EAI Endorsed Transactions
on Creative Technologies
Research Article ICST.ORG

Evaluating music performance and context-sensitivity with Immersive Virtual Environments

Donald Glowinski[1,*], Naëm Baron[1], Kanika Shirole[2], Sélim Yahia Coll[2], Lina Chaabi[2], Tamara Ott[2]
Marc-André Rappaz[3] and Didier Grandjean[1,2]

[1] Swiss Center for Affective Sciences (NCCR), University of Geneva, Campus Biotech, CH-1211 Geneva 20, Switzerland
[2] Neuroscience of Emotion and Affective Dynamics (NEAD) - University of Geneva , 40 bd Pont d'Arves, CH-1205, Geneva, Switzerland
[3]Geneva University of Music (HEM), 12 Rue de l'Arquebuse, CH-1204 Geneva, Switzerland

Abstract

This study explores a unique experimental protocol that evaluates how a musician's sensitivity to social context during performance can be analysed through a combination of behavioral analysis, self-report and Immersive Virtual Environment (IVE). An original application has been developed to create audience of avatars that display different motivational states that are known to affect musician's performance. The musicians' body expressions have then been recorded through a motion capture system and analysed as they relate to audience motivational state. The musician subjective experience has been captured after each performance through semi-structured interviews. Preliminary results depict the strategies implicitly employed by four expert violinists during their performances under the various contexts (empty room and engaged and disengaged audience of avatars). Finally, this study discusses the way to improve methodology, analyses and real-world responses to musician's needs.

Keywords: Virtual Immersive environment, body expressivity, Music Performance.
Received on 14 May 2014, accepted on 29 January 2015, published on 27 February 2015

doi: 10.4108/ct.2.2e3

1. Introduction

In capturing a musician's expressive behavior we develop ways to understand his or her capacity to adapt to social as well as musical contexts. Because music performance often occurs in a public space, audience reactions influence the resulting performance. Musicians themselves are very sensitive to reactions of the audience and have developed specific motor and social capacities that help them to react appropriately in every given context. With this expertise, they develop strategies to control stress and performance anxiety when facing an audience.

A musician may have excellent coping mechanisms and turn his nervous energy into a positive kinetic energy during performance or, alternatively, may become frozen by stage fright. We are interested in the manifestation of this body kinetic energy and how it may help in regulating the interaction between the musician and audience in a performance scenario.

In order to control environmental stimuli such as the audience, light, and space, we used an Immersive Virtual Environments (IVEs) in our experimental design. IVEs allow experimenters to create virtual environments with unlimited configurations that can adapt in real time to user's behavior. Using this system, our study focuses on the behavioral component of body expressivity. Kinetic energy is particularly crucial in understanding broad, unrefined body reactions and helps in deriving a first understanding of the behavioral underpinnings of expressive gesture [1].

In order to take advantage of its wide variety of expressive emotions, we selected a broad range of classical music stimuli from baroque to romantic. Furthermore, the rigorous experimental procedure used to assess expressive reactions expounds upon trends that have been proposed in literature [2]. Based on these premises, this study aims to design, test and explore the possible uses of a multi-method approach utilizing IVEs, behavioral analysis and semi-structured interviews to analyse musicians' performance and their sensitivity to the social context.

[*]Corresponding author. Email: Donald.Glowinski@unige.ch

2. Background

2.1 Controlling environmental variables

Musicians show high sensitivity to specific elements of their performance contexts such as light, space, or audience's emotional reactions. Musicians' capacity to adapt their behavior to different social and environmental situations and disturbances is uniquely developed [3].

The emerging field of creative technology may gain from a better understanding of the underlying processes that help musicians coping with such a variety of situations. Capturing musician's expressive behavior in a naturalistic setting such as a concert hall, is from a technological perspective, highly complex. The work done at the CasaPaganini Institute is remarkable from this point of view because they used a multi-modal setup which allows researchers to record musicians in real-time [4]. However, the capacity to directly modify environmental features like light, space, and audience characteristics to see how musician can adapt to these changes is still understudied.

