Dance to your own drum: identification of musical genre and individual dancer from motion capture using machine learning

Emily Carlson[1], Pasi Saari, Birgitta Burger, Petri Toiviainen

Department of Music, Arts and Culture, University of Jyväskylä, Jyväskylä, Finland

[1]Corresponding author
Emily Carlson
University of Jyväskylä
Department of Music, Arts and Culture
P.O. Box 35
FI-40014
Finland
email: emily.j.carlson@jyu.fi

Abstract

Machine learning has been used to accurately classify musical genre using features derived from audio signals. Musical genre, as well as lower-level audio features of music, have also been shown to have an influence on music induced movement, however the degree to which embodied responses to music are genre-specific has not been explored. The current paper addresses this using motion capture data from participants dancing freely to eight genres (Blues, Country, Dance, Jazz, Metal, Pop, Reggae and Rap). Using a Support Vector Machine (SVM) model, data were classified according to extracted kinematic features by genre, as well as by individual dancer. Against expectations, individual classification was notably more accurate than genre classification, although higher accuracy in classifying movements done to Metal and Jazz parallels some previously findings that these genres were more easily classified by audio signal. Results are discussed in terms of embodied cognition and culture.

Keywords: motion capture, machine learning, embodied cognition

1. Introduction

The universality of music has almost certainly been overstated by well-intended optimists and poets, as well as by researchers who have focused exclusively on phenomena arising from Western musical traditions (Cross et al., 2001). Still, something that can be qualified as music appears to be engaged in by all cultures, and in many cases that something is the use of a regular or isochronous beat that affords synchronization (Nettl, 2000), making it possible that one of the most universal things about music is dance (Richter & Ostovar, 2016). One of the most salient features is its tendency to make us move; the majority of people respond to hearing music with some kind of movement, from simply clapping to a beat to engaging in complex dance movement (Lesaffre et al., 2008).

It seems reasonable to expect that such music-induced movement should be affected by the particular qualities of the music which is influencing the movement—surely one does not move the same way in response to a song by Rage Against the Machine as to one by Bob Dylan—and research has indeed shown that audio features extracted from the acoustic signal of music influence the quality of dancers' movements. Van Dyck et al., (2013) showed that participants modified and increased their dance movements relative to volume of the bass drum. Burger, Thompson, Luck, Saarikallio, and Toiviainen (2013) extracted spectral features and rhythmic features from musical stimuli and compared them with kinematic features extracted from recorded movements of participants dancing to these stimuli. They found that low frequency activity, associated with the presence of kick drum and bass guitar, uniquely related to the speed of head movement, while high frequency activity and beat clarity were associated with a wider variety of movement features including hand distance, hand speed, shoulder wiggle and hip wiggle. These results can be compared to those earlier found by Luck, Saarikallio, Burger, Thompson, and Toiviainen (2010), who noted that Rock music was associated with greater head speed during dance while Jazz was associated with lesser head speed, while Techno, Latin and Metal music were all associated with

specific movement patterns. Such stereotypical movement patterns likely reflect not only audio features of the music, but cultural norms associated with specific genres; for example, there is a close association between Jazz music and swing dancing (Spring, 1997), and between Metal and headbanging (Hudson, 2015). However, such influence may be subject to dancers' familiarity and degree of engagement with the musical culture in question.

Not only is movement a common response to music, some suggest that movement is even necessary to understanding and parsing musical sounds. Godøy, Song, Nymoen, Haugen, and Jensenius (2016) take evidence from studies of music performance, *sound-tracing* studies in which listeners are asked to 'draw' their impressions of audio stimuli, and dance movement studies, and propose that 'any sound event entails an image and context of a body-motion event' and that sound and motion can be considered two aspects of 'the same phenomenon' (p. 214). While this is a bold statement, it does conform with the growing interest among psychology researchers in embodied cognition; that is, the idea that human cognition is not only dependent on the perceptual information gained through the body's sensing of the outside world and of itself, but that cognition and bodily experiences are essentially inseparable (Shapiro, 2007; Wilson & Golonka, 2013).

Embodied cognition, from which embodied music cognition is derived, is a debated idea not so much regarding its validity but regarding how it should be defined, what it means, and how we can best understand and use it in our attempts to understand human nature. To put it in deceptively simple terms, *embodied cognition* refers to the idea that invisible aspects of human experience, namely cognition and emotion, arise from and are characterized by the forms and functions of the human body. Leman (2008) has defined *embodied music cognition* as direct, rather than symbolic, musical experience, where music is considered to be comprised of *moving sonic forms* which the listener parses through a process of corporeal imitation, either internally or externally. Lakoff and Johnson (1999) provide examples of how our bodily experiences of spatial relationships form understanding, such as the idea that up equates with 'more' and down equates with 'less,' which we

experience as part of our sensorimotor reality. Bodily experiences pervade our ability to understand and communicate everything from these most basic concepts to our common metaphors for making sense of our lives as well as, notably in this context, our experiences of music. We *rise* to the occasion; music *rises* in pitch or volume; we *fall* off the band wagon; the melody *drops* to the bass; life is *a journey*; that song *takes us back*; and we *dance* to the beat of our own drum. For Leman and others, such embodied processes define our responses to and ability to understand the acoustic signals which we can imitate with our bodies and thereby experience and understand.

All of this does seem to support the idea that different music should elicit different movement patterns from listeners. However, testing this idea is more complex than it appears on the surface, in part because the idea of what constitutes 'similar' and 'different' in contemporary Western music is a question with no clear-cut answer. With the rise of recording technology, which transformed music in the Western world from an activity into purchasable commodity, came the need to label music effectively so that listeners would know what to buy, a task for which the notion of genre is often employed. Mace, Wagoner, Teachout, and Hodges (2011) have shown that individuals can distinguish between Classical, Jazz, Country, Metal and HipHop with more than 50% accuracy from clips as short as 125 milliseconds. However, while genre labels such as 'Pop' or 'Rock' are commonly used to categorize music, exact boundaries between these categories are unclear, and used inconsistently across listening platforms (Aucouturier & Pachet, 2003; Pachet & Cazaly, 2000). With the advent of the digital age, access to vast libraries of recorded music as both eased and complicated music consumption, as the need for a meaningful method of organizing and labeling millions of digital recordings has increased exponentially. It can be noted here that our ideas of categorization are also influenced by embodied concepts, as we understand and describe categories in terms of physical containers into which we can put appropriate items; Lakoff and Johnson (1999) note this as one of the fundamental ways in which physical reality underlies our cognitive experiences.

Some have sought to enlist computational algorithms in the task of properly categorizing a song into a genre based on information extracted from its audio signal. Tzanetakis and Cook (2002) were among the first to introduce this idea, and have provided a clear framework for the process which many have followed since. First, relevant features must be extracted from the audio signal; various spectral features relating to timbre have been employed in a number of studies (Hartmann, Saari, Toiviainen, & Lartillot, 2013; Holzapfel & Stylianou, 2006), as well as rhythmic features (Genussov & Cohen, 2010; Tzanetakis & Cook, 2002), and pitch features (Tolonen, Member, & Karjalainen, 2000). Once features have been chosen and computed, a method of evaluation should be chosen, such as methods of statistical pattern recognition, linear classifiers, and non-parametric classifiers based on data clustering and nearest neighbors (Tzanetakis & Cook, 2002).

