TOWARDS MULTIMODAL MIR: PREDICTING INDIVIDUAL DIFFERENCES FROM MUSIC-INDUCED MOVEMENT

by

Yudhik Agrawal, Samyak Jain, Emily Carlson, Petri Toiviainen, Vinoo Alluri

in

Centre for Cognitive Science
International Institute of Information Technology
Hyderabad - 500 032, INDIA
October 2020

TOWARDS MULTIMODAL MIR: PREDICTING INDIVIDUAL DIFFERENCES FROM MUSIC-INDUCED MOVEMENT

Yudhik Agrawal[1]   **Samyak Jain**[1]   **Emily Carlson**[2]
Petri Toiviainen[2]   **Vinoo Alluri**[1]

[1] Cognitive Science Lab, International Institute of Information Technology, Hyderabad, India
[2] Department of Music, Art and Culture Studies, University of Jyväskylä, Finland

{yudhik.agrawal, samyak.j}@research.iiit.ac.in, {emily.j.carlson, petri.toiviainen}@jyu.fi
vinoo.alluri@iiit.ac.in

ABSTRACT

As the field of Music Information Retrieval grows, it is important to take into consideration the multi-modality of music and how aspects of musical engagement such as movement and gesture might be taken into account. Bodily movement is universally associated with music and reflective of important individual features related to music preference such as personality, mood, and empathy. Future multimodal MIR systems may benefit from taking these aspects into account. The current study addresses this by identifying individual differences, specifically Big Five personality traits, and scores on the Empathy and Systemizing Quotients (EQ/SQ) from participants' free dance movements. Our model successfully explored the unseen space for personality as well as EQ, SQ, which has not previously been accomplished for the latter. R^2 scores for personality, EQ, and SQ were 76.3%, 77.1%, and 86.7% respectively. As a follow-up, we investigated which bodily joints were most important in defining these traits. We discuss how further research may explore how the mapping of these traits to movement patterns can be used to build a more personalized, multi-modal recommendation system, as well as potential therapeutic applications.

1. INTRODUCTION

From the perspective of most computational analysis, music can be defined as sound, its important features yielding to the decomposition of waveforms. However, for the vast majority of history, musical sound could not be separated from its source; to whatever degree it may have evolved biologically to serve various human functions, music must be regarded as an embodied and socially embedded phenomenon [1–3]. Research has shown intimate links between musical features and human movement, including the reflection of hierarchical rhythmic structures in embodied eigen movements [4], the reflection of higher-level musical structures in group movement to Electronic Dance Movement [5], and reflection of spectral and timbral features of music in dance [6]. Bodily movement is one of the most commonly reported responses to music [7], and movement to music is one of very few universal features of music across cultures [8].

This paper towards Multimodal MIR takes into consideration the multi-modality of music, and takes into account one of the primary aspect of musical engagement, i.e, movement. It is therefore insufficient to consider music only in terms of sound when trying to understand human digital use and interaction with music. This may be especially true in terms of user experience and personalization; human movement in response to music reflects not only the music itself but characteristics of the individual, such as personality [9] and emotion [10]. Indeed, research has shown that music-induced movement is so individual that its features can be used in person-identification with a high degree of accuracy [11]. This is in line with previous research, such as that of Cutting et al. [12] demonstrating that friends can recognise each other from their walk with only point-light (stick figure) displays of movement, without the need for other distinguishing features. This paradoxical balance between universality and individuality in human motoric responsiveness to music poses a challenge for the creation of digital music interfaces which take music-induced movement into account in providing personalized music experiences. Although the concept of an interactive music system has long been proposed that allow music playback to be controlled and altered via human gestures [13], human-movement based interaction techniques and devices are fast gaining importance in the field of HCI [14]. In this context, it makes decoding aspects of a user/individual via human movements a key and useful endeavor, which would then aid in the design of more personalized experiences.

2. RELATED WORK

The specific features used in previous work associated movement with individual differences are quite varied. Satchell et al. [15] examined speed, relative and absolute rotation of the body and found relationships between relative movement of the upper and lower body during walk-

