Validating kinematic displays for the perception of musical performance

Manfred Nusseck[*] Marcelo M. Wanderley[‡] Erwin Schoonderwaldt[‡]
(*) Max Planck Institute for biological cybernetics, Tübingen, Germany
(‡) IDMIL, McGill University, Montreal, Canada
E-mail: manfred.nusseck@tuebingen.mpg.de, marcelo.wanderley@mcgill.ca, schoonderw@music.mcgill.ca.

Abstract

Human gestures contain certain characteristics and meanings in communication and represent a link between intention and body. This paper describes a pilot study investigating the role of ancillary musical gestures in understanding musical meaning from the listener's standpoint. We conducted a perceptual experiment using motion-capture recordings of musicians. Participants were presented video recordings and reconstructed point-light displays of music performances. By asking them to rate certain music-related parameters we found that abstract motions of the point-light displays yielded similar ratings to those of the real recordings. This suggests that pure body motion seems to be sufficient to communicate certain musical impressions.

1. Introduction

Playing musical instruments is an interaction of both sound and body movements. These movements are tightly linked to each other so that ancillary performance gestures seem to have an intrinsic relationship to the musical performance. Why are these gestures performed? What do they communicate to the audience? How important are body motions for the music experience of the audience? In this study, we examined to which extent visual aspects contribute to the communication process between performer and listener and how they influence perceptual and aesthetic judgments of music. This paper describes a pilot-study investigating whether pure body motion supports similar aspects of musical performance as a real video. It validates whether or not kinematic displays of musicians can be sufficient representatives in judging certain musical parameters. The study is part of a larger project addressing the general role of body motions in perceiving musical performances.

Gestures in music are the topic of a large field of research. Previous studies proposed a general categorization into (a) gestures that are involved in playing the instrument and are related directly to the production of the sound, and (b) gestures that are part of the performance but not produced for the purpose of sound generation [1] [11]. The latter are so-called "ancillary gestures" and are considered as being connected to the expressiveness of the performance [5]. These gestures contain body sways and subtle facial motions, such as eyebrow movements or facial expressions. These visual cues have been found to influence musical intelligibility [9].

Investigating performances of clarinet players, Wanderley et al. [11] showed that ancillary gestures occur frequently in musical performances, even though they are not essential for it. Furthermore, they varied considerably across performers, but are more consistent within performers. The findings showed that these gestures are not randomly performed, but more an integral part of the performance process.

During learning of a new musical piece, Chaffin and Logan [2] showed that musicians' performance cues move through stages from basic and effective to expressive, in a more holistic sense. Body movements, however, are also related to performance interpretation. Dahl et al. [4] showed that video sequences of marimba players, where only the upper body was presented, clearly support emotional intentions. With a multi-modal design in which participants either saw, heard, or both saw and heard musicians performing, Vines et al. [10] found that visual perception of musical performance interacts with auditory perception. They also showed that ancillary gestures are linked to structural features of the music, i.e. phrase boundaries or extended notes.

Given that live performances or even video recordings of musicians contain a large amount of contextual and situational information, such as the players clothes, hair style, body shape, or the environmental background (high-level cues), it is difficult to know what influences a perceiver's judgement. To eliminate these high-level information sources, research started to use so-called "point-light displays", adapted by Johansson [6]. With motion-capture systems, the positions of markers, attached to major body joints on the player, are tracked during the performance. If displaying these positions as dots on a screen, it creates a stimulus based solely on the spatio-temporal movements of

Proceedings of ENACTIVE/07
4th International Conference on Enactive Interfaces
Grenoble, France, November 19th-22nd, 2007

body and limbs. Therefore, point-light displays separate performance gestures from all high-level cues so that only the body motion is still present. Previous studies have shown that for human motion recognition point-light display raised similar results as natural videos [8].

Using this technique, Davidson [5] recorded violin players performing three different levels of expression: deadpan, standard, and exaggerated. In a perceptual experiment to recognize expressive intentions, she found that the different expressions were clearly perceivable from point-light displays alone.

Additionally, Davidson [5] found that by presenting the stimuli either visual only, sound only, or multimodal, the gesture of the musicians gave a better indication of expressive intent than the sound. This showed that visual cues seem to have a strong influence on music perception.

