

# EXPLORING MELODY AND MOTION FEATURES IN “SOUND-TRACINGS”

**Tejaswinee Kelkar**

University of Oslo, Department of Musicology  
tejaswinee.kelkar@imv.uio.no

**Alexander Refsum Jensenius**

University of Oslo, Department of Musicology  
a.r.jensenius@imv.uio.no

## ABSTRACT

Pitch and spatial height are often associated when describing music. In this paper we present results from a sound-tracing study in which we investigate such sound–motion relationships. The subjects were asked to move as if they were creating the melodies they heard, and their motion was captured with an infra-red, marker-based camera system. The analysis is focused on calculating feature vectors typically used for melodic contour analysis. We use these features to compare melodic contour typologies with motion contour typologies. This is based on using proposed feature sets that were made for melodic contour similarity measurement. We apply these features to both the melodies and the motion contours to establish whether there is a correspondence between the two, and find the features that match the most. We find a relationship between vertical motion and pitch contour when evaluated through features rather than simply comparing contours.

## 1. INTRODUCTION

How can we characterize melodic contours? This question has been addressed through parametric, mathematical, grammatical, and symbolic methods. The applications of characterizing melodic contour can be for finding similarity in different melodic fragments, indexing musical pieces, and more recently, for finding motifs in large corpora of music. In this paper, we compare pitch contours with motion contours derived from people’s expressions of melodic pitch as movement. We conduct an experiment using motion capture to measure body movements through infra-red cameras, and analyse the vertical motion to compare it with pitch contours.

### 1.1 Melodic Similarity

Marsden disentangles some of our simplification of concepts while dealing with melodic contour similarity, explaining that the conception of similarity itself means different things at different times with regards to melodies [1]. Not only are these differences culturally contingent, but also dependent upon the way in which music is represented as data. Our conception of melodic similarity can

be compared to the distances of melodic objects in a hyperspace of all possible melodies. Computational analyses of melodic similarity have also been essential for dealing with issues regarding copyright infringement [2], “query by humming” systems used for music retrieval [3, 4], and for use in psychological prediction [5].

### 1.2 Melodic Contour Typologies

Melodic contours serve as one of the features that can describe melodic similarity. Contour typologies, and building feature sets for melodic contour have been experimented with in many ways. Two important variations stand out — the way in which melodies are represented and features are extracted, and the way in which typologies are derived from this set of features, using mathematical methods to establish similarity. Historically, melodic contour has been analysed in two principal ways, using (a) symbolic notation, or (b) recorded audio. These two methods differ vastly in their interpretation of contour and features.

### 1.3 Extraction of melodic features

The extraction of melodic contours from symbolic features has been used to create indexes and dictionaries of melodic material [6]. This method simply uses signs such as +/-/=, to indicate the relative movement of each note. Adams proposes a method through which the key points of a melodic contour — the high, low, initial, and final points of a melody — are used to create a feature vector that he then uses to create typologies of melody [7]. It is impossible to know with how much success we can constrain melodic contours in finite typologies, although this has been attempted through these methods and others. Other methods, such as that of Morris, constrain themselves to tonal melodies [8], and yet others, such as Friedmann’s, rely on relative pitch intervals [9]. Aloupis et. al. use geometrical representations for melodic similarity search. Although many of these methods have found robust applications, melodic contour analysis from notation is harder to apply to diverse musical systems. This is particularly so for musics that are not based on western music notation. Ornaments, for example, are easier to represent as sound signals than symbolic notation.