One unique environment that allows for this kind of research exists in the "Espace de Projection (Espro)" of the IRCAM (Institute de Recherche et Coordination Musique Acoustique, Paris). This environment has been conceived and built so that researchers can manipulate more easily form, volume, and acoustical properties [5]. One of the few studies done in this environment [6] attempts to describe the behavior of a pianist playing the same piece of music with different acoustics. Seven professional pianists playing on a Disklavier MIDI grand piano were recorded with synchronous digital audio and midi systems. From this data, a statistical analysis was performed to understand the extent to which performance attributes are correlated to the perceptual dimensions of room acoustics. However, this modular environment is unique and experimental modification remains time consuming and prevents researchers from conducting studies with numerous participants.

When recording musicians in an experimental context, laboratory settings are normally used; however, acoustics are severely compromised sound, artificial to the musician participants. Therefore, the use of an Immersive Virtual Environment (IVE) can be key in developing a laboratory setting that is both easily controlled as well as acoustically and environmentally naturalistic. A virtual environment is defined as a multisensory experience of location or set of locations through artificial electronic means [7]. The observer can navigate in this environment and interactively manipulate it [8]. A virtual environment is a combination of multiple features [9]. Among these we can mention a three-dimensional viewing, a dynamic interface, a closed-loop interaction, an Ego-centered frame of reference, a multimodal interaction and a head-mounted display and tracking. All those features are used to maximize the sense of immersion. To this end, virtual environments are generally displayed by surrounding perceptually the observer in what is called an immersive virtual environment [10]; e.g., either by a CAVE^{TM} system, back-projecting computer that generate visual imagery onto translucent walls, floor and ceiling of a cubical room [11] or by means of shutter glasses providing stereoscopic simulation to create 3D rendering.

IVE is now being used in psychological research [10], [12], and can provide greater experimental control, more precise measurement, ease of replication across participants, and high ecological validity, making it extremely attractive for researchers. This technology also gives researchers the ability to control many stimulus variables, like egocentric distance of an object or its angular size [10]. This also allows for an accurate control of key dimensions affecting behavior and performance and results in emerging research for people suffering from posttraumatic stress [13], phobias such as fear of flying [14], and arachnophobia [15]. They recently also consider organizational research on negotiations [16], or help in determining the conditions that foster prosocial behavior [17], or most related to our thematic of interest, they create a control factor that addresses one's capacities to speak in front of an audience [18]. Social context effects have otherwise been evaluated using interactive setup based on non-invasive technologies (e.g., kinect) [19], [20].

Surprisingly, when considering music performance related to IVE, most studies are oriented toward performative and artistic objectives and include expressive behavioral component in their research, without any relationship with social psychology. Hamilton et al. [21], in particular, developed a series of multi-modal networked musical performance environments where musical works are controlled by motion and gestures generated by in-engine performer avatars in a virtual environment. Most research in this field focused on the quality of the spatialized audio rendering system. The aim of the work, however, is to cultivate an appropriate trade-off between audio quality, spatial precision, and performance, and to obtain convincing acoustic room simulation [22]. Few studies have considered virtual environments in music to test their use in managing music performance anxiety [23] [24]. Benefits to musicians after being exposed in these settings is blended. Psychological and physiological responses to performance stress have shown high inter-individual variability, revealing the multifaceted nature and impact of performance anxiety on musicians. A recent study by Williamon et al. [3] explored how interactive, simulated performance environments could be used to train musicians to perform under pressure without focusing exclusively on anxiety symptoms. The study has focused on responses to simulated recital and audition scenarios with relatively neutral and "well-behaved" virtual observers. Results obtained with 11 advanced violin students showed that simulated environments offered realistic experience of performance contexts and were rated useful for developing performance skills.

Drawing upon this last study, we designed a virtual environment in which only a selective abstraction of key environmental features are provided (consistent across a wide range of Western classical performance venues). Previous studies suggested that a simulated environment only requires few environmental cues directly relevant for the task at hand, as the other imitated environmental features go unnoticed [3].

