Proceedings of the 10th International Conference of Students of Systematic Musicology (SysMus17), London, UK, September 13-15, 2017.
Peter M. C. Harrison (Ed.).

Kinematics Feature Selection of Expressive Intentions in Dyadic Violin Performance

Georgios Diapoulis[1], Marc Thompson[2]

University of Jyväskylä Dept., Music, Arts, and Culture Studies, Finland
[1]gediapou@student.jyu.fi, [2]marc.thompson@jyu.fi

ABSTRACT

There is evidence that bodily movement plays a crucial role in regulating expressivity in music performance. Advances in technologies related to human movement research (e.g. motion capture using infrared cameras) give us the opportunity to study bodily motion with millimeter precision. Consequently, we can extract fine-grained kinematic characteristics and perform statistical learning techniques in order to identify similarities and differences in spatial accuracy of intended expressive movements. In this study, we applied feature extraction and feature generation algorithms to identify the kinematic characteristics that better predict expressive intentions. The results suggest that musical expressivity is not physically rendered in similar movement patterns during perception and during production of dyadic musical performance. We propose that future studies should focus on the interaction between motor experience and visual perception of expressivity.

I. INTRODUCTION

Expressive bodily motion is a fundamental property of music performance, and highly important aspect to deepen our understanding of human interaction (Palmer, 1997). Advances in motion capture technologies make possible to record bodily motion with high spatial and temporal accuracy. Thus, there is growing interest to explain behavioural and affective phenomena based on objective measures of motion capture data. The research program of embodied music cognition suggest that bodily motion has major importance in musical activities (Leman, 2008). Investigating motion is a continuing concern within natural sciences, and the study of physical motion in the 17th century established what is known today as Newtonian physics or classical mechanics. As an analogy, the study of human motion might be the appropriate focal point for sound and reproducible studies within behavioural and cognitive sciences.

Full-body human movement recordings may produce high-dimensional spaces which make the analysis of the data a challenging endeavour. Dimension reduction techniques are based on feature extraction and feature generation. Feature selection is a feature extraction technique which selects the appropriate subset of features that optimize the learning performance. Feature selection is a family of different techniques that may vary from a simplistic exhaustive search of all possible combinations of a set of features, to highly sophisticated techniques. On the other hand, principal component analysis (PCA) is a technique that is used to perform feature generation. PCA generates a new synthetic data set which produces a new transformed coordinate system based on the percent of explained variance. The fundamental difference between PCA and feature selection is that PCA is an unsupervised machine learning technique whereas feature selection is supervised technique.

Computational approaches such as feature selection and feature transformation (e.g. PCA) can provide us with useful insights about the embodiment of expressive performance. More specifically, PCA has been used to detect the dominant modes in movement data (Daffertshofer, Lamoth, Meijer & Beek, 2004, Toiviainen et al., 2010). For example, in a previous study using the same data set (Diapoulis, 2016) we applied joint-PCA on the violin dyads; the first principal component (PC) consisted of movement on the mediolateral axis, the second PC consisted of movement on the anteroposterior axis, and the third PC consisted of movement on the vertical axis. On the other hand, feature selection does not transform the original data, instead it is the process of selecting the appropriate subset of features. Whereas PCA transforms the original dimensions of the movement data and generates a new synthetic data set, feature selection algorithms are used to identify which feature subset can better perform predictions.

Broughton & Davidson (2016) described the expressive moments in marimba performance using Laban movement analysis, and they reported that head nod, head shake, upper body wiggle, and anteroposterior surge, along with a regular sway (anteroposterior movement) are all factors of expressive performance. Bodily sway has been shown in many studies to be a significant factor of communication and interpersonal coordination of leader-follower dynamics (Chang, Livingstone, Bosnyak & Trainor, 2017; Keller & Appel, 2010).