Extraction of contour profiles from audio-based pitch extraction algorithms has been demonstrated in several recent studies [10, 11], including specific genres such as flamenco voice [12, 13]. While such audio-based contour extraction may give us a lot of insight about the musical data at hand,

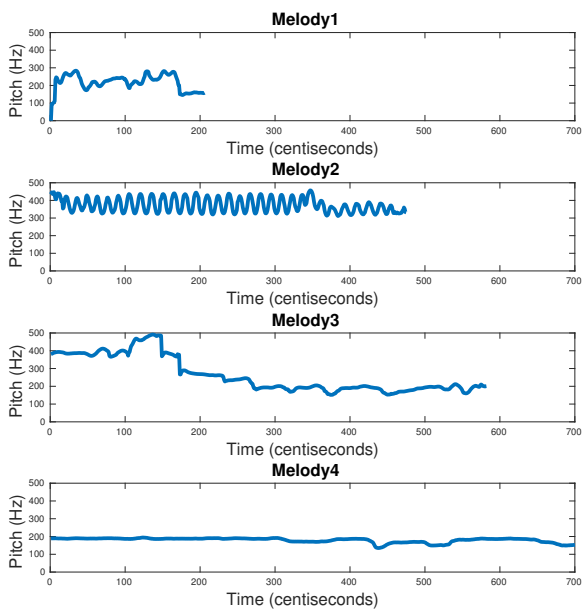


Figure 1. Examples of Pitch features of selected melodies, extracted through autocorrelation.

the generalisability of such a method is harder to evaluate than those of the symbolic methods.

#### 1.4 Method for similarity finding

While some of these methods use matrix similarity computation [14], others use edit distance-based metrics [15], and string matching methods [16]. Extraction of sound signals to symbolic data that can then be processed in any of these ways is yet another method to analyse melodic contour. This paper focuses on evaluating melodic contour features through comparison with motion contours, as opposed to being compared to other melodic phrases. This would shed light on whether the perception of contour as a feature is even consistent, measurable, or whether we need other types of features to capture contour perception.

Yet another question is how to evaluate contours and their behaviours when dealing with data such as motion responses to musical material. Motion data could be transposed to fit the parameters required for score-based analysis, which could possibly yield interesting results. Contour extraction from melody, motion, and their derivatives could also demonstrate interesting similarities between musical motion and melodic motion. This is what this paper tries to address: looking at the benefits and disadvantages of using feature vectors to describe melodic features in a multimodal context. The following research questions were the most important for the scope of this paper:

1. Are the melodic contours described in previous studies relevant for our purpose?
2. Which features of melodic contours correspond to features extracted from vertical motion in melodic tracings?

In this paper we compare melodic movement, in terms of pitch, with vertical contours derived from motion capture recordings. The focus is on three features of melodic contour, using a small dataset containing motion responses of 3 people to 4 different melodies. This dataset is from a larger experiment containing 32 participants and 16 melodies.

## 2. BACKGROUND

### 2.1 Pitch Height and Melodic Contour

This paper is concerned with *melody*, that is, sequences of pitches, and how people trace melodies with their hands. Pitch appears to be a musical feature that people easily relate to when tracing sounds, even when the timbre of the sound changes independently of the pitch [17–19]. Melodic contour has been studied in terms of symbolic pitch [20, 21]. Eitan explores the multimodal associations with pitch height and verticality in his papers [22, 23]. Our subjective experience of melodic contours in cross cultural contexts is also investigated in Eerola’s paper [24].

The ups and downs in melody have often been compared to other multimodal features that also seem to have up-down contours, such as words that signify verticality. This attribute of pitch to verticality has also been used as a feature in many visualization algorithms. In this paper, we focus particularly on the vertical movement in the tracings of participants, to investigate if there is, indeed, a relationship with the vertical contours of the melodies. We also want to see if this relationship can be extracted through features that have been explored to represent melodic contour. If the features proposed for melodic contours are not enough, we wish to investigate other methods that can be used to represent a common feature vector between melody and motion in the vertical axis. All 4 melodies in the small dataset that we create for the purposes of this experiment are represented as pitch in Figure 1.