Our main additional contributions with respect to the performance simulator developed by Williamon et al. consisted of (i) further use of the full resources of the virtual environment to increase the level of immersion experienced by the participants instead of projecting pre-recorded video sessions of life-sized observers onto a screen; (ii) designing a flexible interactive solution for controlling audience to includes a wide range of positive and negative observer responses (iii) combining qualitative and quantitative assessment of the musician performance through semi-structured interview and automatic behavior analysis.

2.2. Music performance and body expressivity

Musicians make bodily gestures and move during performances to express their musical intentions [25]. According to the literature in embodied cognition, music is implicitly associated with gesture on many levels [26]. This association can be functional, for example, when movement relates simply to the production of sound. However, this association can also be of a higher level when movement is related to the emotion that the musician means to convey (e.g., Expressive Gesture, see [4], [27]). To address the communication of emotion in music, Juslin et al. [28] devised a framework to disentangle how body expressivity is encoded by the musician and how it is decoded by an external observer. The efficiency of the emotion communication processes is derived from the mapping between musicians and spectator cues. Using Juslin's model as a scaffold, this study considers how the embodiment of expressivity manifests itself through specific movement features.

Seminal work by Davidson [2] highlights the fact that certain perceptual elements of a musician's gesture are sufficient for the audience to identify a musician's intended expressivity. Gestures contribute to communication of information to the audience as well as to other musicians. For example, information about emotions felt by a particular musician that she or he wants to transmit [25] or about elements of the musical structure [29] are usually conveyed through expressive gestures. Note that the gesture expressivity is not limited to what the musician wants to convey explicitly, but also all implicit information related to the modification of gestures related to emotional processing. A key experimental procedure to investigate changes in expressivity has been initially proposed by Davidson [30]. She suggested three levels of expressivity to clearly identify the relationship between expressive gesture and music performance. These levels are: metronomic (with minimal expressive interpretation of the music), concert-like (consistent with public performance), and emphatic (overstating all aspects of the expressive features). Recent studies [31] have demonstrated that a musician asked to play a song according to these three expressive intentions, tends to move more and accentuates their gestures as they play in a more expressive way. Researchers in this area have recently begun to augment perceptual research by expanding their knowledge and analysis with a quantitative approach. For instance, an increasing number of studies have used motion capture techniques [4], [32]. This system allows for a fine-grained analysis of movement and helps researchers quantify the physical parameters of gesture (amplitude, velocity, speed, smoothness). Several studies have demonstrated that head, shoulders, arms and torso are the body parts which transmit the most expressive information (See [31], [33]–[35]). Specifically, recent findings revealed that the most expressive musical performances are characterized by an increasing amplitude, speed and acceleration in the mentioned body segments [35]. In this context, kinetic energy is particularly crucial in understanding broad, unrefined body reactions and assist in getting a first understanding of the behavioral underpinnings of expressive gesture. Specifically, kinetic energy may help in understanding how musicians cope with the audience [19],[26],[28].