ing in both FFM personality traits and gait, while Michalak et al. [16] were able to associate low mood with lateral body sway and posture. In dance, relevant features have included amount of movement of the whole body relative to itself and to the environment, responsiveness to music features such as tempo [4, 17]. Another area for exploration of individual differences in movement patterns has been in the context of disorders that have altered or impaired movement [18–20]. These links allow us to postulate that movement patterns should give us information related to individual traits and tendencies which can be then linked to music preferences, mood or emotion in relation to music experiences and could have implications for music therapy as well as for music information retrieval. However, as an initial step, there exist no studies that predict personality and empathy as a function of movement patterns. The current study focuses on identifying FFM personality traits, as well as scores on the Empathy Quotient (EQ) and Systemizing Quotient (SQ), from participants' free dance movements to various genres of music. The EQ measures participants' tendency to empathize with others [21], while the SQ measures the tendency to think in terms of systems [22]. These two measures were originally developed to increase understanding of people with ASD, as in this population trait systemizing tends to be very high while empathy tends to be low. However previous work has also used the EQ/SQ to determine how these traits are distributed in the general population. Although previous work has found relationships between empathy and responsiveness to changes in heard music or in dance partner [23, 24], and between EQ/SQ scores and music preferences [25, 26], general movement patterns associated with empathy have not, to the knowledge of the authors, been explored using dance movement, nor have patterns related to systemizing tendencies.

3. METHOD

3.1 Participants

Data acquired was from a previous study [27] comprising data from 73 university students (54 females, mean age = 25.74 years, std = 4.72 years). Thirty-three reported having received formal musical training; five reported one to three years, ten reported seven to ten years, while sixteen reported ten or more years of training. Seventeen participants reported having received formal dance training; ten reported one to three years, five reported four to six years, while two reported seven to ten years. 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. Fifteen participants were excluded from further analysis due to incomplete data. They were asked to listen to the music and move as freely as they desired, but staying within the marked capture space. The aim of these instructions was to create a naturalistic setting, such that participants would feel free to behave as

they might in a real-world situation.

3.2 Apparatus, Stimuli, and Procedure

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 position 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 toe. The stimuli comprised sixteen 35-second excerpts from eight genres, in randomized order: Blues, Country, Dance, Jazz, Metal, Pop, Rap, and Reggae. The stimuli for the experiment were selected using a computational process based on social-tagging and acoustic data. 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. Moreover, investigating movements to multiple genres of music further adds to the generalizability of our findings.
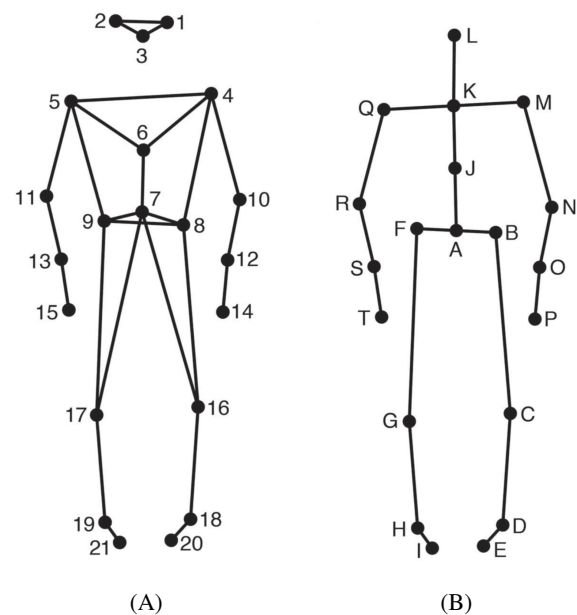

Figure 1: Marker and joint locations (A) Anterior view of the marker locations a stick figure illustration; (B) Anterior view of the locations of the secondary markers/joints used in animation and analysis of the data

3.2.1 Personality and Trait Empathy Measures

The Big Five Inventory (BFI) was used to capture the five predominant personality dimensions, namely, Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism [28]. Trait Empathy was measured using the EQ- and SQ-short form version, developed and validated by Wakabayashi et al. [29], as a result giving an EQ and SQ score per participant.
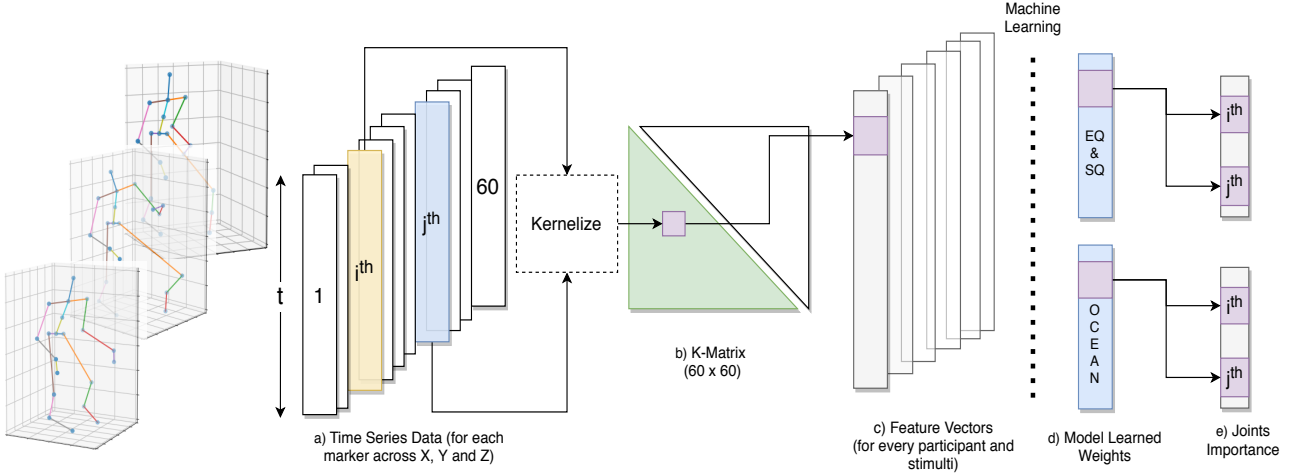
Figure 2: Overview of our Pipeline. Given the position of joints across time frames in 3D Euclidean space(a), we apply pairwise correntropy between time series x_i and x_j and calculate the K-matrix (b). Then, taking the lower triangular part of the symmetric covariance matrix, we get the feature vectors (c). After training the regression model on the feature vectors, we get the weight vector(d). Finally, corresponding weight values from the learned weight vector are mapped to the corresponding joints to get the per-joint importance (e).