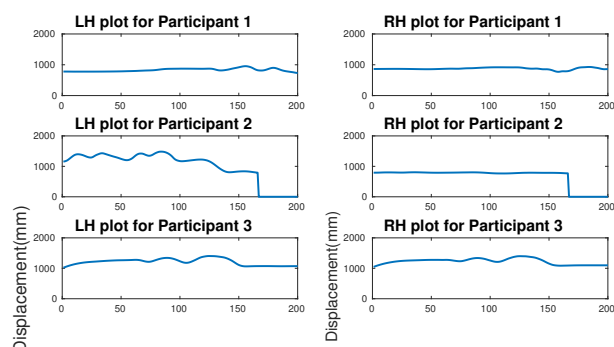


Figure 2. Example plots of some sound-tracing responses to Melody 1. Time (in frames) runs along the x-axes, while the y-axes represent the vertical position extracted from the motion capture recordings (in millimetres). LH=left hand, RH=right hand.



Figure 3. A symbolic transcription of Melody 1, a sustained vibrato of a high soprano. The notated version differs significantly from the pitch profile as seen in Figure 2. The appearance of the trill and vibrato are dimensions that people respond through in motion tracings, that don't clearly appear in the notated version.

	Feature 1	Feature 3
Melody1	[+, -, +, -, +, -],	[0, 4, -4, 2, -2, 4, 0, -9],
Melody2	[+, -, -]	[0, 2, -2, -2, 0, 0],
Melody3	[+, -, -, -, -, -, -],	[0, -2, -4, -1, -1, -1, -4, -2, -3, 0, 0, 0],
Melody4	[+, -, +, -, -, +, -, -]	[0, -2, 2, -4, 2, -2, 4, -2, -2]

Table 1. Examples of Features 1 and 3 for all 3 melodies from score.

## 2.2 Categories of contour feature descriptors

In the following paragraphs, we will describe how the feature sets selected for comparison in this study are computer. The feature sets that come from symbolic notation analysis are revised to compute the same features from the pitch extracted profiles of the melodic contours.

### 2.2.1 Feature 1: Sets of signed pitch movement direction

These features are described in [6], and involve a description of the points in the melody where the pitch ascends or descends. This method is applied by calculating the first derivatives of the pitch contours, and assigning a change of sign whenever the spike in the velocity is greater than or less than the standard deviation of the velocity. This helps us come up with the transitions that are more important to the melody, as opposed to movement that stems from vibratos, for example.

### 2.2.2 Feature 2: Initial, Final, High, Low features

Adams, and Morris [7, 8] propose models of melodic contour typologies and melodic contour description models that rely on encoding melodic features using these descriptors, creating a feature vector of those descriptors. For this study, we use the feature set containing initial, final, high and low points of the melodic and motion contours computed directly from normalized contours.

### 2.2.3 Feature 3: Relative interval encoding

In these sets of features, for example as proposed in Friedman, Quinn, Parsons, [6, 9, 14], the relative pitch distances are encoded either as a series of ups and downs, combined with features such as operators ( $\uparrow, \downarrow$ ) or distances of relative pitches in terms of numbers. Each of these methods employs a different strategy to label the high and low



Figure 4. Lab set-up for the Experiment with 21 markers positioned on the body. 8 Motion capture cameras are hanging on the walls.

points of melodies. Some rely on tonal pitch class distribution, such as Morris's method, which is also analogous to Schenkerian analysis in terms of ornament reduction; while others such as Friedmann's only encode changes that are relative to the ambit of the current melodic line. For the purposes of this study, we pick the latter method given as all the melodies in this context are not tonal in the way that would be relevant to Morris.

## 3. EXPERIMENT DESCRIPTION

The experiment was designed so that subjects were instructed to perform hand movements 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 locations of all of their previous movements, thus giving us an insight of the perceptually salient properties of the melodies that they choose to represent.