Computational classification of genre based on audio signal is an attractive method for managing the large databases of recorded music. However, the factors that differentiate one genre from another are composed of more than acoustic features; Tzanetakis and Cook (2002) note that genres 'arise through complex interaction between the public, marketing, historical and cultural factors,' (p. 293). Evidence from attempts to classify genre based on acoustic signals, exploration of genre as a socially determined phenomenon via analysis of social-tagging data, and manual attempts at creating genre taxonomies suggest that some genres are more acoustically distinctive than others (Aucouturier & Pachet, 2003; Carlson, Saari, Burger, & Toiviainen, 2017; Hartmann et al., 2013; Holzapfel & Stylianou, 2006; Scaringella, Zoia, & Mlynek, 2006; Sordo, Celma, Blech, & Guaus, 2008). What this means in terms of embodied responses to music of different genres is still an open question. Although there is, as described above, evidence of some commonality between individuals in response to acoustic signals (e.g., Burger et al., 2013; Godøy et al., 2016), there is also evidence that human individual differences, as in differences in personality, influence the characteristics of these embodied responses to music (Carlson, Burger, London, Thompson, & Toiviainen, 2016; Carlson, Burger, & Toiviainen, 2018; Luck et al., 2010). In fact, individuality of

movement patterns was among the first topics studied after Johansson demonstrated that humans can perceive human movement from video in which only lights placed on key joints were visible (see Figure 1) (Johansson, 1973), Cutting and Kozlowski (1977) demonstrating that friends could recognize each other from their walk with only such point-light (or stick figure) displays of movement, without the need for other distinguishing features.

Subsequent research has shown that humans have a surprisingly robust ability for recognizing individuals based on their movements (as from point-light animations, see Methods section), absent any information about other physical characteristics such as size, shape, clothing or facial features (Bläsing & Sauzet, 2018; Swedish & Troje, 2007; Troje, Westhoff, & Lavrov, 2005), as well as more abstract characteristics such as the walker's mood (Michalak et al., 2009), personality (Satchell et al., 2017), and even vulnerability to physical attack (Gunns, Johnston, & Hudson, 2002). Troje and Chang (2013) explored the underlying mechanisms of this by manipulating point-light animations of multiple individuals walking such that all size, shape, and gait frequency information was removed, and further decomposed stimuli systematically into harmonic components. They found that, while participants' accuracy in identifying individuals from these manipulated stimuli decreased as more harmonic information was removed, participants were still able to correctly identify individuals well above chance level when the first harmonic was removed, despite this harmonic accounting for 91% percept of the variance in walking patterns (p. 246). The individuality of movement extends to more complex activities than walking; Sevdalis and Keller (2009) showed that individuals were able to recognize their own motion-captured movements of not only walking, but also of clapping to a beat and dancing. Bläsing and Sauzet (2018) expanded on this idea and found that, when participants were asked to identify dance movements they had either created themselves while blindfolded, a learned movement, or movements they had merely watched, participants better recognized and were more likely to

associate themselves with movements they had created, even in the absence of visual memory of the movement.

Just as with musical genre, the identification of individuals is these days the purview of both man and machine. Readers familiar with large social media sites will be familiar with the application of face-recognition technology, which has been successfully used to analyze relationships between facial features (Guo et al., 2000). Other work has focused on identifying human actions such as running, kicking, or throwing across large samples of different recordings (Guo, Li, & Chan, 2000). However, while humans are able to recognize individuals from whole body movement without the need for other information, computer vision approaches often include analysis of color and shape, which in turn leads to difficulties re-identifying individuals who may, for example, be wearing different clothing from one day to the next (Poppe, 2010).

The above literature review raises some interesting questions about musical genre and its relationship to music-induced movement, as well as about the relationship between individuals' apparently unique movement patterns and the effects of distinct musical stimuli. While there is evidence that both audio features and the genre category of a song can influence how it is embodied by listeners, only a limited number of specific genres has been studied in this regard. For example, Luck et al. (2010) examined the influence of Jazz, Techno, Latin, Funk, Pop and Rock, using only non-vocal excerpts, while Solberg & Jensenius (2017) focused exclusively on Electronic Dance Music (EDM), and Burger and Toiviainen (2018) examined participants' movements in response to EDM compared to Latin, Funk and Jazz. The aim of the current study, therefore, is to explore the distinctions between common Western musical genres in terms of how they are embodied by participants within a free dance movement setting, in which participants are allowed to move as they desire without pre-choreographed constraints. As a comparatively large body of work exists using machine learning to differentiate between genres, the current study will similarly employ machine learning to explore the degree to which genre can be distinguished from the bodily

movements of participants. As previous work also highlights bodily moment as a robust means by which humans can distinguish between others based on movement, this study will also explore the degree to which such individuality of movement is present in individuals' movements across multiple genres.

Within a framework of embodied music cognition, and in light of previous research has shown both that audio features and genre influence music-induced movement (Burger, Thompson, Saarikallio, Luck, & Toiviainen, 2010; Luck et al., 2010) and that genre can be distinguished to a large degree by analysis of audio signals (Genussov & Cohen, 2010; Tzanetakis & Cook, 2002), we hypothesized that machine learning analysis of kinematic features extracted from music-induced movement could be used to identify the genre of the heard musical stimulus. Given strong evidence for the presence of individually identifiable movement features in both music and non-music settings (Karkavitsas & Tsihrintzis, 2011; Nanni, Costa, Lumini, Kim, & Baek, 2016), we also hypothesized that identification of individuals through machine learning analysis of kinematic features would be possible. As previous research has shown that some genres are easier than others to distinguish using machine learning, we further expect that there will be variation between genres in the accuracy of both genre and individual identification.

2. Methods

2.1 Motion Capture Study

A motion capture study was designed to collect free dance movement data from participants using naturalistic (commercially available) musical stimuli representing different genres (see section 2.1.4).

2.1.1 Participants

A total of 73 participants (54 females) completed the motion capture experiment. Participants ranged in age from 19 to 40 years (M = 25.74, SD = 4.72). Thirty held Bachelor's degrees while 16 held Master's degrees. Thirty-three reported having received some formal musical training; five

reported one to three years, ten reported seven to ten years, while 16 reported ten or more years of training. Seventeen participants reported having received some formal dance training; ten reported one to three years, five reported four to six years, while two reported seven to ten. Participants were of 24 different nationalities, with Finland, the United States and Vietnam being the most frequently represented. For attending the experiment, participants received two movie ticket vouchers each. All participants spoke and received instructions in English.

2.1.2 Apparatus

Participants' movements were recorded using a twelve-camera optical motion capture system (Qualisys Oqus 5+), tracking at a frame rate of 120 Hz, the three-dimensional positions of 21 reflective markers attached to each participant. Markers were located as follows (L = left, R = right, F = front, B = back) 1: LF head; 2: RF head; 3: B head; 4: L shoulder; 5: R shoulder; 6: sternum; 7: stomach; 8: LB hip; 9: RB hip; 10: L elbow; 11: R elbow; 12: L wrist; 13: R wrist; 14: L middle finger; 15: R middle finger; 16: L knee; 17: R knee; 18: L ankle; 19: R ankle; 20: L toe; 21: R tow, visible in Figure 1A. The musical stimuli were played in a random order via four Genelec 8030A loudspeakers and a sub-woofer. The direct (line-in) audio signal of the playback and the synchronization pulse transmitted by the Qualisys cameras when recording were recorded using ProTools software so as to synchronize the motion capture data with the musical stimulus afterwards.