The perception of expressive performance is associated with a wide variety of movement patterns, but there is consensus in literature that bodily sway is a dominant component of expressive gestures (Broughton & Davidson, 2016; Dahl & Friberg, 2007; Diapoulis, 2016). On the other hand, there are no studies that attempt to identify which bodily parts can better discriminate different expressive manners. The present study fills a gap in the literature by shifting the focus on the kinematic features that discriminate the different expressive manners. Thus, we make use of third-person objective movement measures to classify intersubjective experience of expressive intentions. An important point that we have to clarify is that the focus is *on intended* and *not on perceived* expressivity in music performance. That is, our aim is to identify which kinematic features account for the embodiment of intended expressivity in dyadic music performance.

II. METHODS

A. Participants and Procedure

Three violin dyads participated in this study (6 musicians total; 4 females; age: M = 24.1, SD = 1.7). The violinists were recruited from student populations at the University of Jyväskylä and the Jyväskylä University of Applied Science. Musicians had received on average 15.8 (SD = 2.3) years of instrumental training on the violin.

The violin dyads performed while standing and looking at each other as shown in Figure 1. The dyads performed a short piece arranged for two violins: "De Kleinste", composed by J. Beltjens (16 bars, 6/8 time signature), and the score is available in Diapoulis (2016). After a short rehearsal period, each dyad performed the piece nine times in a 3 × 3 task design: three expressive intentions (deadpan, normal, exaggerated) performed using three timing conditions (60-BPM, 90-BPM, free tempo). In the current study, we ignored the effect of tempo, as a factor that might have an effect on the classification of different expressive conditions.

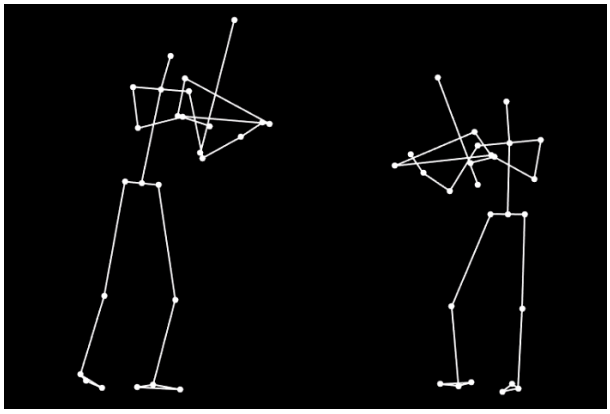

Figure 1. Snapshot from a dyadic performance.

B. Apparatus

Optical motion capture data was produced using 8 Qualisys Oqus infrared cameras at 120 Hz sampling rate. Twenty-six markers were placed on the joints of each musician, and five markers were placed on the violin (2 on the bow, and 3 on the violin itself). The data was labeled within Qualisys' Track Manager software and analyzed in MATLAB using functions within the MoCap Toolbox (Toiviainen & Burger, 2010), and the MATLAB statistics and machine learning toolbox.

C. Experimental Design

This study is based on the experimental design that was reported in Diapoulis (2016). The aforementioned study had two experiments; a motion capture experiment of dyadic violin performance, and a perceptual experiment of evaluating expressivity in performance. Figure 1 shows a screenshot from a perceptual stimulus. In the current study, we have used the motion capture segments that we used as stimuli for perceptual evaluation of expressivity. No perceptual data are used in the current study. As noted in Diapoulis (2016) the total number of perceptual stimuli was 72 segments (3×2×3×4); three dyads, two expressive conditions (deadpan and exaggerated), three modalities (audiovisual, audio-only and visual-only), and four melodic segments. The decision to eliminate the normal expressive manner, was done based on Thompson and Luck (2012). In this study, the authors reported that there is no consistency in the embodiment of normal and exaggerated piano performance. The decision to take the two extreme expressive conditions (deadpan and exaggerated) was done with a view to reduce the average duration of the perceptual experiment, due to the fact that the perceptual experiment was web-based (online) and we didn't provide any incentives to the participants (for details see Diapoulis, 2016).