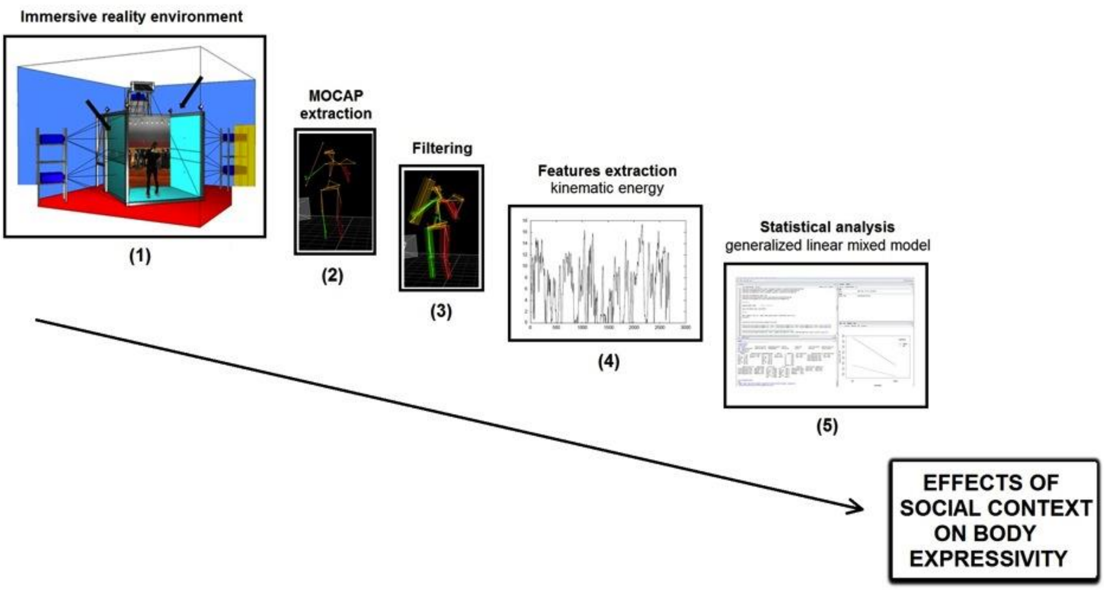


Figure 1. Visualization of the data processing and analysis steps in evaluation of the difference in musician's behavioral response according to the expressive conditions (metronomic, concert-like and emphatic) and social contexts (absent, disengaged and engaged audience).

3. Method

3.1 Participants

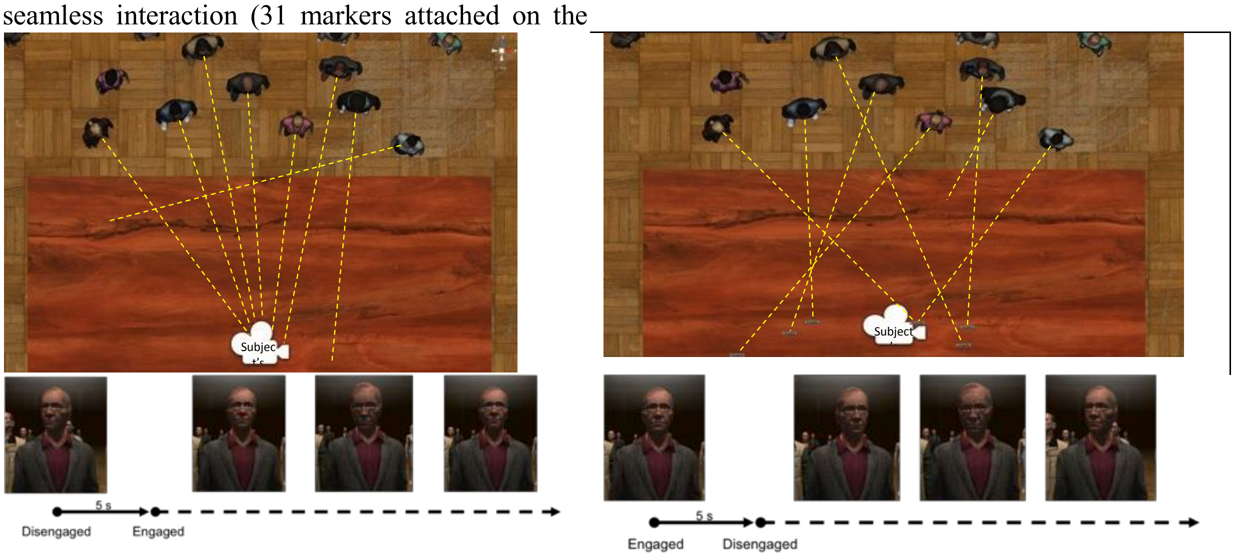
Four violinists from the Geneva University of music participated to this study (3 females, mean age = 22).