[Insert Figure 1 about here]

2.1.3 Stimulus selection

The 35-second stimuli for the experiment were selected using a computational process based on social-tagging and acoustic data. Social tags are defined as "free text labels that are applied to items such as artists, albums and songs" (Lamere, 2008, pp 101), the possibility of which is provided by music-listening platforms such as Last.fm. The selection pipeline was designed to select naturalistic stimuli that were uncontroversially representative of their respective genres, which would also be

appropriate to use in a dance setting. A total of 2407 tracks were collected from Last.fm which also were tagged by users with one and only one genre label (e.g. ‘Country’ or ‘Jazz’); these labels were derived from the revised version of the Short Test of Music Preferences, or STOMP-R (Bläsing & Sauzet, 2018; Troje & Chang, 2013). Tracks were also required to have been tagged by Last.fm users with at least one dance-related term, such as ‘danceable,’ ‘dancing,’ ‘head banging,’ or ‘headbanging,’ and were retained only if they had a non-zero danceability score according to Echo Nest (which is determined by computational analysis of a given track’s acoustic features including beat strength, tempo and loudness), and only if the track’s tempo fell between 118-132 BPM. Four randomly selected excerpts from each genre were checked for tempo and stylistic consistency by the researchers, leaving 16 stimuli from 8 genres: Blues, Country, Dance, Jazz, Metal, Pop, Rap, and Reggae. The details of the stimuli are given in Table 1:

[Insert Table 1 about here]

For a complete description of this stimuli-selection methodology, see Carlson et al. (2017).

2.1.4 Procedure

Groups of three or four dancers at a time attended the experiment and were instructed to move freely to the randomized musical stimuli, as they might in a dance club or party setting. They moved first individually (without seeing any other dancers) and in dyads, although only individual data is considered in the current analysis. In each condition (individual and all possible dyads), all 16 stimuli were heard in randomized order. Participants were asked to listen to the music and move freely as they desired, staying within the marked capture space. The aim of these instructions was to create a naturalistic paradigm, such that participants would feel free to behave as they might in the real world. To limit the effects of fatigue, participants were informed that they were free to ask for a break or stop the experiment at any time, and were additionally offered water, juice and biscuits as light refreshment.

2.1.5 Preprocessing Mocap Data

Using the Motion Capture (MoCap) Toolbox (Burger & Toiviainen, 2013) in MATLAB, movement data of the 21 markers were first trimmed to match the duration of the musical excerpts. Due to small recording errors (i.e., the mocap recorded being stopped too quickly), several files were 33-seconds in length while the majority were 35-seconds. Gaps in the data were linearly filled. Following this, the data were transformed into a set of 20 secondary markers – subsequently referred to as joints. The locations of these 20 joints are depicted in Figure 1B. The locations of joints B, C, D, E, F, G, H, I, M, N, O, P, Q, R, S, and T are identical to the locations of one of the original markers, while the locations of the remaining joints were obtained by averaging the locations of two or more markers; Joint A: midpoint of the two back hip markers; J: midpoint the shoulder and hip markers; K: midpoint of shoulder markers; and L: midpoint of the three head markers. The instantaneous velocity of each marker in each direction was calculated Instantaneous velocity was estimated by time differentiation followed by the application of a 2nd-order Butterworth filter with a cutoff frequency of 24 Hz for all participants for all 16 stimuli (see Toiviainen and Burger, 2013). Subsequently, the data were converted into local coordinate system, in which the root marker (Figure 1B, marker A) is defined as the origin and the line connecting the hip markers the mediolateral axis, to allow for comparison between individuals who may have been oriented differently in the capture space.

Full details of the experiment can be found in Carlson, Burger and Toiviainen (2018).

2.2. Machine Learning Analysis

A machine learning model involving feature extraction, feature selection, and classification was employed and evaluated using cross validation. Two classification tasks were undertaken: participant classification and genre classification.

The classification method employed for the current analysis is that of Support Vector Machines (SVM), which has become popular over the early 2000s. This method is based on the relatively straightforward idea that, given two classes of data (for example, half the points represent

Mozart String Quartets and the other half songs by Metallica, see Figure 2) graphed onto some two-dimensional space based on some of their identifiable features (number of electric instruments and average amount of head banging during performances), there are theoretically infinite lines that *could* be drawn to divide the data. However, there is only one line that *optimally* divides the two classes such that the introduction of new data is most likely to be classified correctly. SVM is used to identify the line that best divides as the one that is maximally distant from the nearest data point of each classes, providing, essentially, the largest possible buffer space between the two classes; this buffer space defines the Optimal Separating Hyperplane, or OSH. Not only does the OSH separate the classes of data, but it minimizes the risk that new data would be incorrectly classified (G. Guo et al., 2000; Mamonne, Turchi, & Cristianini, 2009).

[Insert Figure 2 about here]

A small tweak to the above example should easily convince the reader of the importance of feature selection. If instead of graphing our data based on electronic instruments or crowd head-banging behavior, the features we used were number of musicians performing and number of musicians wearing black, it would be virtually impossible for our algorithm to distinguish between performances of Mozart string quartets and performances by Metallica. The problem of classification becomes more serious and more abstract when the features available for analysis are limited to the recorded acoustic signals themselves, or indeed to dancers' movements. It should be noted that SVM is not limited to two or even three-dimensional space, nor to linearly-separable classes, so real-world attempts to distinguish genres based on multiple extracted audio features can become highly complex and more challenging to interpret (for a more formal overview of SVM, see Estes, 1962, or Mamonne, Turchi, & Cristianini, 2009).

2.2.1 Kinematic Feature Extraction

The kinematic feature used in this analysis was the covariances of velocity between all the marker time series in each direction (X, Y and Z) within each participant for each stimulus; that is, the

degree to which the movement of any two of a participants' markers covaried with each other across the entire stimulus in any of the three dimensions, resulting in a marker by marker covariance matrix for each participant. Such covariance features have been previously yielded high performance in classification tasks with reasonable computational complexity (Ergezer & Leblebicioğlu, 2018; K. Guo, Ishwar, & Konrad, 2009; Tuzel, Porikli, & Meer, 2008). Furthermore, such relationships between markers were suggested by Troje et al. (2005) to play at least some role in human perceptual identification of individuals. In addition to using linear covariance, however, we also tested a nonlinear measure of covariance, defined using a Radial Basis Function (RBF) kernel (that is, a specific mathematical algorithm used to transform data into the desired form) in the covariance computation, which was normalized according to the time-series length (denoted by d) to facilitate different lengths of some stimuli. This measure, also referred to as the *correntropy* between time series x_i and x_j (Liu, Member, Pokharel, & Principe, 2007), was computed as follows:

$$K(\boldsymbol{x}_i, \boldsymbol{x}_j) = e^{-\|\boldsymbol{x}_i - \boldsymbol{x}_j\|_2^2/(2\sigma^2 T^2)}$$

where $\|\boldsymbol{x}_i - \boldsymbol{x}_j\|_2 = \sqrt{\sum_{t=1}^{T} |x_i(t) - x_j(t)|^2}$ is the L2 norm, also referred to as the Euclidian norm as it is calculated as the Euclidian distance from the origin. This yielded for each participant per stimulus a symmetric covariance matrix whose lower triangular part was subsequently vectorized to produce a feature vector of length 1596. Correntropy calculated using an RBF kernel provides an alternate measure of similarity compared to linear covariance, in which similarity decays as a function of the distance between two vectors (data vectors in this case representing marker movement in a given dimension), in the shape of a bell-curve (e.g., a Gaussian distribution) rather than a straight line. The degree to which two markers covary are quite literally 'graded on the curve,' with a steeper curve resulting in highly covarying pairs being marked as even higher than they would be in a linear measure, and vice versa for low covarying pairs.

The steepness of the curve in the above computation is determined by the value of the sigma parameter, as this affects the distribution of the produced features, where high values yield negatively skewed feature value distribution, and vise-versa for low values. Thus, the data discriminability with respect to the feature values can be low if a sigma value that is either too high or too low is used. Because of this, it was necessary to include RBF-kernel optimization within our machine learning model, such that the features used to discriminate between classes were useful for the relevant tasks. Our solution for determining an appropriate value for sigma was to optimize the sigma value by minimizing the skewness of each feature separately using the downhill simplex algorithm. Each of these kernel-optimized features were normalized into zero mean and unit standard deviation. For comparison, data were also analyzed using linear covariances.

2.2.2 Feature Selection

After the RBV covariance matrix was extracted for each participant, distinguishing features (that is, pairs of markers) were further selected for analysis using SVM. In machine learning, models are regularized through the use of a penalty term, which is applied in feature selection to control the model complexity and prevent overfitting (that is, the creation of a model with so many features as to be impractical to generalize beyond the current classification problem). Simply put, regularization limits the number of features used within a model to those of sufficient importance such that the model does not become too specific. In this case, an SVM classifier using the linear kernel and L1 norm as the penalty was used for feature selection (also known as LASSO penalty) (Zhu, Rosset, Hastie, & Tibshirani, 2003). The L1 norm of the feature weights, that is, the sum of the feature weights, is used in this model as a regularizer. The difference between L1 and L2 norms is best imagined visually; given a point in two-dimensional space, the L2 norm is the diagonal distance from the origin to that point, while the L1 norm is the distance from the origin as drawn by two perpendicular lines two perpendicular lines aligned with the coordinate axes. For this reason, the L1 norm is sometimes referred to as a Manhattan or Taxicab norm, as a competent taxi driver

would, one hopes, follow straight roads to a destination rather than barreling a Euclidian diagonal, (or via any random pair of perpendicular line) through parks and buildings, no matter how much shorter that route.

The linear SVM is defined to solve the following optimization problem consisting of the penalty and loss terms (Zhu, Rosset, Hastie, & Tibshirani, 2003):

$$\min_{w} \frac{1}{2} \|\boldsymbol{w}\|_n + C \sum_{i=1}^{m} loss(y_i, f_w(\boldsymbol{x}_i))$$

where $\boldsymbol{w} \in \mathbb{R}^d$ is the feature weight vector (normal of the separating hyperplane), $\|*\|_n$ is the norm used, *C* is a cost parameter controlling the effect of the error term, $\{\boldsymbol{x}_i, y_i\} \in \mathbb{R}^d$are the *m* training examples and their binary labels, respectively, and $f_w(\boldsymbol{x}_i) = w \cdot \boldsymbol{x}_i$ is the classifier that is learnt. In this paper we use the hinge loss, $\max(0,1 - y * f(\boldsymbol{x}))^n$, as the loss function.

The L1 norm SVM, as opposed to the standard L2 norm SVM, (i.e., using $\|\boldsymbol{w}\|_1$ for regularization in the above formula) tends to yield sparse features (many feature weights close to zero), which makes it applicable for feature selection. The L1 norm SVM is also able to handle a larger number of irrelevant features than the L2 norm SVM without overfitting (Ng, 2004). Here we used a multiclass one-versus-all strategy for classification, which yields a feature weight matrix with a value for each class/feature. Typically, to determine which features are selected, the L1 norm is computed across the classes, and features with norm higher than a specified threshold, are retained. However, instead of using a norm threshold as a parameter, here we used the number of features to retain as the parameter to select features having the highest norm. The optimal sigma values, when optimized over all RBF 3D velocity features and samples ranged from 0.41 to 12.1, (mean=5.05, std=2.15). The free model parameters of all stages were optimized based on each training fold in the cross-validation, independently from the respective test folds.

2.2.3 Classification

Linear SVM with L2 (Euclidian norm) penalty was used to classify the data based on the selected features. The L2 norm SVM, in which the regularizer term is calculated as the square root of the sum of squares, is generally more efficient at handling a data where feature redundancy is not of concern (Ng, 2004).

For participant classification, a leave-one-genre-out cross validation was employed. This enables us to see how well the model learns participant-specific, idiosyncratic movement patterns that generalize to new musical genres. For genre classification, an 8-fold cross validation was employed to have an equal number of folds as in the participant classification (as there were eight genres). Here at each fold, the data was split so that the participants in the respective training and test sets were not overlapping. This enabled us to see how well the model learns genre-typical movement patterns that generalize to new participants

3. Results

3.1 Model Classification Accuracy

The cost parameter was set to the default value 1.0, and the number of features retained varied exponentially between 1 and 256. For genres, using the linear covariance features, the highest accuracy of 23.5% was reached at 26 features, and adding more features sets into 20% accuracy level. For an 8-class classification problem, chance level accuracy is 12.5%. While our model did provide accuracy above chance level, this accuracy is low compared to classification of genre based on audio signal (Holzapfel & Stylianou, 2006; Kujala et al., 2009; Tzanetakis & Cook, 2002). For participants, the highest obtained accuracies are at the 80-85% range, well above the 1.37% chance accuracy for a 73-class classification task.

When using the RBF covariance features the machine learning pipeline was identical, except for the addition of the kernel optimization stage. For genres, results were the same as for linear covariance. For participants, the results using Correntropy (RBV) matrixes show that adding more features improves the results until convergence after 107 features at a notably high 94-95% level. At 107

features, the accuracy of classification was 94.1%, approximately 10% higher than the accuracy found for linear covariance features. It is notable how much better our model performed when classifying individuals rather than the genres to which they were moving, especially given the comparatively higher level of chance-accuracy for the genre problem (12.5% compared to 1.37%). These results are shown in Figure 3.

[Insert Figure 3 about here]

3.2 Evaluation of Model Feature Selection by Comparison to PCA

In further analysis only correntropy matrixes were used due to their better performance over linear features. To confirm the feasibility of the L1 (Taxicab)-norm SVM-based feature selection, a cross-validation experiment was conducted replacing the feature selection stage by Principal Components Analysis (PCA), where the correntropy matrixes were projected to its principal components. The range of number of components used was the same as that for the number of selected features in the experiment above. The results, shown in Figure 4, show that the PCA projection yields lower or similar accuracies than feature selection. The main benefit of the feature selection is the easier interpretability of classifier model features.