### 3.1 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 were taken from real recordings, containing complete phrases. This retained the melodies in the form that they were sung and heard in, thus preserving their ecological quality. The choice of vocal melodies was both to eliminate the effect of words on the perception of music, but also to eliminate the possibility of imitating the sound-producing actions on instruments ("air-instrument" performance) [25].

There was a pause before and after each phrase. The phrases were an average of 4.5 seconds in duration (s.d. 1.5s). These samples were presented in two conditions: (1) the real recording, and (2) a re-synthesis through a sawtooth wave from an autocorrelation analysis of the pitch profile. There was thus a total of 32 stimuli per participant.

The sounds were played at comfortable listening level through a Genelec 8020 speaker, placed 3 metres ahead of the participants at a height of 1 meter.

### 3.2 Participants

A total of 32 participants (17 female, 15 male) were recruited to move to the melodic stimuli in our motion capture lab. The mean age of the participants was 31 years ( $SD=9$ ). The participants were recruited from the University of Oslo, and included students, and employees, who were not necessarily from a musical background.

The study was reported to and obtained ethical approval from the Norwegian Centre for Research Data. The participants signed consent forms and were free to withdraw during the experiment, if they wished.

### 3.3 Lab set-up

The experiment was run in the fourMs motion capture lab, using a Qualisys motion capture system with eight wall-mounted Oqus 300 cameras (Figure 3.1, 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.

### 3.4 Procedure

The participants were asked to trace all 32 melody phrases (in random order) as if their hand motion was ‘producing’ the melody. The experiment lasted for a total duration of 10 minutes. After post processing the data from this experiment, we get a dataset for motion of 21 markers while the participants performed sound-tracing. We take a subset of this data for further analysis of contour features. In this step, we extract the motion data for the left and the right hands from a small subset of 4 melodies performed by 3 participants. We focus on the vertical movement of both the hands given as this analysis pertains to verticality of pitch movement. We process these motion contours along with the pitch contours for the 4 selected melodies, through 3 melodic features as described in section 2.2.

## 4. MELODIC CONTOUR FEATURES

For the analysis, we record the following feature vectors through some of the methods mentioned in section 1.2. The feature vectors are calculated as mentioned below:

**Feature 1** Signed interval distances: The obtained motion and pitch contours are binned iteratively to calculate average values in each section. Mean vertical motion for all participants is calculated. This mean motion is then binned in the way that melodic contours are binned. The difference between the values of the successive bins is calculated. The sign of this difference is concatenated to form a feature vector composed of signed distances.

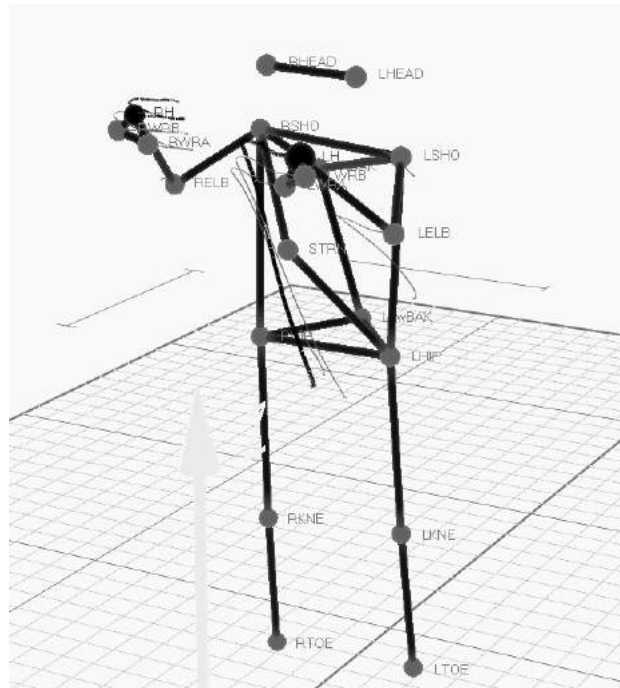


Figure 5. Example of post-processed Motion Capture Recording. The markers are labelled and their relative positions on the co-ordinate system is measured.