3.2 Musical Stimuli and Procedure

Before the study, participants received 3 excerpts from 3 different pieces (one from Beethoven, one from Tchaikovsky and one from Bach) to learn by heart (duration ~ 30 sec.). Excerpts were chosen because they were not positioned in the middle of a musical phrase, each excerpt's end matched the end of the phrase, and successfully conveyed a musical thought even when the excerpt was decontextualized. Moreover, musicians received a document describing the three expressivity modes in order to familiarize themselves with them. They were asked, in a pseudorandom order, to interpret each music piece twice according to the three selected expressive intentions: metronomic, concert-like and emphatic [8]. Each piece was performed in three environmental conditions: (i) engaged, (ii) disengaged audience and (ii) empty space. They were also asked to come wearing black or dark slim clothes, without glasses or make-up, and with their hair tied back. Just before starting to play and at the end of each excerpt, musicians did the T-pose as a reference. They were also asked to stay in the middle of the stage during each performance. Semi-structured interviews were conducted after each performance with a fairly open framework to allow for focused, conversational, two-way communication between the experimenter and the musician.

3.3 Apparatus

The Brain and Behavior Laboratory-Immersive System (BBL-IS) has 4 sides presenting seamless and perspectively coherent 3D images. The acrylic coated screens from DaLite are the most visible part of the system; they present the image with a high contrast ratio and brightness transfer from the video projectors. The biggest is 2.8 m wide and is 2.4 m in height. There are seven TITAN QUAD 3D from Digital Projection capable of showing high-resolution images at a very high speed (120 image per second). The screens offer brightness up to 1605 cd/m and a resolution of 4 Arcmin/OLP. The BBL-IS simulation scenes are powered by a workstation cluster structured with one master node and seven client nodes. Each node is a Z800 workstation from HP company with a 12-core processor, 12 Gb of RAM and a professional nVidia Quadro graphic card capable of driving stereoscopic 3D devices. To react to each user's movements, the BBL-IS uses a Vicon optical motion tracking system composed of height Bonita 3 cameras. The system is able to perform full-body motion capture in real time for seamless interaction (31 markers attached on the body, 3 markers on the violin and 2 markers on the bow). Selected body parts have been chosen following recent literature on analysis of music performance; for example, the head is known to convey expressive information, as well as arms and torso [13]. We recorded the body movement to examine the dynamic of movements according to the virtual situation (see Figure 1).



(A) Engaged audience **(B) Disengaged audience**

Figure 2. In the engaged audience (A), a large part of the avatars is gazing at the musician, a small part of them looks at random targets; in the same time, a small part of the avatars looks at the musician for brief moments. In the disengaged audience (B), each avatar looks at a random target moving in random direction around the musician [range: ±4 m (left-right); ±3 m (up-down); ±1 m (front-back)] ; in the same time, a small part of the avatars looks at the musician for brief moments.

3.4 Procedure to manipulate virtual audience (engaged / disengaged)

Three conditions related to the social attitudes are adopted by avatars in a 3-D virtual immersive environment (empty room, engaged and disengaged audience). Modeling of the audience state (engaged vs. disengaged) carefully considers smooth transitions of avatar's behavior. Specifically, engaged behavior, or audience attention increases through convergence of individual gazes [36] focusing on the musician. On the other hand, a disengaged audience is rendered by allowing the gaze of avatar to wander around as normally observed in distracted people.

To increase the realism of avatar's behavior, each avatar has random idle animation of their body while looking at the musician to approximate usual fluent behavior of people during a concert. Furthermore, the group's behavior is animated when a few of the avatars do not follow the current trend as to avoid unnatural uniformity [37]. For example, in the engaged audience case, while most of the avatars are focusing on the musician, a small part (5%-10% distributed over the different rows) looks at others in the audience or away from the performer brief moments. Finally, the distance between the musician and the audience mirrors the usual concert environment of the participants (about 3 meters between the musician and the first row of spectators).

Pseudo-code for the transition from disengaged to engaged audience:

Avatars[] : list of avatars *Steps[]* : list of step range *InterStepTimeRange* : time range between steps *Target* : object to look at when engaged	Index = 0 Foreach *Step* in *Steps* *Count* = Random(*Step.min, Step.max*) For *Index*: *Index* -> *Index+Count* *Avatars*[*Index*].LookAt(*Target*) EndFor Wait(Random(*InterStepTimeRange.min, InterStepTimeRange.max*)) EndForeach

Figure 3. Example views of the conditions: empty room, engaged/disengaged audience.