[Insert Figure 4 about here]

3.3 Evaluation of Model Cost (C) Parameter

The cost parameter C in SVM controls the cost of misclassification (by weighting the error term, and consequently down-weighting the regularization), and therefore lower C values may lead to more complex models which cannot then be easily generalized to other sets of data (i.e., the OSH would too narrowly conform around the individual data points specific to this set). The C parameter was in our initial analysis set to the default value C=1.0. In a typical machine learning scenario, the C parameter is optimized with respect to model accuracy using a hyper parameter optimization strategy such as grid search (varying the values) in an inner cross-validation loop. To see how the C value affects the participant classification results, the value was varied from 0.01 to 1000. As shown

by the results, which can be seen in Figure 5, the value has only minor effect on the results, and the default value is close to the optimal. This shows that a model with medium complexity is feasible for the classification task.

[Insert Figure 5 about here]

In summary, as shown by the various trials with different model parameters and the comparisons to other feature sets, the results demonstrate the high robustness of the kernel covariance features at representing dance move patterns for participant classification, and the feasibility of the classification pipeline employed.

3.4 Accuracy by Participants and Genres

If a participant's dance moves are idiosyncratic, the patterns of their movements should be invariable across genres, and therefore should yield high individual classification accuracy for that participant. Conversely, such non-genre-specific movement patterns would likely result in low genre classification accuracy for that participant. To examine this relationship between the participant and genre classification accuracies, we computed correlation between the participant-wise accuracies of participant and genre classification. The correlations are significantly negative, as expected, $r = -.31$, $p < .01$.

Correspondingly, if a specific genre elicits genre-typical movement patterns, this should yield high genre classification accuracies of that genre, and conversely, low participant classification accuracies of that genre. This is demonstrated by the genre-wise participant and genre classification accuracies, where Metal and Jazz were found to elicit the most genre-typical dance moves, and vice versa for Dance, Pop, and Blues, the results of which can be seen in Figure 6.

[Insert Figure 6 about here]

3.5 Importance of Features

To further explore the unexpectedly accurate classification of individual participants, we chose to examine which features optimally classify participants. The participant classification model pipeline

with the 107 selected kernel covariance features was run on the full data set, and the feature importance scores were computed as the L2 norms of the L2 norm SVM classifier feature weights over the classes (participants). The results can be seen in Figure 7:

[Insert Figure 7 about here]

The results show a very general pattern of relationships between medio-lateral (ML) and anterior-posterior (AP) movement across various markers as distinguishing features between dancers, while marker movement in the vertical (V) direction was distinguishable more often in relation to other vertical movement. The markers/dimensions which were overall most important for classification of individuals, calculated by taking the mean over feature matrix columns, are shown in Table 2:

[Insert Table 2 about here]

4. Discussion

The current paper employed a novel approach to exploring the relationship between musical genre and music-induced movement. Participants' free, improvised dance movements were captured while dancing to musical stimuli representing eight genres selected using a data-driven, social-tagging method. Both linear covariance and non-linear correntropy were calculated from dancers' whole-body movements for each stimulus, and an established machine learning algorithm was applied with the aim of classifying the correct genre to which the movement was generated and the correct participant who generated the movement. Contrary to our expectations, person classification was notably more accurate than genre classification, despite chance level being much lower for person classification (1.37% compared to 12.5%). However, some genres were also more recognizable than others from dancers' movements. To the authors' knowledge, this is the first study to attempt to classify genres or individuals in the context of free dance movement.

Greater classification accuracy was achieved using correntropy, calculated using an RBF kernel, than when using the linear covariance as a movement feature. This suggests that nonlinear features are particularly able to capture relevant identifying characteristics of movement. In this context, we can interpret the correntropy measure as being defined more by similarities between markers than by dissimilarities; that is, when markers are moving dissimilarly, it is of little consequence exactly how much their two time-series differ. If, for instance, the RBF kernel used is very narrow (that is, the curve represents a steep decay), correntropy can be thought of as measuring the proportion of time for which the two time-series are very similar to each other.

Human capacity to identify individuals from point light displays of movement has been explored in the literature (Swedish & Troje, 2007; Troje & Chang, 2013; Troje et al., 2005; Ueda, Yamamoto, & Watanabe, 2018), as have kinematic features that commonly characterize traits such as gender, mood, and personality in human movement (Bhowmik, Ghosh, Debsinha, Kajal, & Professor, 2016; Røislien et al., 2009; Troje et al., 2005). However, computational classification of individuals based on movement features has largely been studied in relation to practical applications, such as security surveillance, where video data is more commonly used than three-dimensional motion capture data, necessitating the inclusion of shape and color features (Bhowmik et al., 2016; Moeslund, Hilton, & Krüger, 2006). The current study is the first known to the authors to classify individuals using only movement-related features derived from a free dance movement setting. The surprisingly accurate results suggest that individuality is partly encoded into the covariance between the three-dimensional movement of certain body parts.

The model achieved its best fit at 107 features, most of which were related to the head and limbs. Analysis of the chosen features revealed that head, shoulder and knees were important markers in distinguishing between individuals, and that discriminative features often occurred between adjacent joints (e.g. the right wrist and right elbow) within the same dimension and between key left and right markers such as the shoulders and knees. Although the complexity of this

model creates challenges for interpretation, the general picture that emerges is a mathematically and mechanically reasonable explanation but still a surprisingly telling one in actuality. Of course, in theory, different individuals' movements may covary differently between any markers in any dimensions, but as it is highly unlikely that participant' were consciously controlling these aspects of their movements, the fact that these movement features could be used to accurately classify individuals across various musical stimuli suggests that we each have our own 'motoric fingerprint' which is evidenced in our free dance movements, regardless of what music is playing.

Just what genre of music *was* playing was, against our expectations, not classified very accurately. Although our model did manage to perform better than chance for an 8-class classification problem, at best its overall accuracy rate was less than 25%, well below accuracy rates for most models which classify genre from acoustic signals, which often have accuracy rates of more than 80% (e.g., Holzapfel & Stylianou, 2006; Nanni et al., 2016). The results suggest that an individual's motoric fingerprint has a stronger influence on her dance movement than the specific music to which she is dancing. This results is supported by previous work, for example that of Troje & Chang (2013), showing that there are highly individual characteristics of biological movement, and also by work showing a large degree of consistency of individual movement strategy in skilled drummers across conditions (Dahl, 2011; Danielsen, Waadeland, Sundt, & Witek, 2015) However, these results do not necessarily indicate that the body is not involved in parsing sound or genre. Rather, comparison of the current results with previous work involving classification of genre from audio signals reveal some interesting parallels that can be interpreted through a lens of embodied music cognition.