**Feature 2** Initial, Final, Highest, Lowest vector: These features were obtained by calculating the four features mentioned above as indicators of the melodic contour. This method has been used to form a typology of melodic contours.

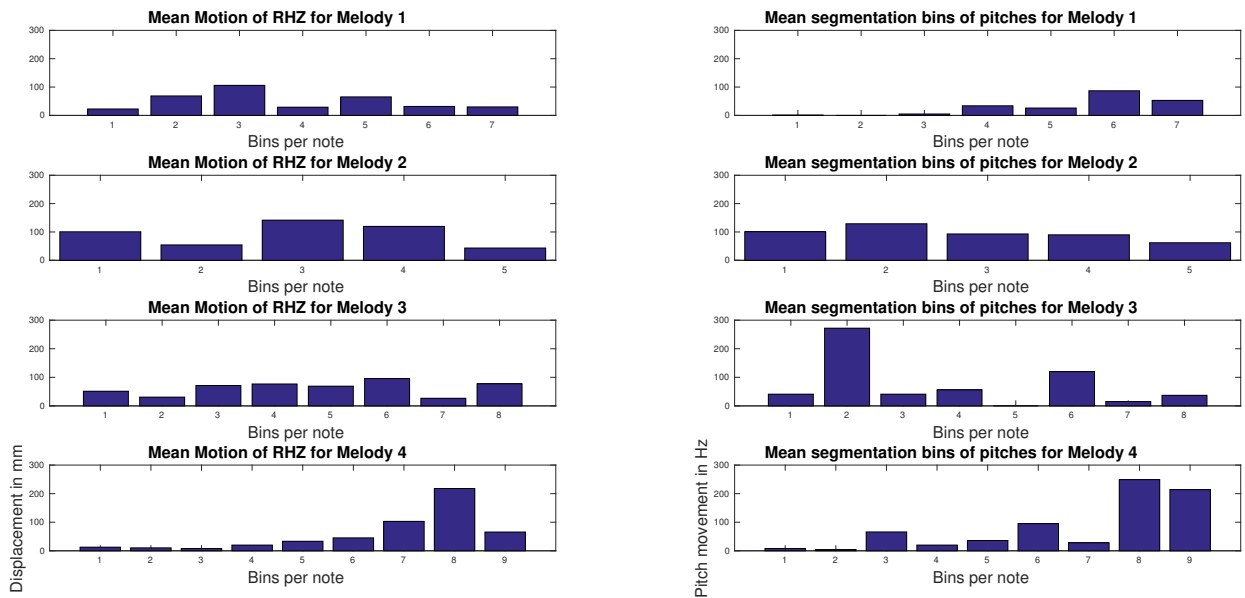
**Feature 3** Signed relative distances: The obtained signs from Feature 1 are combined with relative distances of each successive bin from the next. The signs and the values are combined to give a more complete picture. Here we considered the pitch values at the bins. These did not represent pitch class sets, and therefore made the computation ‘genre-agnostic.’

Signed relative distances of melodies are then compared to signed relative distances of average vertical motion to obtain a feature vector.

## 5. RESULTS

### 5.1 Correlation between pitch and vertical motion

Feature 3, which considered an analysis of signed relative distances had the correlation coefficient of 0.292 for all 4 melodies, with a p value of 0.836 which does not show a confident trend. Feature 2, containing a feature vector for melodic contour typology, performs with a correlation coefficient of 0.346, indicating a weak positive relationship, with a p value of 0.07, which indicates a significant positive correlation. This feature performs well, but is not robust in terms of its representation of the contour itself, and fails when individual tracings are compared to melodies, yielding an overall coefficient of 0.293.



(a) Motion Responses

(b) Melodic Contour Bins

Figure 6. Plots of the representation of features 1 and 3. These features are compared to analyse similarity of the contours.