3.3 Feature Extraction

The analysis and prediction pipeline is illustrated in Figure 2. To facilitate extraction of kinematic features using the MATLAB Motion Capture (MoCap) Toolbox [30], a set of 20 secondary markers, subsequently referred to as joints, was derived from the original 21 markers. The locations of these 20 joints are depicted in Figure 1. 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 24Hz [30].

The features used in our analysis is the co-variances of position and velocity. The co-variances between all marker time series in each direction (X, Y and Z) within each participant for each stimulus. We used a non-linear measure to calculate covariance between the markers. This method, referred to as correntropy between time series x_i and x_j [31], is given by:

$$K(x_i, x_j) = e^{\frac{-||x_i - x_j||_2^2}{2\sigma^2 T^2}} \qquad (1)$$

where $||x_i - x_j||_2$ is L2-norm between x_i and x_j, σ is a constant, 12.0 in our case and T is the length of the time-series. The L2-norm is divided by T to normalize according to time series length since it has different lengths with varying stimuli. Since the number of joints are 20 and each joint has three coordinates, the dimension of K

would be 60×60. The lower triangular part excluding the diagonal elements of the symmetric covariance matrix was vectorised to produce a feature vector of length 1770 for each participant and for each stimuli.

We also run our experiments using the Normalized feature vectors calculated by using Position and Velocity, we employed standard Gaussian Normalization technique:

$$\hat{X} = \frac{X - \mu(X)}{\sigma(X)} \qquad (2)$$

where $\hat{X}$ is the feature vector, $\mu(X)$ is the mean and $\sigma(X)$ is standard deviation.

3.4 Model Regression

The most common regression model for value prediction tasks used is Linear Regression. The goal here is to find an optimal line that minimizes the total prediction error. But this model treats its parameters as unknown constants whose values must be derived. Moreover, the weights become sensitive when the dataset size is large. So to prevent the model from overfitting, we took principal components of the features to train the model (For the result sections, we will be considering 243 components for position data and 137 components for velocity data which gave us the best results). We also approached this problem by using Bayesian Regression other than Principal Component Regression (PCR) [1] .

In Bayesian Regression the parameters are treated as random variables belonging to an underlying distribution. Depending on the dataset, we can be more or less certain about the weights. Since, the parameters of the model belong to a distribution, the predictions of the model also belong to a distribution. So we have confidence bounds on

---

[1] Detailed analysis of Principal Component Regression (PCR) and Bayesian Regression are discussed in the supplementary.

our predictions. Therefore, they are better at representing the uncertainty of a model's predictions.

3.4.1 Personality and Trait Empathy Prediction

The features extracted are used to train five different Bayesian Regression models to predict each of the five personality traits - Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism (OCEAN). The features extracted are used to train both regression models to predict EQ and SQ respectively. The model is trained and evaluated on the described dataset.

3.5 Visualizing the Weight Vector

To interpret the coefficients (also known as the weights or model parameters) of the regression models, we add the value of the feature vector to the corresponding joints. In our algorithm, we first find the index in the 60×60 matrix and then add the absolute value to those joints.

The sign of the coefficient indicates the direction of the relationship but the magnitude preserves the importance. After that, Min-Max Normalization is applied to bring the values in the range (0, 1) for better visualizing the same variable across similar tasks.

$$\overline{JI}[i] = \left(\frac{JI[i] - min(JI)}{max(JI) - min(JI)} \right) \forall JI[i] \qquad (3)$$

where $\overline{JI}$ represent the normalized Joint Importance Vector. Algorithm 1 describes the pseudo-code to get the importance of joints from the weights of the trained regression model.