Figure 4. Details of a disengaged audience.

3.5 Experimental design

The experimenters have applied a factorial design where each experimental condition in the design represented a gradation in expressivity (within variable, 3 levels: metronomic, concert-like and emphatic), a social context (within variable, 3 levels: empty room, engaged, disengaged audience) and a music style (within variable, 3 levels: Bach, Tchaikovsky, Beethoven). The dependent variable refers to the kinetic energy of the musician's body movement.

4 Analysis and results

4.1 Focus on kinetic energy of body movement

This analysis consisted of the approximation of kinetic energy of each body segment as well as the computation of the overall kinetic energy of the musician. We considered the 31 markers placed on the upper-part of the musicians' body (see figure 1) and from the 2 markers placed on the violin's bow. Mocap data have been recorded at a sampling rate of 120 Hz.

The Kinetic Energy (KE) is the overall energy spent by the user during movement, estimated as the total amount of displacement in all of the tracked points. Given the 3D musician and tracking information, let:

$$v_i(f) = \sqrt{\dot{x}_i(f) + \dot{y}_i(f) + \dot{z}_i(f)} \quad (1)$$

denote the velocity's magnitude of the *i-th* tracked point at time frame *f*. We then define *KE(f)* index at the frame f, as an approximation of the body kinematic energy, the weighted sum of each joint's kinetic energy.

4.2 Statistical analysis revealing personal bias of each musician

The kinetic energy measure was calculated for the four musicians. A linear mixed model (LMM) approach was chosen as statistical model to compare musicians' kinetic energy values across conditions. LMM was chosen as it allows for handling correlated data and unequal variance [38]. To control the inflation of type I error probability due to multiple comparisons, the Bonferroni correction was applied to adjust the p-values required for statistical significance.

Applied on the full set of 220 samples (55 for each musician), LMM identified significant main effects of Expressivity (Emphatic > Concert-like > Metronomic, $p < 0.001$), Musician ($p < .05$) and Music score (Bach > Tchaikovsky > Beethoven, $p < .05$) and a marginal effect of the Social context (Empty > Disengaged audience > Engaged audience, $p = .057$). Two significant interaction effects have been identified: Expressivity × Musician ($p < .05$) and Expressivity × Musician × Social ($p < 0.05$). Bonferroni-corrected post-hoc analyses were performed to assess specific difference among interaction effects. We review in the following the main effect of Expressivity and the related Expressivity × Musician × Social.

Main effects of Expressivity

Results showed that the experimental conditions of Expressivity had a significant main effect on kinetic values: considering musicians altogether, for all social contexts, over all trials, kinetic energy values in the emphatic were significantly higher than in concert and metronomic conditions [$F_{(2,110)}=229.578$, $p < .05$], see Figure 5.

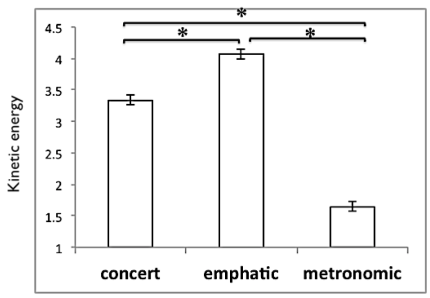
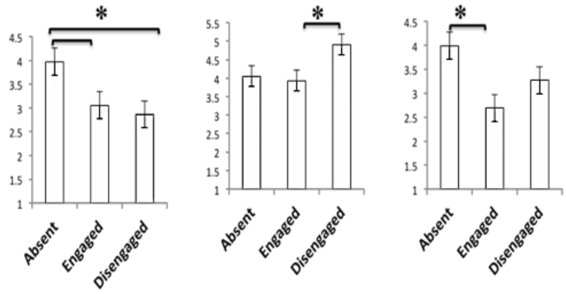


Figure 5. Main effect of the Expressivity condition on kinetic energy values.

Expressivity × Musician × Social Interactions
Post-hoc analysis of the Expressivity x Musician x Social interaction revealed that kinetic energy values of musician 1 is significantly higher when there is no audience in the concert-like condition, whereas this interaction effect happens for musician 3 in the emphatic condition.