The results show that some genres were more successfully classified by examining dancers' movements than others, namely Metal (53% accuracy) and Jazz (35% accuracy). This is in line with previous audio-based classification studies, which have routinely shown differences between genres in ease of classification. Tzanetakis and Cook (2002), for example, found that Jazz was classified

noticeably better than other genres. Holzapfel and Stylianou (2006) report the highest classification rate for Metal, as do Wülfing and Riedmiller (2012). Metal and Jazz (along with Classical) were also classified most accurately by Hartmann et al. (2013). These similarities arise despite differences in the acoustic features and classification methods used, corroborating the evidence that there is indeed something particularly 'genre-like' about these genres, especially compared to genres that have regularly proven more difficult to classify, such as Rock and Pop. One interpretation, in keeping with a framework of embodied music cognition as conceptualized in the works of Leman (2008) and Godøy et al. (2016) is that these genres tend to supply listeners with perceptually distinctive audio stimuli than that of some other genres, naturally eliciting different movement patterns that distinguish the results from the less differentiable Pop and Dance. Jazz music tends to feature a fairly unique set of instruments—keyboard, drums, bass, and saxophone are archetypal to jazz—while Metal, though sharing the guitar, drums, bass and vocals of many Rock-related styles, is characterized by noisiness and spectral rolloff the audio signal (Ajoodha, Klein, & Rosman, 2017). One might question whether these characteristic timbral features are necessarily processed by listeners through different movements than others. It is also worth asking whether the sonic forms afforded by Jazz and Metal are particularly distinctive from those of other genres, making the empathic imitation of them similarly more distinctive?

Although the unique musical features of Metal and Jazz undoubtedly do play a role in influencing movement, it is important to mention the obvious extramusical factors that may also have affected participants responses to these genres. Some genres have been previously associated with stereotypical movements; Luck et al. (2010), for example, found evidence that Techno, Latin and Metal were all associated with specific movement patterns, the latter with recognizable 'headbanging' patterns. Although this could again relate to the acoustic qualities of the heard music, the role of culture in driving such stereotypes must be considered in interpreting the current results. Musical genres often exist as part of a culture that includes visual elements, such as clothing or

makeup, and associated dance movements (e.g., Jaimangal-Jones, Pritchard, & Morgan, 2015). Familiarity with these visual elements of culture is likely to have influenced participants' embodied responses to some genres.

Metal seems to be a central piece of a definite subculture, which has been analyzed from various sociological perspectives (Bryson, 1996; Lacourse, Claes, & Villeneuve, 2001; Straw, 1984). Snell and Hodgetts (2007) explored the formation of identity within the social context of Metal music scenes, and found a close association between stereotypical movements such as headbanging and participants' sense of communal bonding and shared experience. They refer to 'the sharing of [Metal] music through dance as a way of reaffirming a sense of belonging, shared experience and support' (p. 434). The current results suggest that, while dancing to Metal, participants moved in more similar ways to one another, allowing for more accurate classification, possibly supporting the idea of stereotypical dance in Metal serving identification with a group. Jazz, which has its origins in its own subculture, has been historically associated with a number of distinctive dance movements, such as the Charleston and other types of swing dancing, which through revivals and specific efforts at cultural preservation are not unlikely to be at least cursorily familiar to Western listeners today (Monaghan, 2002; Lena & Peterson, 2008). These influences may be less prominent in responses to genres that are more mainstream, such as Pop, Hip-Hop and Dance, which may account for the differences in identification accuracy of these genres compared to Metal and Jazz.

Given these cultural characteristics, it is possible that participants' movements were affected by familiarity with norms specific to the subcultures invoked by these two musical genres, quite possibly even if they themselves did not identify with that culture. Familiarity may arise, for example, through exposure to music videos, films, and other types of visual media, as well as through direct engagement with others in the cultural context. This does not need to be considered a conflicting interpretation to the idea that particular acoustic features were embodied by participants

while listening to these genres, in line with a framework of embodied music cognition. It does, however, invite consideration of the idea that an embodied response to music—that is, the use of bodily movement to process and parse the incoming audio signals—means, in naturalistic settings at least, the embodiment of music on many levels of abstraction higher than sound. Previous work has shown that emotion, mood, and personality are embodied by our music-induced movements (Camurri, Lagerlöf, & Volpe, 2003; Dyck, Maes, Hargreaves, Lesaffre, & Leman, 2013; Luck et al., 2010), as well as our own cultures (Himberg & Thompson, 2011). It is not a very far leap, then, to expect that cultural information embedded into musical genre is also embodied. Viewed in this light, the process of dancing to a song is not only the process of interaction between a complex acoustic signal and an even more complex human nervous and muscular system; it is simultaneously the interaction of a unique person and their memories, beliefs, and preferences with a culturally-defined set of extramusical associations and expectations (Shevy, 2008). Finnish participants, who comprised a majority of the current study's participants, may be particularly familiar with, and therefore predisposed to enjoy, Metal music (Carlson et al., 2017), which could have yielded a more prominent embodied manifestation of Metal culture in the current results than would be found elsewhere in the world.

The current results suggest that the unique role of the individual and the role not only the participants' culture, but the cultural affordance of the music itself, should be taken into consideration in designing and interpreting studies related to embodied music cognition. These results also show that, while analysis focused on identifying group means and general tendencies are a common approach for such studies, there are both quantitative and theoretical insights to be gained from the application of analysis methods which highlight individual differences. The notable individuality of movement patterns shown here should be explored with further research, for example by using stimulus manipulations other than genre, or considering individual differences at the level of personality or culture. Future research is also necessary to examine genres and their

associated dance movements at the level of sub-genre, to explore the relationships between genre preferences and movement patterns, and to explore embodied responses to audio excerpts that are ambiguous or multi-faceted in terms of genre. The current results regarding genre also merit further research, particularly of the influences of music preference, culture, and familiarity with a given genre and its extramusical associations.

This work was supported by funding from the Academy of Finland, project numbers 272250, 299067 and 274037.

References

Ajoodha, R., Klein, R., & Rosman, B. (2017). Single-labelled Music Genre Classification Using Content-Based Features. In C. Senac, T. Pellegrinit, F. Mouret, & J. Pinquier (Eds.), *Proceedings of the 15th International Workshop on Content-Based Multimedia Indexing. Florence, Italy*: University of Toulouse, Toulouse, France.

Aucouturier, J.-J., & Pachet, F. (2003). Representing Musical Genre: A state of the art. *Journal of New Music Research, 32*(1), 83–93. https://doi.org/10.1076/jnmr.32.1.83.16801

Bhowmik, S., Ghosh, A. K., Debsinha, J., Kajal, R., & Professor, A. (2016). A Literature Survey on Human Identification by Gait. *Imperial Journal of Interdisciplinary Research, 2*(7), 2454–1362. Retrieved from http://www.onlinejournal.in

Bläsing, B. E., & Sauzet, O. (2018). My action, my self: Recognition of self-created but visually unfamiliar dance-like actions from point-light displays. *Frontiers in Psychology, 9*(10), 1–9. https://doi.org/10.3389/fpsyg.2018.01909

Bryson, B. (1996). " Anything But Heavy Metal ": Symbolic Exclusion and Musical Dislikes. *American Sociological Review, 61*(5), 884–899.