Algorithm 1 Joint Importance

Result: Calculate a vector J of 20 dimensions representing importance of each joint.

W is the weight vector; J is the importance vector initialised with 0; S contains lower triangular indices excluding diagonal indices; 0-indexing is followed;

1: $S \leftarrow LowerTriangularIndices(60 \times 60)$
2: $N \leftarrow S.length()$
3: **for** $k = 0 : N - 1$ **do**
4:     $(i, j) := S(k)$
5:     $(\hat{i}, \hat{j}) \leftarrow IndexToJoint(i, j)$
6:     $J(\hat{i}) := J(\hat{i}) + |W(k)|$
7:     $J(\hat{j}) := J(\hat{j}) + |W(k)|$
8: **end for**
9: **return** J

After getting a vector of 20 dimension, we reduce it to 12 before visualizing joint importance. We did this by taking the average of joints which occur in pairs eg. (L shoulder, R shoulder), (L wrist, R wrist).

3.5.1 Evaluation Metric

(a) Root Mean Square Error (RMSE): It computes a risk metric corresponding to the expected value of the root of squared (quadratic) error or loss.

(b) R^2 Score: It represents the proportion of the variance(of y) that has been explained by the independent variables in the model. [2]

As the square root of a variance, RMSE can be interpreted as the standard deviation of the unexplained variance, and has the useful property of being in the same units as the response variable and at the same time the R^2 helps us evaluate the goodness of fit in capturing the variance in training data. We calculate RMSE and R^2 on multiple splits so that we get an average estimate of the accuracy.

3.6 Results

3.6.1 EQ and SQ

The results for EQ prediction are in Table 1 and SQ prediction are in Table 2. The results are calculated using 5-fold cross validation. The range of EQ and SQ is 0-80. The boldface values represent the best score. The 'N' in the tables denote that Gaussian Normalization was applied on the features. We trained and evaluated two different models for each of the aforementioned tasks. We can see that using position data, instead of velocity data, to generate the feature vectors, gave us the best results. Also, we can see that the Bayesian Regression gave better results than Principal Component Regression on both metrics. So from here on, we will be using Bayesian Regression for other prediction and analysis tasks.

Input	PCR		Bayesian Ridge	
	RMSE	R^2	RMSE	R^2
Position	3.071	0.708	**2.722**	**0.771**
Position(N)	3.201	0.684	2.733	0.765
Velocity	4.938	0.249	4.343	0.423
Velocity(N)	4.583	0.353	4.015	0.503

Table 1: Prediction Results for Empathizing Quotient

Input	PCR		Bayesian Ridge	
	RMSE	R^2	RMSE	R^2
Position	2.398	0.781	**2.161**	**0.867**
Position(N)	2.363	0.786	2.502	0.838
Velocity	4.448	0.252	3.832	0.469
Velocity(N)	4.211	0.329	3.714	0.552

Table 2: Prediction Results for Systemizing Quotient

3.6.2 Personality Regression

The results for OCEAN value prediction on Dataset can be found in Table 3. The results are calculated using 5-fold cross validation. The range of the personality values is 1.0-5.0. We can see that using position data to extract features gave the best results on predicting all five personality traits on the dataset. We can concur that using position data instead of velocity data in the kernelized space is better for these regression tasks.

---

[2] Detailed explanation of metrics is provided in the supplementary.

Input	Openness		Conscientiousness		Extraversion		Agreeableness		Neuroticism	
	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2
Position	**0.197**	**0.776**	**0.317**	**0.760**	**0.384**	0.743	**0.252**	**0.776**	**0.384**	**0.758**
Position(N)	0.227	0.740	0.332	0.690	0.414	**0.756**	0.273	0.716	0.390	0.739
Velocity	0.332	0.464	0.487	0.415	0.556	0.523	0.440	0.335	0.557	0.483
Velocity(N)	0.304	0.527	0.426	0.543	0.501	0.623	0.408	0.442	0.461	0.654

Table 3: Prediction Results for Five Personality Traits using Bayesian Regression

3.6.3 Joints' Importance

For evaluating joint importance we used learned weights of the model using position data across different tasks. For the purpose of analyzing the importance of the joints, we reduced them to 12 by taking the average for those joints which occur in pairs eg. (L shoulder, R shoulder). This was also done for hips, knee, ankle, toe, elbow, wrist, and finger. Altogether the results in characterizing an individual trait is dominated by the limbs than the core of the body.