## 5.2 Confusion between tracing and target melody

As seen in the confusion matrix in Figure 7, the tracings are not clearly classified as target melodies by direct comparison of contour values itself. This indicates that although the feature vectors might show a strong trend in vertical motion mapping to pitch contours, this is not enough for significant classification. This demonstrates the need for having feature vectors that adequately describe what is going on in music and motion.

## 6. DISCUSSION

A significant problem when analysing melodies through symbolic data is that a lot of the representation of texture, as explained regarding Melody 2, gets lost. Vibratos, ornaments, and other elements that might be significant for the perception of musical motion can not be captured efficiently through these methods. However, these ornaments certainly seem salient for people’s bodily responses. Further work needs to be carried out to explain the relationship of ornaments and motion, and this relationship might have little or nothing to do with vertical motion.

We also found that the performance of a tracing is fairly intuitive to the eye. The decisions for choosing particular methods of expressing the music through motion do not appear odd when seen from a human perspective, and yet characterizing what are significant features for this cross-modal comparison is a much harder question.

Our results show that vertical motion seems to correlate with pitch contours in a variety of ways, but most significantly when calculated in terms of signed relative values. Signed relative values, as in Feature 3, also maintain the context of the melodic phrase itself, and this is seen to be significant for sound-tracings. Interval distances matter

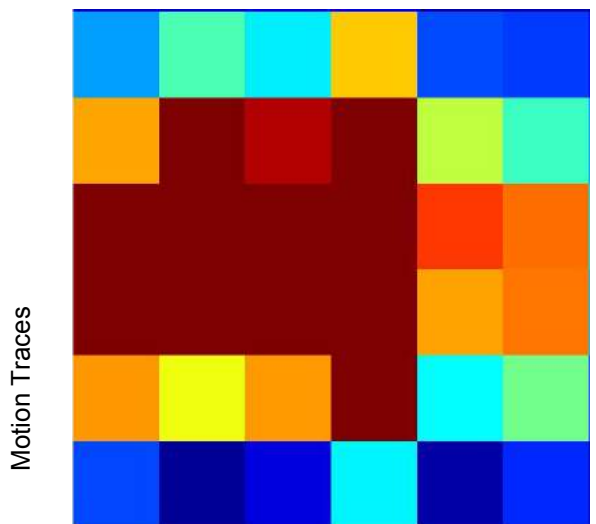
less than the current ambit of melody that is being traced.

Other contours apart from pitch and melody are also significant for this discussion, especially timbral and dynamic changes. However, the relationships between those and motion were beyond the scope of this paper. The interpretation of motion other than just vertical motion is also not handled within this paper.

The features that were shown to be significant can be applied for the whole dataset to see relationships between vertical motion and melody. Contours of dynamic and timbral change can also be interesting to compare with the same methods against melodic tracings.

## 7. REFERENCES

- [1] A. Marsden, “Interrogating melodic similarity: a definitive phenomenon or the product of interpretation?” *Journal of New Music Research*, vol. 41, no. 4, pp. 323–335, 2012.
- [2] C. Cronin, “Concepts of melodic similarity in music copyright infringement suits,” *Computing in musicology: a directory of research*, no. 11, pp. 187–209, 1998.
- [3] A. Ghias, J. Logan, D. Chamberlin, and B. C. Smith, “Query by humming: musical information retrieval in an audio database,” in *Proceedings of the third ACM international conference on Multimedia*. ACM, 1995, pp. 231–236.
- [4] L. Lu, H. You, H. Zhang *et al.*, “A new approach to query by humming in music retrieval.” in *ICME*, 2001, pp. 22–25.



Target Melodies including synthesized variants

Figure 7. Confusion matrix for Feature 3, to analyse the classification of raw motion contours with pitch contours for 4 melodies.