In this study we conducted a perceptual experiment investigating if pure body motions can be sufficient to allow rating of music-related parameters. Since the visual aspects seem to be an important source of information, we compared ratings of point-light displays with real videos.

2. Experiment

2.1. Method

2.1.1. Stimuli

To create the set of stimuli we used motion-capture data of four different clarinet players (3 male, 1 female) performing Brahms Sonata 1, Op 120. The recordings were done at the IDMIL Lab of McGill University. In addition to the motion-capture, the players were recorded using a camcorder. All players stood in front of a music stand and were asked to play naturally (i.e., as during a concert performance). We used clarinet players, as their visible body movements are not directly related to sound production. The study therefore directly investigates music-related ancillary gestures.

For the experiment, we used only the first phrase of the piece (12-15 seconds), since this was the most consistent part across players and was performed without in-between breathing to eliminate this as a possible cue.

For the first condition we showed digital video recordings of the players. For the other conditions, the motion-capture data was taken to create "stick figures" of the musicians. A selection of certain data points from the players was taken to connect them with a white line building a sparse shape of the person.

The second condition presented the stick figures of the players from the same viewpoint as shown in the video recordings (90 degree from the left). In the third condition, we showed the stick figures from the front (see Figure 1). While possible side-to-side motions of the player were difficult to detect in the side view, due to the abstract and transparent presentation, in the frontal view possible front-back motions were less recognizable.

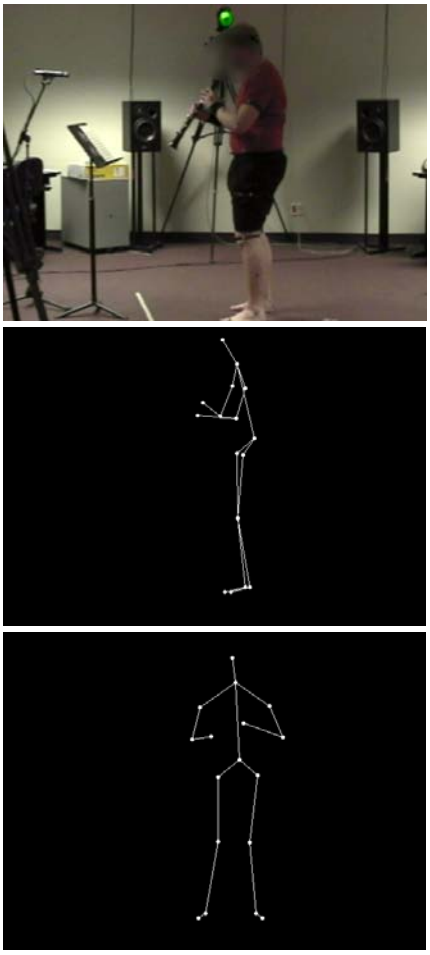

Figure 1. Screenshots of one player: (A) video recording, (B) stick figure side view, and (C) stick figure frontal view[1]

2.1.2. Design

The experiment was designed and conducted with the Psychophysics Toolbox (PTB-3)[2] in Matlab on an Apple G4 notebook. The videos were presented on the screen with 560x420 pixel size and 25 fps. Sound was provided through headphones.

Participants had to press the space bar to start the video sequence. The movies were shown in randomized order and could only be seen once. After the sequence ended, the participants were asked to rate the *tension*, the *intensity*, the *fluency*, and the *musician's professionalism* in a 7-point Likert scale. After the four ratings were made, the participants were able to start the next trial with the space bar. The movies were shown without repetition.

[1] The face in Figure A is blurred to hide performers' identity, though there was no blurring during stimulus presentation in the experiment.
[2] http://www.psychtoolbox.org

The different scales of rating were introduced to the participants as follows:

- *Tension*: With this scale, a more general musical impression of the player has to be judged. The concept of musical tension is a complex phenomenon and hard to describe formally. The participants were told to rate in this scale the relaxation, stress, and immersion of the performance and the performer. A high tension refers to a feeling of excitement, whereas a low tension refers to uncertainty and relaxation. Vines et al. [10] found that different manners of movements elicit different tension ratings. Here, we wanted to find out if tension is contained in the body motion or needs other visual cues.
- *Intensity*: This scale relates to the expressiveness and the emotion. High intensity meant that the player exaggerated the performance in both fields in relation to the piece. Unexpressive and emotionless performances should be rated with low intensity. This scale relates to Davidson [5].
- *Fluency*: In this scale, participants were asked to rate the smoothness of the performance. If the player drew a clear bow over the whole phrase it should be rated with high fluency. Otherwise, a low fluency was related to a more jerky performance. Here, we wanted to see if the body motion supports the impression of fluent performances.
- *Professionalism*: Finally, the participants were asked to rate the musician's ability to play the instrument. Here, the extremes referred either to a player who seemed to be a beginner on the instrument or to a professional player. This scale measures whether high-level visual cues influence the judgment of the musicians' expertise and ability.

Additionally, the participants were told to answer intuitively. Furthermore, they were asked to use the "4" if they think the performance was played in a standard and neutral way related to the piece. The extremes of each scale were judged relative to a neutral performance by the player.

2.1.3. Participants

Ten graduate students from the McGill Music Lab attended this pilot (age 25-35). Most of them were music educated and none of them was a clarinet player. They were not paid for their participation.

2.2. Results and discussions

The mean ratings split by the different presentation styles and scales of rating are shown in Figure 2. Overall, the ratings seem to be rather similar and are distributed in a close range around the 4 as the middle answer (3.7 to 4.7). A two-way analysis of variance (ANOVA) was run for each scale for the factors presentation and player.

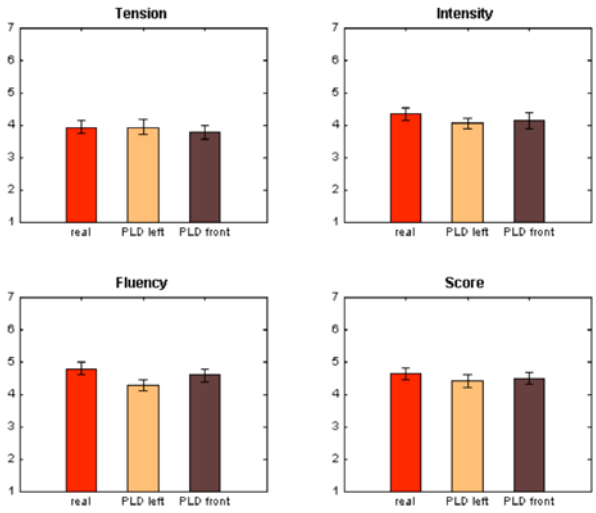

Figure 2. Mean ratings of the participants for each scale and presentation style (PLD: point-light display). Error bars depict standard error of the mean.

For all scales we found no significant effect for the presentation ($F(2,18)<3.0$, $p>0.1$), indicating that the individual ratings were not influenced by the different display styles. This suggests that the abstract display of pure body motions provides similar impressions for judging these scales to those of the real videos. Furthermore, different viewing angles did not seem to influence these experiences.

A significant effect of player was found only for tension ($F(3,27)=3.6$, $p<0.02$) and for intensity ($F(3,27)=4.6$, $p<0.01$), but not for fluency and professionalism ($F(3,27)<1.0$, $p>0.4$). This effect was mostly due to one player, who was rated lower than the others. Interestingly, for all scales there was no significant effect for the interaction of player and presentation style ($F(6,54)<0.8$, $p>0.6$), indicating that the individual player was rated in a similar way for the different presentation styles.

Although all players performed naturally, they played the piece in a very individual way. For instance, the average tempo of the performances varied considerably (Player 1: 97 bpm, Player 2: 113 bpm, Player 3: 118 bpm , Player 4: 91 bpm). Even then, the findings show that the listeners got rather neutral impressions for each player and rated them in a similar way across the scales.

To test if the different scales were judged independently, we also analyzed correlations between the scales (see Table 1). Overall, the correlation coefficient (R^2) is rather low, indicating that the scales were treated separately and did not interact with each other. For fluency and professionalism the correlation is about 0.5, suggesting that fluent and smooth performances were sometimes dependent on a higher level of instrumental expertise.