From the relative joint importance depicted in Figure 3, we observe that 'Ankle', Elbow' and 'Shoulder' play an important role in determining EQ and SQ of an individual, whereas 'Neck' and 'Torso' have a negligible contribution. We also infer that 'Finger', 'Hip', and 'Knee' are more crucial joints for predicting EQ than for SQ whereas 'Elbow' holds significantly higher importance for predicting SQ than for EQ.

Figure 4 displays the relative joint importance of personality along with the mean plotted in each sub-figure. The farther away from the mean the joint importance value for an individual joint is, the more important it is in characterizing that trait. Some similarities in the joint importance profiles across the personality traits can be attributed to the inherent correlation that exists among them [3] . We observe that it is the 'Finger', 'Elbow', and 'Knees' that contribute to Feature Importance whereas 'Root', 'Neck' and 'Torso' have negligible contribution. For characterizing Conscientiousness, 'Shoulders', 'Knees' and 'Neck' play a crucial role while 'Head' and 'Toe' plays an important role for Extraversion. For Agreeableness, 'Neck' and 'Wrists' have relatively less importance as compared to other joints whereas, 'Wrists' play an important role in Openness. Finally, there are no significant defining features for Neuroticism, which indicates that their expression in Dance Movements through Music-Induced Movements are very limited.
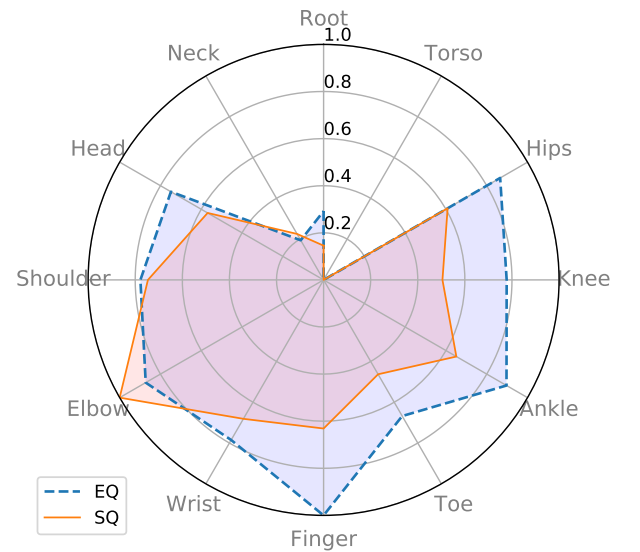
4. DISCUSSION

Music experiences are highly embodied, making it necessary to consider individual embodied responses to music in developing more advanced personalized user experiences. The current study is among the first to the authors' knowledge to use participants' free dance movements to predict personality traits, and both the Empathizing and Systemizing Quotients (EQ/SQ).

Co-variance between joint velocities has previously been used to identify an individual from their free dance movements with a high degree of accuracy [11]. The results of the current analysis show co-variance to be a useful feature in predicting individual differences. However, we achieved considerably better prediction accuracy by using position data than velocity data.

Overall, the limbs of the body seemed to have more importance in predicting individual traits than the core body. This is in line with the fact that gesture plays an important role in communication [32], and as specifically regards the EQ/SQ, as these tests were originally developed in conjunction with studies of ASD, in which gesture and imitative movement appear to be compromised [33]. Although the sample used in the current study comprised typically functioning (non-ASD) participants, the accuracy of prediction of EQ/SQ scores in this analysis is worth highlighting in light of recent work suggesting the existence of motor signatures unique to ASD, detectable from whole body movements as well as data drawn from participants' interaction with tablets [19, 34].

The specific markers that were important in the prediction of individual traits in some cases corroborates previous work, and in some cases contradicts it. Luck et al. [9] found correlations between Extraversion and speed of head movement, which supports the current finding that the head is of particular importance in identifying Extraver-

Figure 3: Relative importance of Joints in EQ and SQ Tasks using the Position Data.