It is also in an empty room that values of musicians 4 are significantly higher, but only in the emphatic conditions. Values of musician 2 are significantly higher in front of a disengaged audience. On the opposite, Musician 3 shows no significant differences whatever the social context is (see Figure 6).



a b c

Figure 6. Main effects of social contexts: the bias of individual sensitivity: (a) musician 1 in concert style, (b,c) musician 2 and 3 in emphatic style.

4.3. Outcomes from the semi-structured interviews

Semi-structured interviews were instructive in investigating how musicians perceived each simulated social context. The four violinists found the simulations to be realistic and convincing. They all mentioned that their sensitivity to the avatars' behavior increased over the experiment. They all started the experiment noticing some avatars that captured their attention (e.g., "the woman on the left") before becoming aware that the audience as a whole could display a positive or a negative attitude. Musicians also mentioned the "paradoxical" situations of playing the selected music piece with a metronomic style, i.e., as if they were rehearsing for themselves, in front of an engaged audience. This discrepancy between playing style and social context was actually key in our protocol to put in evidence the possible impact of social context on musician's behavior. With respect to the "invasiveness" of the setup, musicians did not complain about the on-body markers that were disseminated over their body and on their instrument. Two, however, commented on the minor inconvenience of wearing the stereo glasses that prevented them to clearly see their hand on the violin neck and fingerboard.

In terms of potential uses, musicians positively acknowledged the virtue of using such immersive virtual environment to develop their performance skills and highlight their weakness. They complained about not having enough opportunities to be exposed to performance situations (e.g., on stage, in front a jury) and they clearly enjoyed the idea of accessing a realistic simulation with pre-defined outcomes combined with an in-depth quantitative monitoring/profiling of their performance.

4.4. Discussion

This study revealed empirical evidence of body kinematic changes underlying emotional expressivity. These results are consistent with the literature, see for example [2], [39] showing that the three expressive styles are characterized by different body movements and especially values of kinetic energy that are increasing gradually from metronomic, concert-like to emphatic conditions.

However, these results also show that there are some differences in each performer's sensitivity towards the social contexts presented. For example, some of the musicians were more affected by the audience's motivational state (engaged or disengaged), while others were more impacted by the simulation of playing to an empty room. One participant in our sample has not been impacted by any of the social stimuli though she was clearly aware of the variations in avatars behavior (see section 4.3). This first set of results confirmed that kinetic energy can be successfully considered as a right manipulation check for our novel multi-method approach of musician performance.

Three factors may explain such differences in social context sensitivity. This difference may revolve around issues of attention. Keller et al. [33], for example, revealed how individual differences in auditory imagery and movement kinematics can influence one's capacity to play with others [40]. Musicians may become so deeply focused on the task at hand (e.g. playing with correct technique, being musical, applying the expressive style requested, etc.) that they may, in turn, appear less sensitive to social environment.

Secondly, social context sensitivity could be influenced by inter-individual differences in empathetic responses. Empathy, one's proclivity to adopt another's perspective could also explain the observed difference [41]. From this perspective, the kinetic-related results may seem counterintuitive if you do not consider the personal self-reflexivity of each musician. One might expect for example that the musician would be more attentive and engaged with an attentive audience, and more stiffer and inhibit their movement with a disengaged one. However, results revealed an increased tendency in the kinetic energy when the audience was absent or disengaged, or when expressive style was incongruent with the actual situation (e.g., metronomic in front of an engaged audience). These results were confirmed by the participants during the post-performance interview, when they referred to their difficulty to face an "empty" large room or to cope with a disengaged, annoyed audience, by trying to impact upon them and compensating by a further physical engagement.