Table 1. Correlations between different scales

R^2	Intensity	Fluency	Profession
Tension	0.24	0.01	0.06
Intensity	XX	0.14	0.34
Fluency	0.14	XX	0.49

3. Conclusions

In this paper we investigated the influence of presentation style on the perception of musical performances. Participants were shown either the real video or point light displays of clarinet players. They were then asked to rate the tension, the intensity, and the fluency of the performance, along with the player's professionalism.

Overall, the results show that the different presentation styles were rated similarly, suggesting that the presentation of pure kinematic body motions of a musician seem to generate the same impressions as a real video, particularly for the scales used. Furthermore, presenting a frontal or a side view of the musician did not change the aesthetic and musical experience of the performance. This suggests that both frontal and side motions contain indicators related to similar musical impressions.

In summary, we have shown that music-related judgments could be sufficiently communicated through pure body motions of the musician. This experiment, however, still leaves open questions. Where exactly is the information located? Which motion raises what kind of experiences? How far does pure body motion effect the communication of complex aspects of musical performances? Further experiments in this project currently address these questions

4. Future work

The major goal of the project is to investigate the influence of body motions on music perception. Since this experiment has shown that the point-light displays carry sufficient information to give certain musical impressions, current studies are using these point-light displays to examine the role of different body parts and the relationship between sound and appropriate motions. Physical analyses of the body motions are also being considered in order to correlate them with perceptual ratings.

Furthermore, it is planned to use this method of musical ratings to validate the believability of virtual musicians. Mazzarino et al. [7] used quantitative gesture analysis to investigate similarities between animated characters and real recorded motions. They developed an inverse kinematics model using prioritized constraints to animate and reconstruct virtual musicians. Using this method of measuring musical impressions we aim to investigate perceptual aspects of the animation model as it is presented to listeners. Findings from this can help in designing and improving believable human computer interfaces.

Finally, this project will contribute general insights into how we perceive musical performance and whether ancillary gestures provide additional information to the listener.

5. Acknowledgments

We would like to thank Barbara Mazzarino and Ronan Boulic for helpful input and discussions. This research has been supported by the E.U. Network of Excellence on ENACTIVE Interfaces[3].

References

[1] Cadoz, C. and Wanderley, M.M. "Gesture: music", In: M.M. Wanderley and M. Battier (Eds.), Trends in gestural control of music, Paris, IRCAM/Centre Pompidou, 2000

[2] Chaffin, R. and Logan, T. "Practicing perfection: How concert soloists prepare for performance", Advances in Cognitive Psychology, 2 (2-3), 113-130, 2003

[3] Clarke, E.F. and Davidson, J.W. "The body in music as mediator between knowledge and action", In: W. Thomas (Ed.) "Composition, Performance, Reception: Studies in the Creative Process in Music", Oxford University Press, 74-92, 1998

[4] Dahl, S. and Friberg, A. "Expressiveness of musician's body movements in performances on marimba", Gesture Workshop at the 5th International Workshop on Gesture and Sign Language Based Human-Computer Interaction, Genoa, 2004

[5] Davidson, J. "Visual perception of performance manner in the movements of solo musicians", Psychology of Music, 21, 103-113, 1993

[6] Johansson, G. "Visual perception of biological motion and a model for its analysis", Perception & Psychophysics, 14, 201-211, 1973

[7] Mazzarino, B., Peinado, M., Boulic, R., Volpe, G., and Wanderley, M.M., "Improving the believability of virtual characters using quantitative gesture analysis", International Gesture Workshop, Lisbon, Portugal, 2007

[8] Scully, D.M. "Visual perception of technical execution and aesthetic quality in biological motion", Human Movement Science, 5, 185-206, 1986

[9] Thompson, W.F., Graham, P., and Russo, F.A., "Seeing music performance: visual influences on perception and experiences", Semiotica, 156 (1/4), 177-201, 2005

[10] Vines, B.W., Krumhansl, C.L., Wanderley, M.M., and Levitin, D.J. "Cross-modal interactions in the perception of musical performance", Cognition,, 101, 80-113, 2006

[11] Wanderley, M.M., Vines, B.W., Middleton, N., McKay, C., and Hatch, W. "The Musical Significance of Clarinetists' Ancillary Gestures: An Exploration of the Field", Journal of New Music Research, 34:1, 97 – 113, 2005

[3] http://www.enactivnetwork.org