---

[3] The table for Spearman Correlation between the personality traits is provided in the supplementary material.
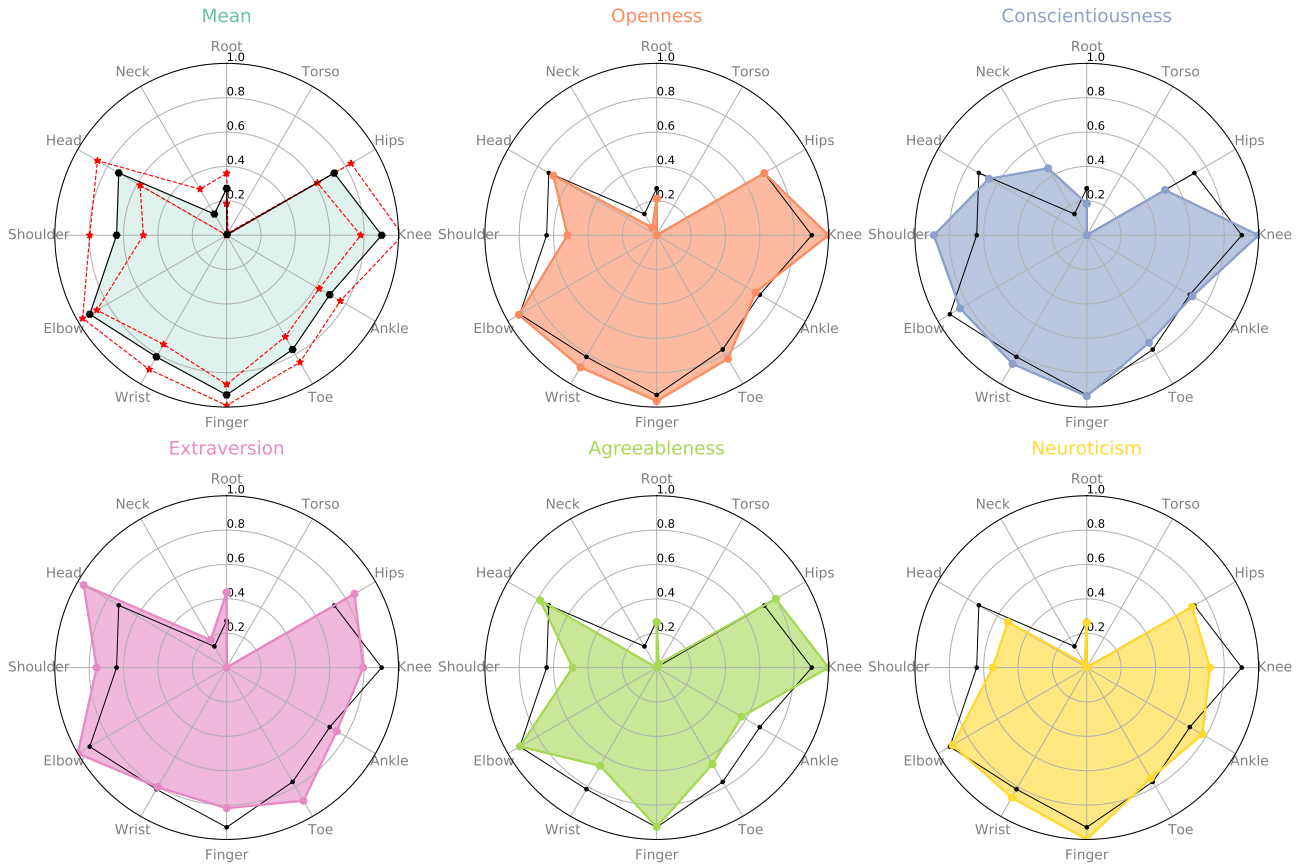
Figure 4: Relative importance of Joints of the five personality traits(Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism) using the Position Data. The black line indicates the mean importance of the corresponding joint marker. The red dotted line in the top left sub-figure indicates the standard deviation about the mean.

sion. Carlson et al. [17] found that, compared to Conscientiousness, the core body was more important in responsiveness to musical tempo in relation to Extraversion, which is partly supported by the slightly greater importance of the finger and wrist markers to Extraversion in our study, but partly contradicted by the importance of the shoulder marker in Extraversion. The difference between findings may relate to the use in the current study of position rather than velocity or acceleration data; that is, core body posture while moving to music may be more indicative of Conscientiousness than core body movement. EQ scores were more related to head, finger, hips and lower limb joints than SQ scores, which may be partly attributed to gender-typical movement patterns as females tend to score higher on the EQ than males [21, 35].

Several limitations of the current study should be noted. First, the majority of participants were from European or North American countries, and all eight music stimuli were of Western origin, limiting the degree to which results can be generalized cross-culturally. Secondly, There may exist potential bias due to gender imbalance. Future work could include separate analysis performed within gender categories. And lastly, participants' preferences for heard stimuli were not included in our model. This would be an important feature to focus on in future work, as preference

and enjoyment are highly relevant to personalized MIR.

Further extension of this work could help to make music recommendation systems more robust. Previous work has considered the relationship between personality and music preference [25, 36], while Greenberg et al. [26] explored the relationship between music preference and empathizing-systematizing theory, suggesting even that music may play a role in increasing empathy in people with empathy-related disorders, such as ASD. However, the relationship between embodiment, personality and musical experiences requires further exploration.