D. Movement Analysis

All the analysis is based only on motion captured data. In the pre-processing stage of the movement analysis we reduce the 26 markers to 20 joint markers for each violinist, ignoring the markers on the violin. Then we connected the 20 joint markers in order to create stick figures as shown in Figure 2. This was done to facilitate presentation view and had no effect on the movement analysis. Preliminary movement analysis showed that the musicians embodied the different levels of expressivity by moving with more kinetic energy in the more exaggerated expressive conditions. This preliminary result was interpreted as evidence that the assigned linguistic descriptions (i.e. normal, deadpan, exaggerated) had causal effect on the embodiment of the musicians' expressive intentions.

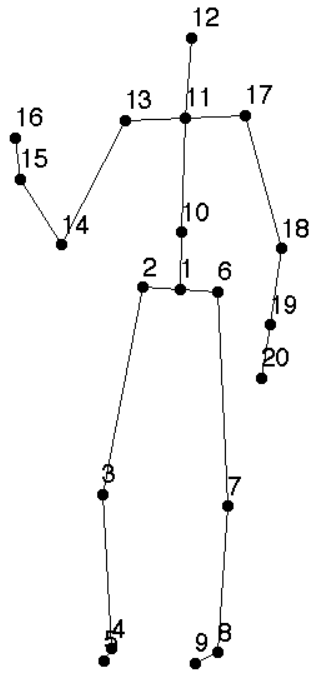

Figure 2. Stick figure of violinist, this is the back view of the performer.

For the movement analysis, we assigned to the stick figures frontal view in respect to the two markers on the hips. This step was done in order to standardize the motion capture data. Then we segmented each motion capture recording in four parts based on the score, and we computed the velocities for every marker (120 timeseries = 20 markers x 3 dimensions x 2 performers). The next step was to concatenate the two timeseries for each segment, in order to treat the dyad as a whole. First, we computed the velocities and then we applied concatenation of co-performers in order to eliminate the

possibility of applying derivation on non-continuous timeseries, which implements noise in variance. This error was done by the first author in Diapoulis (2016), and the result was that the joint-PCA produced five dimensions for explained variance of 95%.

E. Kinematic Features and Statistical Learning

The statistical analysis was based on the global descriptor of standard deviation for each segment. We applied forward sequential feature selection (FSFS) using cross-validation, in order to identify which markers can better predict the different expressive intentions. For that purpose, we evaluate the performance of both linear and quadratic classifiers of discriminant analysis.

Moreover, we also applied FSFS and backward sequential feature selection (BSFS) on transformed kinematics that we generated by applying joint-PCA on a small subset of ancillary markers (head, root, left and right shoulder). For this purpose, we followed the feature extraction process that we already described, but we focused on the subset of ancillary markers and we applied joint-PCA in advance of calculating the statistical moment of standard deviation. The decision to focus on the subset of ancillary markers, was done due to the fact that the first three principal components generated new synthetic dimensions that describe movement on the mediolateral, anteroposterior and vertical axis (see Introduction). The computational procedure is shown in Figure 3.

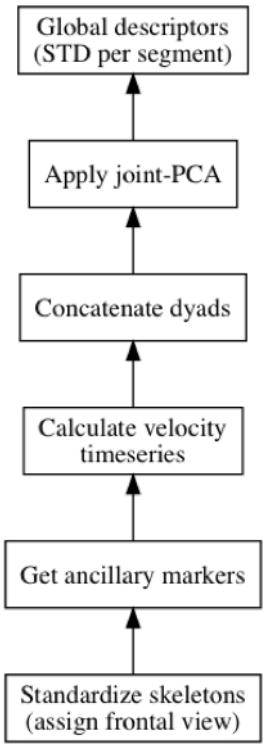

Figure 3. Computational process for generating features based on joint principal component analysis.