Burger, B., Thompson, M. R., Luck, G., Saarikallio, S., & Toiviainen, P. (2013). Influences of Rhythm- and Timbre-Related Musical Features on Characteristics of Music-Induced Movement. *Frontiers in Psychology, 4,* 183. https://doi.org/10.3389/fpsyg.2013.00183

Burger, B., Thompson, M. R., Saarikallio, S., Luck, G., & Toiviainen, P. (2010). Influence of musical features on characteristics of music-induced movements. In S. M. Demorest, S. J. Morrison, & P. S. Campbell (Eds.), *Proceedings of the 11th International Conference on M usic perception and Cognition (ICMPC11)* (pp. 425–428). Seattle, Washington.

Burger, B., & Toiviainen, P. (2013). MoCap Toolbox -A Matlab toolbox for computational analysis of movement data. In R. Bresin (Ed.), *Proceedings of the Sound and Music Computing Conference* (pp. 172–178).

Burger, B., & Toiviainen, P. (2018). Embodiment in Electronic Dance Music: Effects of musical content and structure on body movement. *Musicae Scientiae*. https://doi.org/10.1177/1029864918792594

Camurri, A., Lagerlöf, I., & Volpe, G. (2003). Recognizing emotion from dance movement: comparison of spectator recognition and automated techniques. *International Journal of Human-Computer Studies, 59*(1), 213–225.

Carlson, E., Burger, B., London, J., Thompson, M. R., & Toiviainen, P. (2016). Conscientiousness and Extraversion relate to responsiveness to tempo in dance. *Human Movement Science, 49*, 315–325. https://doi.org/10.1016/j.humov.2016.08.006

Carlson, E., Burger, B., & Toiviainen, P. (2018). Dance Like Someone is Watching. *Music & Science, 1,* 205920431880784. https://doi.org/10.1177/2059204318807846

Carlson, E., Saari, P., Burger, B. B., & Toiviainen, P. (2017). Personality and musical preference using social-tagging in excerpt-selection. *Psychomusicology: Music, Mind, and Brain, 27*(3), 203–212. https://doi.org/10.1037/pmu0000183

Cross, I., Fitch, W. T., Aboitiz, F., Iriki, A., Jarvis, E. D., Lewis, J., … Trehub, S. E. (2001). Culture and evolution. *Annals of the New York Academy of Sciences, 930*(1), 28–42. https://doi.org/10.7551/mitpress/9780262018104.003.0021

Cutting, J. E., & Kozlowski, L. T. (1977). Recognizing friends by their walk: Gait perception without familiarity cues. *Bulletin of the Psychonomic Society, 9*(5), 353–356.

Monaghan, T. (2002). Why study the lindy hop*? Dance Research Journal, 33*(2), 124–127.

Danielsen, A., Waadeland, C. H., Sundt, H. G., & Witek, M. A. (2015). Effects of instructed timing and tempo on snare drum sound in drum kit performance. *The Journal of the Acoustical Society of America*, *138*(4), 2301-2316.

Dahl, S. (2011). Striking movements: A survey of motion analysis of percussionists. *Acoustical Science and Technology*, *32*(5), 168-173.

Dyck, E. Van, Maes, P.-J., Hargreaves, J., Lesaffre, M., & Leman, M. (2013). Expressing induced emotions through free dance movement. *Journal of Nonverbal Behavior, 37*(3), 175–190.

Dyck, E. Van, Moelants, D., Demey, M., Deweppe, A., Coussement, P., & Leman, M. (2013). On the impact of the bass drum on human dance movement. *Music Perception, 30*(4), 349–359.

Ergezer, H., & Leblebicioğlu, K. (2018). Time series classification with feature covariance matrices. *Knowledge and Information Systems.* https://doi.org/10.1007/s10115-017-1098-1

Estes, W. K. (1962). Learning Theory. *Annual Review of Psychology, 13(*1), 107–144. https://doi.org/10.1146/annurev.ps.13.020162.000543

Genussov, M., & Cohen, I. (2010). Musical genre classification of audio signals using geometric methods. *European Signal Processing Conference, 10*(5), 497–501.

Godøy, R. I., Song, M., Nymoen, K., Haugen, M. R., & Jensenius, A. R. (2016). Exploring sound-motion similarity in musical experience. *Journal of New Music Research, 45*(3), 210–222. https://doi.org/10.1080/09298215.2016.1184689

Gunns, R. E., Johnston, L., & Hudson, S. M. (2002). Victim selection and kinematics: A point-light investigation of vulnerability to attack. *Journal of Nonverbal Behavior, 26*(3), 129–158. https://doi.org/10.1023/A:1020744915533

Guo, G., Li, S. Z., & Chan, K. (2000). Face recognition by support vector machines. *Proceedings Fourth IEEE International Conference on Automatic Face and Gesture Recognition.* Grenoble, France: IEEE.

Guo, K., Ishwar, P., & Konrad, J. (2009). Action recognition in video by covariance matching of silhouette tunnels. *Proceedings of SIBGRAPI 2009 - 22nd Brazilian Symposium on Computer Graphics and Image Processing*. https://doi.org/10.1109/SIBGRAPI.2009.29

Hartmann, M., Saari, P., Toiviainen, P., & Lartillot, O. (2013). Comparing timbre-based features for musical genre classification. *Proceedings of the International Conference on Sound and Music Computing,* 707–714.

Himberg, T., & Thompson, M. R. (2011). Learning and synchronising dance movements in South African songs - Cross-cultural motion-capture study. *Dance Research, 29(*2), 305–328. https://doi.org/10.3366/drs.2011.0022

Holzapfel, A., & Stylianou, Y. (2006). Musical Genre Classification using Nonnegative Matrix Factorization-Based Features. *Robotics & Automation Magazine, IEEE, 13*(2), 99–110. https://doi.org/10.1038/nature09973

Johansson, G. (1973). Visual perception of biological motion and a model for its analysis. *Perception & Psychophysics*, *14*(2), 201–211. https://doi.org/10.3758/BF03212378

Jaimangal-Jones, D., Pritchard, A., & Morgan, N. (2015). Exploring dress, identity and performance in contemporary dance music culture. *Leisure Studies*, *34*(5), 603-620.

Karkavitsas, G. V., & Tsihrintzis, G. A. (2011). Automatic music Genre Classification using hybrid Genetic Algorithms. *Smart Innovation, Systems and Technologies, 11* SIST, 323–335. https://doi.org/10.1007/978-3-642-22158-3_32

Kujala, J., Aho, T., & Elomaa, T. (2009). A walk from 2-norm SVM to 1-norm SVM. *Proceedings - IEEE International Conference on Data Mining, ICDM*, 836–841. https://doi.org/10.1109/ICDM.2009.100

Lacourse, E., Claes, M., & Villeneuve, M. (2001). Heavy metal music and adolescent suicidal risk. *Journal of Youth and Adolescence, 30*(3), 321–332.

Lakoff, G., & Johnson, M. (1999). *Philosophy in the Flesh*. New York: Basic books.

Leman, M. (2008). *Embodied music cognition and mediation technology*. MIT Press.

Lena, J. C., & Peterson, R. A. (2008). Classification as culture: types and trajectories of music genres. *American Sociological Review, 73*(5), 697–718.