To conclude, this study represents an early step towards multimodal MIR. To make this approach applicable to personalized gesture-based retrieval systems, it can be extended to monocular video captured by accessible devices such as a mobile phone camera. This approach would be feasible due to recent progress in the area of 3D human pose estimation in predicting the body joint coordinates from a monocular video [37–39]. This would then allow future recommendation systems to take embodied processes into account, resulting in better and more responsive personalized experiences.

5. REFERENCES

[1] J. C. Bispham, "The human faculty for music: What's special about it?" Ph.D. dissertation, University of Cambridge, 2018.

[2] I. Cross, "The evolutionary nature of musical meaning," *Musicae scientiae*, vol. 13, no. 2_suppl, pp. 179–200, 2009.

[3] J. Richter and R. Ostovar, ""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*, vol. 10, p. 485, 2016.

[4] P. Toiviainen, G. Luck, and M. R. Thompson, "Embodied meter: hierarchical eigenmodes in music-induced movement," *Music Perception*, vol. 28, no. 1, pp. 59–70, 2010.

[5] R. T. Solberg and A. R. Jensenius, "Pleasurable and intersubjectively embodied experiences of electronic dance music," *Empirical Musicology Review*, vol. 11, no. 3-4, pp. 301–318, 2017.

[6] B. Burger and P. Toiviainen, "Embodiment in electronic dance music: Effects of musical content and structure on body movement," *Musicae Scientiae*, p. 1029864918792594, 2018.

[7] M. Lesaffre, L. D. Voogdt, M. Leman, B. D. Baets, H. D. Meyer, and J.-P. Martens, "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*, vol. 59, no. 5, pp. 695–707, 2008.

[8] B. Nettl, "An ethnomusicologist contemplates universals in musical sound and musical culture," *The origins of music*, vol. 3, no. 2, pp. 463–472, 2000.

[9] G. Luck, S. Saarikallio, B. Burger, M. R. Thompson, and P. Toiviainen, "Effects of the big five and musical genre on music-induced movement," *Journal of Research in Personality*, vol. 44, no. 6, pp. 714–720, 2010.

[10] G. Luck, S. Saarikallio, B. Burger, M. Thompson, and P. Toiviainen, "Emotion-driven encoding of music preference and personality in dance," *Musicae Scientiae*, vol. 18, no. 3, pp. 307–323, 2014.

[11] E. Carlson, P. Saari, B. Burger, and P. Toiviainen, "Dance to your own drum: Identification of musical genre and individual dancer from motion capture using machine learning," *Journal of New Music Research*, pp. 1–16, 2020.

[12] J. E. Cutting and L. T. Kozlowski, "Recognizing friends by their walk: Gait perception without familiarity cues," *Bulletin of the psychonomic society*, vol. 9, no. 5, pp. 353–356, 1977.

[13] M. Subotnick, "Interactive music playback system utilizing gestures," Nov. 1 2001, uS Patent App. 09/835,840.

[14] M. Gillies, "Understanding the role of interactive machine learning in movement interaction design," *ACM Transactions on Computer-Human Interaction (TOCHI)*, vol. 26, no. 1, pp. 1–34, 2019.

[15] L. Satchell, P. Morris, C. Mills, L. O'Reilly, P. Marshman, and L. Akehurst, "Evidence of big five and aggressive personalities in gait biomechanics," *Journal of nonverbal behavior*, vol. 41, no. 1, pp. 35–44, 2017.

[16] J. Michalak, N. F. Troje, J. Fischer, P. Vollmar, T. Heidenreich, and D. Schulte, "Embodiment of sadness and depression—gait patterns associated with dysphoric mood," *Psychosomatic medicine*, vol. 71, no. 5, pp. 580–587, 2009.

[17] E. Carlson, B. Burger, J. London, M. R. Thompson, and P. Toiviainen, "Conscientiousness and extraversion relate to responsiveness to tempo in dance," *Human movement science*, vol. 49, pp. 315–325, 2016.

[18] M. J. de Dreu, A. Van Der Wilk, E. Poppe, G. Kwakkel, and E. E. van Wegen, "Rehabilitation, exercise therapy and music in patients with parkinson's disease: a meta-analysis of the effects of music-based movement therapy on walking ability, balance and quality of life," *Parkinsonism & related disorders*, vol. 18, pp. S114–S119, 2012.

[19] A. Anzulewicz, K. Sobota, and J. T. Delafield-Butt, "Toward the autism motor signature: Gesture patterns during smart tablet gameplay identify children with autism," *Scientific reports*, vol. 6, no. 1, pp. 1–13, 2016.

[20] E. B. Torres, M. Brincker, R. W. Isenhower III, P. Yanovich, K. A. Stigler, J. I. Nurnberger Jr, D. N. Metaxas, and J. V. José, "Autism: the micro-movement perspective," *Frontiers in integrative neuroscience*, vol. 7, p. 32, 2013.

