Segmentation of Bodily Gestures Induced by Music
Juan Ignacio Mendoza & Marc Richard Thompson
Department of Music, Art and Culture Studies; University of Jyväskylä, Finland
juanmendo@student.jyu.fi, marc.thompson@jyu.fi

BACKGROUND
In line with the Embodied Music Cognition train of thought (Leman, 2008), it has been argued that a person’s spontaneous movement when listening to music can reflect the person’s perception of the music. The correspondence between music and bodily movement has been studied under the term musical gesture (Schneider, 2010). The first stage in perception of gesture is the identification of when and where it starts and ends, a process called segmentation (Kahol, Tripathi & Panchananath, 2004). Modelling perceived segmentation of bodily gestures serves to a better understanding of human perception. Also, it allows a computing machine to segment gestures in the same way as a human being.

AIMS
This research project is aimed to observe, model and predict the perceived segmentation of bodily gestures induced by music. The project is composed by three stages: Building a multimodal database, the collection of ground truth and the development of an automatic system that performs segmentation of bodily gesture.

MULTIMODAL DATABASE
Naive participants spontaneously move to music excerpts between 40 and 100 s. For each music excerpt they are recorded in two conditions: free movement and ‘dancing with one arm’. In the two conditions they wear an optical motion capture suit with reflective markers. In the second condition they hold an accelerometer with the hand of the arm that moves.

Recorded data modalities are:
- Optical motion capture (3D position)
- Accelerometer
- Video
- Audio (music excerpts)

Free movement Dancing with one arm

GROUND TRUTH
Semi-expert annotators watch videos of the database and indicate where there is a change of gesture. Gesture is defined as a new pattern. Thus, if movement repeats consecutively, no indication shall be made until a new pattern is perceived.
The annotation task is done in two conditions:
Real-time annotation
While watching and listening, the annotator presses a button where a change in gesture is perceived.
Non-real-time annotation
Video is presented without audio and the annotator can scroll back and forward to accurately place a marker at a perceived change of gesture.

RESULTS AND FUTURE WORK
Preliminary results have been obtained by a constrained brute-force search for greatest similarity between computed and perceived segmentation boundaries. The latter corresponds to responses of one annotator, for a video of one participant in the condition ‘dancing with one arm’ to one musical excerpt. The search was made without combining functions (Step 5), revealing that kurtosis, skewness, interquartile range and root mean square (RMS), without use of EMD (Step 2), are satisfactory predictors for isolated regions of boundaries. The closest similarity for the full sequence of boundaries was still not satisfactory. This suggests that each windowed function characterises a specific kind of gesture. Future work will evaluate the system with more of the collected motion data and ground truth data, combining functions and optimising the search with a genetic algorithm, as done in research on auditory segmentation (Hartmann, Lartilott & Toivainen, 2016).

REFERENCES
Segmentation of Bodily Gestures Induced by Music
Juan Ignacio Mendoza$^1$ & Marc Richard Thompson$^2$
Department of Music, Art and Culture Studies; University of Jyväskylä, Finland
jmendoz@student.jyu.fi, mthompson@jyu.fi

BACKGROUND
In line with the Embodied Music Cognition train of thought (Leman, 2008), it has been argued that a person's spontaneous movement when listening to music can reflect the person's perception of the music. The correspondence between music and bodily movement has been studied under the term musical gesture (Schneider, 2010). The first stage in perception of a gesture is the identification of where it starts and ends, a process called segmentation (Kahol, Tripathi, & Panchananathan, 2004). Modelling perceived segmentation of bodily gestures serves to a better understanding of human perception. Also, it allows a computing machine to segment gestures in the same way as a human brain.

AIMS
This research project is aimed to observe, model and predict the perceived segmentation of bodily gestures induced by music. The project is composed by three stages: Building a multimodal database, the collection of ground truth and the development of an automatic system that performs segmentation of bodily gesture.

MULTIMODAL DATABASE
Naïve participants spontaneously move to music excerpts between 40 and 100 s. For each music excerpt they are recorded in two conditions: free movement and 'dancing with one arm'. In the two conditions they wear an optical motion capture suit with reflective markers. In the second condition they hold an accelerometer with the hand of the arm that moves.

Recorded data modalities are:
- Optical motion capture (3D position)
- Accelerometer
- Video
- Audio (music excerpts)

GROUNDF TRUTH
Semi-expert annotators watch videos of the database and indicate where there is a change of gesture. Gesture is defined as a new pattern. Thus, if movement repeats consecutively, no indication shall be made until a new pattern is perceived.

The annotation task is done in two conditions:
- Real-time annotation: Video is presented with audio. While watching and listening, the annotator press a button where a change in gesture is perceived.
- Non-real-time annotation: Video is presented without audio and the annotator can scroll back and forward to accurately place a marker at a perceived change of gesture.

RESULTS AND FUTURE WORK
Preliminary results have been obtained by a constrained brute-force search for greatest similarity between computed and perceived segmentation boundaries. The latter corresponds to responses of one annotator; for a video of one participant in the condition 'dancing with one arm' to one musical excerpt. The search was made without combining functions (Step 5), revealing that kurtosis, skewness, interquartile range, centered zero-crossings, local extrema, integration, spectral entropy, spectral centroid, spectral flatness, and window length are satisfactory predictors for isolated regions of boundaries.

REFERENCES