- [5] N. N. Vempala and F. A. Russo, "Predicting emotion from music audio features using neural networks," in *Proceedings of the 9th International Symposium on Computer Music Modeling and Retrieval (CMMR)*. Lecture Notes in Computer Science London, UK, 2012, pp. 336–343.
- [6] D. Parsons, *The directory of tunes and musical themes*. Cambridge, Eng.: S. Brown, 1975.
- [7] C. R. Adams, "Melodic contour typology," *Ethnomusicology*, pp. 179–215, 1976.
- [8] R. D. Morris, "New directions in the theory and analysis of musical contour," *Music Theory Spectrum*, vol. 15, no. 2, pp. 205–228, 1993.
- [9] M. L. Friedmann, "A methodology for the discussion of contour: Its application to schoenberg's music," *Journal of Music Theory*, vol. 29, no. 2, pp. 223–248, 1985.
- [10] J. Salamon and E. Gómez, "Melody extraction from polyphonic music signals using pitch contour characteristics," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 20, no. 6, pp. 1759–1770, 2012.
- [11] R. M. Bittner, J. Salamon, S. Essid, and J. P. Bello, "Melody extraction by contour classification," in *Proc. ISMIR*, pp. 500–506.
- [12] E. Gómez and J. Bonada, "Towards computer-assisted flamenco transcription: An experimental comparison of automatic transcription algorithms as applied to a cappella singing," *Computer Music Journal*, vol. 37, no. 2, pp. 73–90, 2013.
- [13] J. C. Ross, T. Vinutha, and P. Rao, "Detecting melodic motifs from audio for hindustani classical music." in *ISMIR*, 2012, pp. 193–198.
- [14] I. Quinn, "The combinatorial model of pitch contour," *Music Perception: An Interdisciplinary Journal*, vol. 16, no. 4, pp. 439–456, 1999.
- [15] G. T. Toussaint, "A comparison of rhythmic similarity measures." in *ISMIR*, 2004.
- [16] D. Bainbridge, C. G. Nevill-Manning, I. H. Witten, L. A. Smith, and R. J. McNab, "Towards a digital library of popular music," in *Proceedings of the fourth ACM conference on Digital libraries*. ACM, 1999, pp. 161–169.
- [17] K. Nymoen, "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*, 2011.
- [18] M. B. Küssner and D. Leech-Wilkinson, "Investigating the influence of musical training on cross-modal correspondences and sensorimotor skills in a real-time drawing paradigm," *Psychology of Music*, vol. 42, no. 3, pp. 448–469, 2014.
- [19] G. Athanasopoulos and N. Moran, "Cross-cultural representations of musical shape," *Empirical Musicology Review*, vol. 8, no. 3-4, pp. 185–199, 2013.
- [20] M. A. Schmuckler, "Testing models of melodic contour similarity," *Music Perception: An Interdisciplinary Journal*, vol. 16, no. 3, pp. 295–326, 1999.
- [21] J. B. Prince, M. A. Schmuckler, and W. F. Thompson, "Cross-modal melodic contour similarity," *Canadian Acoustics*, vol. 37, no. 1, pp. 35–49, 2009.
- [22] Z. Eitan and R. Timmers, "Beethovens last piano sonata and those who follow crocodiles: Cross-domain mappings of auditory pitch in a musical context," *Cognition*, vol. 114, no. 3, pp. 405–422, 2010.
- [23] Z. Eitan and R. Y. Granot, "How music moves," *Music Perception: An Interdisciplinary Journal*, vol. 23, no. 3, pp. 221–248, 2006.
- [24] T. Eerola and M. Bregman, "Melodic and contextual similarity of folk song phrases," *Musicae Scientiae*, vol. 11, no. 1 suppl, pp. 211–233, 2007.
- [25] R. I. Godøy, E. Haga, and A. R. Jensenius, "Playing air instruments: mimicry of sound-producing gestures by novices and experts," in *International Gesture Workshop*. Springer, 2005, pp. 256–267.