[21] S. Baron-Cohen and S. Wheelwright, "The empathy quotient: an investigation of adults with asperger syndrome or high functioning autism, and normal sex differences," *Journal of autism and developmental disorders*, vol. 34, no. 2, pp. 163–175, 2004.

[22] S. Baron-Cohen, J. Richler, D. Bisarya, N. Gurunathan, and S. Wheelwright, "The systemizing quotient: an investigation of adults with asperger syndrome or high–functioning autism, and normal sex differences," *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, vol. 358, no. 1430, pp. 361–374, 2003.

[23] J. M. S. Bamford and J. W. Davidson, "Trait empathy associated with agreeableness and rhythmic entrainment in a spontaneous movement to music task:

Preliminary exploratory investigations," *Musicae Scientiae*, vol. 23, no. 1, pp. 5–24, 2019.

[24] E. Carlson, B. Burger, and P. Toiviainen, "Dance like someone is watching: A social relations model study of music-induced movement," *Music & Science*, vol. 1, p. 2059204318807846, 2018.

[25] E. Carlson, P. Saari, B. Burger, and P. Toiviainen, "Personality and musical preference using social-tagging in excerpt-selection." *Psychomusicology: Music, Mind, and Brain*, vol. 27, no. 3, p. 203, 2017.

[26] D. M. Greenberg, S. Baron-Cohen, D. J. Stillwell, M. Kosinski, and P. J. Rentfrow, "Musical preferences are linked to cognitive styles," *PloS one*, vol. 10, no. 7, 2015.

[27] E. Carlson, B. Burger, and P. Toiviainen, "Empathy, entrainment, and perceived interaction in complex dyadic dance movement," *Music Perception: An Interdisciplinary Journal*, vol. 36, no. 4, pp. 390–405, 2019.

[28] O. P. John, S. Srivastava *et al.*, "The big five trait taxonomy: History, measurement, and theoretical perspectives," *Handbook of personality: Theory and research*, vol. 2, no. 1999, pp. 102–138, 1999.

[29] A. Wakabayashi, S. Baron-Cohen, S. Wheelwright, N. Goldenfeld, J. Delaney, D. Fine, R. Smith, and L. Weil, "Development of short forms of the empathy quotient (eq-short) and the systemizing quotient (sq-short)," *Personality and individual differences*, vol. 41, no. 5, pp. 929–940, 2006.

[30] B. Burger and P. Toiviainen, "Mocap toolbox-a matlab toolbox for computational analysis of movement data," in *10th Sound and Music Computing Conference, SMC 2013, Stockholm, Sweden*. Logos Verlag Berlin, 2013.

[31] W. Liu, P. P. Pokharel, and J. C. Príncipe, "Correntropy: Properties and applications in non-gaussian signal processing," *IEEE Transactions on Signal Processing*, vol. 55, no. 11, pp. 5286–5298, 2007.

[32] S. Goldin-Meadow, "Talking and thinking with our hands," *Current directions in psychological science*, vol. 15, no. 1, pp. 34–39, 2006.

[33] A. F. d. C. Hamilton, R. M. Brindley, and U. Frith, "Imitation and action understanding in autistic spectrum disorders: how valid is the hypothesis of a deficit in the mirror neuron system?" *Neuropsychologia*, vol. 45, no. 8, pp. 1859–1868, 2007.

[34] D. Wu, J. V. José, J. I. Nurnberger, and E. B. Torres, "A biomarker characterizing neurodevelopment with applications in autism," *Scientific reports*, vol. 8, no. 1, pp. 1–14, 2018.

[35] N. F. Troje, "Decomposing biological motion: A framework for analysis and synthesis of human gait patterns," *Journal of vision*, vol. 2, no. 5, pp. 2–2, 2002.

[36] P. J. Rentfrow and S. D. Gosling, "The do re mi's of everyday life: the structure and personality correlates of music preferences." *Journal of personality and social psychology*, vol. 84, no. 6, p. 1236, 2003.

[37] D. Pavllo, C. Feichtenhofer, D. Grangier, and M. Auli, "3d human pose estimation in video with temporal convolutions and semi-supervised training," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 7753–7762.

[38] A. Venkat, C. Patel, Y. Agrawal, and A. Sharma, "Humanmeshnet: Polygonal mesh recovery of humans," in *Proceedings of the IEEE International Conference on Computer Vision Workshops*, 2019, pp. 0–0.

[39] Y. Cheng, B. Yang, B. Wang, and R. T. Tan, "3d human pose estimation using spatio-temporal networks with explicit occlusion training," *arXiv preprint arXiv:2004.11822*, 2020.