III. RESULTS

We remind the reader that the motion capture segments were identical to the perceptual stimuli that we used in Diapoulis (2016). The mean duration of each perceptual stimulus was 8.89 seconds, whereas the mean duration for the exaggerated condition was 9.20 seconds and the mean duration for the deadpan condition was 8.58 seconds.

A. Kinetic Energy

For each dyadic performance, we extracted the instantaneous kinetic energy for each performer using the method used by Toiviainen, Luck & Thompson (2010). The total kinetic energy was estimated as the sum of both performers' translational and rotational energy of each marker. We trimmed each performance from five to twenty-five seconds, and we estimated the kinetic energy within this time span. The total mean instantaneous kinetic energy across all segments for all dyads per expressive condition was .31, .80, 1.20 Joules for deadpan, normal and exaggerated expressive intentions respectively. This measure provides an estimation of the overall physical activity, and provide us the initial evidence to continue to further analysis.

B. Principal component analysis

We applied joint-PCA on the ancillary markers, of head, root, left and right shoulder of the timeseries data. We selected this small subset of ancillary markers, because ancillary gestures have been proposed that play a crucial role in the perception of expressivity (Thompson and Luck, 2012; Wanderley 2002). Furthermore, joint-PCA produced four synthetic dimensions that explained more than 95% of variance (see Table 1), and the first three principal component consist of movements on different axes, which makes the interpretation of the components trivial (see Introduction). Figure 4 shows the principal component loadings matrix based on varimax rotation. The latter is a linear transformation which rotates the coordinate system in order to maximize the explained variance.

Principal Components	PC1	PC2	PC3	PC4
Percent of explained variance	72.0	12.7	6.3	4.5

Table 1. Explained variance of the first four principal components.

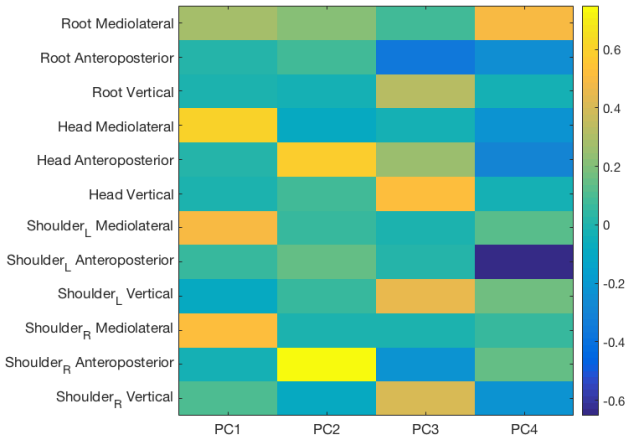

Figure 4. Principal component loadings matrix based on varimax rotation.

C. Feature Selection

We applied feature selection on two sets of kinematic features; the extracted and the generated kinematics. The extracted kinematics described a 60-dimensional space. The global descriptor of standard deviation was extracted for every marker for each segment. The set of the generated features, described a four-dimensional space, based on the global descriptors of standard deviation that was calculated from the joint-PCA for each segment.

1) FSFS on the extracted kinematics. We applied forward sequential feature selection on the global descriptors of all the markers. Our analysis, showed that FSFS using 6-fold cross-validation on a quadratic discriminant classifier predict the expressive intentions of deadpan and exaggerated with 100.0% accuracy based on the confusion matrix. This prediction performed using the kinematic features of standard deviation of the *left knee* and the *head* on the *vertical axis*. We also performed FSFS based on linear discriminant classifier. This approach predicted the expressive intentions with 98.6% accuracy, based on the kinematics of the *right ankle* and *head* on the *vertical axis*.

2) FSFS and BSFS on the generated kinematics. We applied FSFS and BSFS based on the kinematics that were generated from joint-PCA. This analysis showed that the third principal component (PC3) was the best predictor of expressive intentions for both FSFS and BSFS. Using quadratic discriminant analysis (QDA) the accuracy was 97.2% and using linear discriminant analysis (LDA) the accuracy was 93.0%.

