextual similarity of folk song phrases. *Musicae Sci.* **2007**, *11*, 211–233.

35. Eerola, T. Are the emotions expressed in music genre-specific? An audio-based evaluation of datasets spanning classical, film, pop and mixed genres. *J. New Music Res.* **2011**, *40*, 349–366.

36. Salamon, J.; Peeters, G.; Röbel, A. Statistical Characterisation of Melodic Pitch Contours and its Application for Melody Extraction. In Proceedings of the ISMIR, Porto, Portugal, 8–12 October 2012; pp. 187–192.

37. Bittner, R.M.; Salamon, J.; Bosch, J.J.; Bello, J.P. Pitch Contours as a Mid-Level Representation for Music Informatics. In Proceedings of the Audio Engineering Society Conference: 2017 AES International Conference on Semantic Audio, Erlangen, Germany, 22–24 June 2017.

38. Rao, P.; Ross, J.C.; Ganguli, K.K.; Pandit, V.; Ishwar, V.; Bellur, A.; Murthy, H.A. Classification of melodic motifs in raga music with time-series matching. *J. New Music Res.* **2014**, *43*, 115–131.

39. Gómez, E.; Bonada, J. Towards computer-assisted flamenco transcription: An experimental comparison of automatic transcription algorithms as applied to a cappella singing. *Comput. Music J.* **2013**, *37*, 73–90.

40. Walker, P.; Bremner, J.G.; Mason, U.; Spring, J.; Mattock, K.; Slater, A.; Johnson, S.P. Preverbal infants are sensitive to cross-sensory correspondences much ado about the null results of Lewkowicz and Minar (2014). *Psychol. Sci.* **2014**, *25*, 835–836.

41. Timmers, R.; Li, S. Representation of pitch in horizontal space and its dependence on musical and instrumental experience. *Psychomusicol. Music Mind Brain* **2016**, *26*, 139–148.
42. Leman, M. *Embodied Music Cognition and Mediation Technology*; MIT Press: Cambridge, MA, USA, 2008.
43. Godøy, R.I. Motor-mimetic music cognition. *Leonardo* **2003**, *36*, 317–319.
44. Jensenius, A.R. Action-Sound: Developing Methods and Tools to Study Music-Related Body Movement. Ph.D. Thesis, University of Oslo, Oslo, Norway, 2007.
45. Jensenius, A.R.; Kvifte, T.; Godøy, R.I. Towards a gesture description interchange format. In Proceedings of the 2006 Conference on New Interfaces for Musical Expression, IRCAM—Centre Pompidou, Paris, France, 4–8 June 2006; pp. 176–179.
46. Gritten, A.; King, E. *Music and Gesture*; Ashgate Publishing, Ltd.: Hampshire, UK, 2006.
47. Gritten, A.; King, E. *New Perspectives on Music and Gesture*; Ashgate Publishing, Ltd.: Surrey, UK, 2011.
48. Molnar-Szakacs, I.; Overy, K. Music and mirror neurons: From motion to 'e'motion. *Soc. Cognit. Affect. Neurosci.* **2006**, *1*, 235–241.
49. Sievers, B.; Polansky, L.; Casey, M.; Wheatley, T. Music and movement share a dynamic structure that supports universal expressions of emotion. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 70–75.
50. Buteau, C.; Mazzola, G. From contour similarity to motivic topologies. *Musicae Sci.* **2000**, *4*, 125–149.
51. Clayton, M.; Leante, L. Embodiment in Music Performance. In *Experience and Meaning in Music Performance*; Oxford University Press: New York, NY, USA, 2013; pp. 188–207.
52. Rahaim, M. *Musicking Bodies: Gesture and Voice in Hindustani Music*; Wesleyan University Press: Middletown, CT, USA, 2012.
53. Huron, D.; Shanahan, D. Eyebrow movements and vocal pitch height: Evidence consistent with an ethological signal. *J. Acoust. Soc. Am.* **2013**, *133*, 2947–2952.
54. Savage, P.E.; Tierney, A.T.; Patel, A.D. Global music recordings support the motor constraint hypothesis for human and avian song contour. *Music Percept. Interdiscip. J.* **2017**, *34*, 327–334.
55. Godøy, R.I.; Haga, E.; Jensenius, A.R. Exploring Music-Related Gestures by Sound-Tracing: A Preliminary Study. 2006. Available online: https://www.duo.uio.no/bitstream/handle/10852/26899/Godxy_2006b.pdf?sequence=1&isAllowed=y (accessed on 16 January 2018).
56. Glette, K.H.; Jensenius, A.R.; Godøy, R.I. Extracting Action-Sound Features From a Sound-Tracing Study. 2010. Available online: https://www.duo.uio.no/bitstream/handle/10852/8848/Glette_2010.pdf?sequence=1&isAllowed=y (accessed on 16 January 2018).
57. Küssner, M.B. Music and shape. *Lit. Linguist. Comput.* **2013**, *28*, 472–479.
58. Roy, U.; Kelkar, T.; Indurkhya, B. TrAP: An Interactive System to Generate Valid Raga Phrases from Sound-Tracings. In Proceedings of the NIME, London, UK, 30 June–3 July 2014; pp. 243–246.
59. Kelkar, T. Applications of Gesture and Spatial Cognition in Hindustani Vocal Music. Ph.D. Thesis, International Institute of Information Technology, Hyderabad, India, 2015.
60. Nymoen, K.; Godøy, R.I.; Jensenius, A.R.; Torresen, J. Analyzing Correspondence Between Sound Objects and Body Motion. *ACM Trans. Appl. Percept.* **2013**, *10*, 9.
61. 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 (MIRUM '11), Scottsdale, AZ, USA, 30 November 2011; ACM: New York, NY, USA, 2011; pp. 39–44.
62. Ollen, J.E. A Criterion-Related Validity Test of Selected Indicators of Musical Sophistication Using Expert Ratings. Ph.D. Thesis, The Ohio State University, Columbus, OH, USA, 2006.
63. Nymoen, K.; Torresen, J.; Godøy, R.; Jensenius, A. A statistical approach to analyzing sound-tracings. In *Speech, Sound and Music Processing: Embracing Research in India*; Springer: Berlin/Heidelberg, Germany, 2012; pp. 120–145.
64. Schmuckler, M.A. Components of melodic processing. In *Oxford Handbook of Music Psychology*; Oxford Uniersity Press: Oxford, UK, 2009; p. 93.
65. Godøy, R.I. Gestural affordances of musical sound. In *Musical Gestures: Sound, Movement, and Meaning*; Routledge: New York, NY, USA, 2010; pp. 103–125.
66. Paschalidou, S.; Clayton, M.; Eerola, T. Effort in Interactions with Imaginary Objects in Hindustani Vocal Music—Towards Enhancing Gesture-Controlled Virtual Instruments. In Proceedings of the 2017 International Symposium on Musical Acoust, Montreal, QC, Canada, 18–22 June 2017.

67. Paschalidou, S.; Clayton, M. Towards a sound-gesture analysis in Hindustani Dhrupad vocal music: Effort and raga space. In Proceedings of the ICMEM, University of Sheffield, Sheffield, UK, 23–25 March 2015; Volume 23, p. 25.

68. Pearson, L.; others. Gesture in Karnatak Music: Pedagogy and Musical Structure in South India. Ph.D. Thesis, Durham University, Durham, UK, 2016.