Lesaffre, M., De Voogdt, L., Leman, M., De Baets, B., De Meyer, H., & Martens, J. P. (2008). How potential users of music search and retrieval systems describe the semantic quality of music.

Journal of the American Society for Information Science and Technology, 59(5), 695–707. https://doi.org/10.1002/asi.20731

Liu, W., Member, S., Pokharel, P. P., & Principe, J. C. (2007). Correntropy: Properties and Applications in Non-Gaussian Signal Processing. *IEEE Transactions on Signal Processing,* 55, 5286–5298.

Luck, G., Saarikallio, S., Burger, B., Thompson, M. R., & Toiviainen, P. (2010). Effects of the Big Five and musical genre on music-induced movement. *Journal of Research in Personality, 44*(6), 714–720. https://doi.org/10.1016/j.jrp.2010.10.001

Mace, S. T., Wagoner, C. L., Teachout, D. J., & Hodges, D. A. (2011). Genre identification of very brief musical excerpts. *Psychology of Music, 40*(1), 112–128.

Mamonne, A., Turchi, M., & Cristianini, N. (2009). Support vector machines. *Wiley Interdisciplinary Reviews: Computational Statistics, 1*(3), 283–289.

Michalak, J., Troje, N. F., Fischer, J., Vollmar, P., Heidenreich, T., & Schulte, D. (2009). Embodiment of sadness and depression--gait patterns associated with dysphoric mood. *Psychosomatic Medicine, 71*(5), 580–587. https://doi.org/10.1097/PSY.0b013e3181a2515c

Moeslund, T. B., Hilton, A., & Krüger, V. (2006). A survey of advances in vision-based human motion capture and analysis. *Computer Vision and Image Understanding, 104*(2-3 SPEC. ISS.), 90–126. https://doi.org/10.1016/j.cviu.2006.08.002

Nanni, L., Costa, Y. M. G., Lumini, A., Kim, M. Y., & Baek, S. R. (2016). Combining visual and acoustic features for music genre classification. *Expert Systems with Applications, 45*, 108–117. https://doi.org/10.1016/j.eswa.2015.09.018

Nettl, B. (2000). An ethnomusicologist contemplates universals in musical sound and musical culture. In *The origins of music* (pp. 463–472). Cambridge, MA: MIT Press.

Ng, A. Y. (2004). Feature selection, L1 vs. L2 regularization, and rotational invariance. *Proceedings of the 21st International Conferences on Machine Learning.*

Pachet, F., & Cazaly, D. (2000). A taxonomy of musical genres. *Content-Based Multimedia Information Access Conference,* 1238–1245. Paris, France.

Poppe, R. (2010). A survey on vision-based human action recognition. *Image and Vision Computing, 28*(6), 976–990. https://doi.org/10.1016/j.imavis.2009.11.014

Richter, J., & Ostovar, R. (2016). "It Don't Mean a Thing if It Ain't Got that Swing"– an alternative concept for understanding the evolution of dance and music in human beings. *Frontiers in Human Neuroscience, 10*(October), 1–13. https://doi.org/10.3389/fnhum.2016.00485

Røislien, J., Skare, Ø., Gustavsen, M., Broch, N. L., Rennie, L., & Opheim, A. (2009). Simultaneous estimation of effects of gender, age and walking speed on kinematic gait data. *Gait & Posture, 30*(4), 441–445.

Satchell, L., Morris, P., Mills, C., O'Reilly, L., Marshman, P., & Akehurst, L. (2017). Evidence of Big Five and Aggressive Personalities in Gait Biomechanics. *Journal of Nonverbal Behavior, 41*(1), 35–44. https://doi.org/10.1007/s10919-016-0240-1

Scaringella, N., Zoia, G., & Mlynek, D. (2006). Automatic genre classification of music content. *IEEE Signal Processing Magazine, 23*(2), 133–141.

Sevdalis, V., & Keller, P. E. (2009). Self-recognition in the perception of actions performed in synchrony with music. *Annals of the New York Academy of Sciences, 1169,* 499–502.

Shapiro, L. (2007). The Embodied Cognition Research Programme. *Philosophy Compass, 2*(2), 338–346. https://doi.org/10.1111/j.1747-9991.2007.00064.x

Shevy, M. (2008). Music genre as cognitive schema: extramusical associations with country and hip-hop music. *Psychology of Music, 36*(1997), 477–498.

Snell, D., & Hodgetts, D. (2007). Heavy metal, identity and the social negotiation of a community practice. *Journal of Community & Applied Social Psychology, 17,* 430–434.

Solberg, R. T., & Jensenius, A. R. (2017). Pleasurable and intersubjectively embodied experiences of electronic dance music. *Empirical Musicology Review, 11*(3–4), 1–17.

Sordo, M., Celma, O., Blech, M., & Guaus, E. (2008). The quest for musical genres: Do the experts and the wisdom of crowds agree? *9th International Conference on Music Information Retrieval,* 255–260.

Straw, W. (1984). Characterizing rock music cultures: The case of heavy metal. *Canadian University Music Review, (5),* 104. https://doi.org/10.7202/1013933ar

Swedish, W., & Troje, N. F. (2007). Kinematic cues for person identification. *Perception & Psychophysics, 69*(2), 241–253.

Tolonen, T., Member, S., & Karjalainen, M. (2000). A computationally efficient multipitch analysis model - *Speech and Audio Processing, IEEE Transactions on. 8*(6), 708–716.

Troje, N. F., & Chang, D. H. F. (2013). Shape-independent processing of biological motion. *People Watching: Social, Perceptual, and Neurophysiological Studies of Body Perception.* https://doi.org/10.1093/acprof:oso/9780195393705.003.0006

Troje, N. F., Westhoff, C., & Lavrov, M. (2005). Person identification from biological motion: Effects of structural and kinematic cues. *Perception and Psychophysics, 67*(4), 667–675. https://doi.org/10.3758/BF03193523

Tuzel, O., Porikli, F., & Meer, P. (2008). Pedestrian detection via classification on Riemannian manifolds. *IEEE Transactions on Pattern Analysis and Machine Intelligence.* https://doi.org/10.1109/TPAMI.2008.75

Tzanetakis, G., & Cook, P. (2002). Musical genre classification of audio signals. *In IEEE Transactions on Speech and Audio Processing.* https://doi.org/10.1109/TSA.2002.800560

Ueda, H., Yamamoto, K., & Watanabe, K. (2018). Contribution of global and local biological motion information to speed perception and discrimination. *Journal of Vision, 18*(3), 2. https://doi.org/10.1167/18.3.2

Wilson, A. D., & Golonka, S. (2013). Embodied cognition is not what you think it is. *Frontiers in Psychology, 4.*

Wülfing, J., & Riedmiller, M. (2012). Unsupervised learning of local features for music classification. *Proceedings of the 13th International Society for Music Information Retrieval Conference (ISMIR 2012),* 139–144. https://doi.org/10.1016/j.msea.2015.07.064

Zhu, J., Rosset, S., Hastie, T., & Tibshirani, R. (2003). 1-norm Support Vector Machines. NIPS'03 *Proceedings of the 16th International Conference on Neural Information Processing Systems, 49–46.* Whister, BC, Canada: MIT press.