IV. DISCUSSION

The aim of the study was to identify which kinematics features can better discriminate performances of deadpan and exaggerated expressive intentions. Three violin dyads participated and they performed a short composition. The instruction given to the violists was to perform the piece under three expressive manners (for details see in Methods, subsection Experimental Design). We segmented the song in four melodic segments based on the score (for detailed information see Diapoulis, 2016), and for each segment we both extracted and generated global descriptors based on velocity timeseries data. The statistical moment of standard deviation was the most appropriate descriptor of expressivity.

Our goal was to identify which kinematic features can better predict intended expressivity in musical dyads. Thus, our focus was to use a variety of machine learning techniques in order to predict the qualities of deadpan and exaggerated expressive intentions. For that purpose, we used both supervised and unsupervised algorithms. Forward sequential feature selection using QDA showed that the velocities of the left knee and the head across the vertical axis are the most important kinematic features. Using LDA the kinematic feature of the left knee was replaced by vertical motion of the right ankle. Furthermore, we applied both FSFS and BSFS on the transformed kinematics (i.e. PCA). Once again, movement on the vertical axis showed to be the most important predictor of expressive intention.

The aforementioned evidence raises questions whether or not the intended expressivity shares the same movement patterns as perceived expressivity. The perception of expressive bodily motion seems to had major influences from body sway. Our analysis shows that the production of deadpan and exaggerated expressive performance can better discriminated based on movement on the vertical axis. Thus, the results suggest that bodily movement based on motoric experience might not align with visual perception of expressive music performance.

V. CONCLUSION

We presented evidence that intended expressive performance might not share the same movement patterns with visual perception of expressivity. Future studies should focus on the comparison of expert musicians and non-musicians populations in order to study the interaction between motor experience and visual perception of expressive music performance. Ultimately, the focus should be placed on kinematic correlates of intended and perceived expressivity in music performance. Data collection is ongoing and future reports will include more violin dyads.

REFERENCES

Broughton, M. C., & Davidson, J. W. (2016). An Expressive Bodily Movement Repertoire for Marimba Performance, Revealed through Observers' Laban Effort-Shape Analyses, and Allied Musical Features: Two Case Studies. *Frontiers in Psychology,* 7.

Daffertshofer, A., Lamoth, C. J., Meijer, O. G., & Beek, P. J. (2004). PCA in studying coordination and variability: a tutorial. *Clinical biomechanics*, 19(4), 415-428.

Dahl, S., & Friberg, A. (2007). Visual Perception of Expressiveness in Musicians9 Body Movements. *Music Perception: An Interdisciplinary Journal*, 24(5), 433-454.

Diapoulis, G. (2016). *Exploring the perception of expressivity and interaction within musical dyads.* (Master's thesis, University of Jyväskylä, Jyväskylä, Finland). Retrieved from http://r.jyu.fi/jXo

Keller, P. E., & Appel, M. (2010). Individual differences, auditory imagery, and the coordination of body movements and sounds in musical ensembles. *Music Perception: An Interdisciplinary Journal*, 28(1), 27-46.

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

Palmer, C. (1997). Music performance. *Annual review of psychology*, 48(1), 115-138.

Thompson, M. R., & Luck, G. (2012). Exploring relationships between pianists' body movements, their expressive intentions, and structural elements of the music. *Musicae Scientiae*, 16(1), 19-40.

Toiviainen, P., & Burger, B. (2010). Mocap toolbox manual. Online at:http://www.jyu.fi/music/coe/materials/mocaptoolbox/MCTmanual.

Toiviainen, P., Luck, G., & Thompson, M. R. (2010). Embodied meter: hierarchical eigenmodes in music-induced movement. *Music Perception: An Interdisciplinary Journal*, 28(1), 59-70.

Wanderley, M. M. (2002). Quantitative analysis of non-obvious performer gestures. *Lecture notes in computer science*, 241-253.