Performance anxiety may also be a factor to consider when exploring the differences in social context sensitivity. This large inter- intra- individual variability in how people experience the social context actually reminds the large difference observed with the psychological symptoms of performance anxiety [3]. Since the music students have a high level skills such as fine motor dexterity, coordination, attention and memory, as well as the aesthetic and interpretive skills required of a musician, high level music students often develop a sense of perfectionism. Because their training is extremely demanding they partake in intense self-critique and develop acute anxiety in performance scenari. This distress impairs efficient functioning when the musician engages in goal directed behavior and increases the extent to which processing is influenced by outside stimuli. Furthermore, this anxiety increases participants' attention to threat-related stimuli and may inhibit an appropriate reaction to the social environment [42].

Suggested Improvements. Our main objective was to evaluate a novel experimental procedure that combines the use of behavioral analysis, semi-structured interviews and an immersive virtual environment to give a comprehensive view of the musician performance and his or her sensitivity to a specific environment. These preliminary results revealed that the selected experimental procedure is worth being followed. Further studies are however needed to establish the benefits objectively by considering the following points.

A larger population of musicians should be recruited. Pre-post experiments that compare the amelioration of performance before and after the use of IVE should be carried out in the relevant contexts (e.g., concert or public examination) to evaluate whether the skills acquired during the simulation can be transferred in real-world situations. Behavioral monitoring should be combined with psychophysiological and questionnaire-based measures to systematically address the psychological specificities of each musician (e.g., personality through the Big Five Inventory by [43], symptoms of anxiety by [44].

The degree of immersion as experienced by the musician should also be objectively assessed. As stated in [12], there are dimensions such as the extent to which a display system can deliver an inclusive, extensive, surrounding and vivid illusion of virtual environment to a participant. Specifically, the sense of presence seems key for such an evaluation. The ITC-Sense of Presence Inventory (ITC-SOPI [45]) in particular has emerged as the reference as it taps all the possible manifestations of different content areas (e.g., sense of space, involvement and attention). Such information about the musician experience could provide insights to compare simulated environments. In this study, the choice of the CAVE^{TM} system with respect to Head-Mounted-Display (HMD) solutions was mainly motivated by the non-invasiveness of the setup. Actually, the actual HMD available on the market would not allow the musician to play normally as in the real world. In addition, it would have been uncomfortable for them not to be able to see their own body when playing (see section 3.4). A virtual representation of their body could have been implemented and rendered in real-time in the system. However, due to the limited resources of the system already involved in the real-time animation of an audience of avatars, this further implementation could have increased the latency of the interaction and therefore affected a crucial component of the sense of presence and effective immersion. A new generation of HMD device, lighter and less-invasive should however be considered as they may offer a more immersive and cheaper solution, that allows for more convenient and affordable replication of our protocol. The approaches mentioned have resulted in quality, portable simulations that are lower in cost. Thus, this protocol will be widely accessible to the research community.

The number of avatar (= 33) has been decided based on the following aspects: the limit of the computational resources which allows for a total number of 33 different, realistic expressive avatars whose behavior could be manipulated in real-time.

The realism of the scene which combines the dimension of the virtual space of a concert hall that would be familiar with the musicians' experience and the ratio between the number of avatar and the available space which may refer to concert-like environment in the real world (e.g., neither too close nor too distant one from another). This configuration allowed for rendering a minimal but realistic variety of behavior in the audience. Future experiment may include the manipulation of the number of avatar with respect to the rendered space to evaluate extreme but possible situations (e.g., a room with nearly nobody attending vs. crowded environment where you can hardly distinguish space between the participants). The acoustic presence of the avatar should also be considered: cough sounds, noisy breathing, chattering, murmuring or explicit judgments that are known to disrupt the musician attention.

5 Conclusion

This study has presented the unique experimental procedure developed to investigate the effects of an audience's attention on a musician's body expressivity. By exploiting the Virtual Immersive Environment (IVE) and a motion capture system, we managed to record and analyse the musician's expressive behavior as they relate to audience's motivational state. Preliminary results revealed intriguing inter-individual differences in term of context-sensitivity, confirming previous findings related to performance anxiety [3].

Future work includes the development of a novel, in-depth analysis of the interaction between musician and the audience by describing expressive gesture with a higher number of features such as expansion and symmetry of posture. Musician's experience should also be investigated through formal self-report that draws upon research on empathy [46], personality [43], state anxiety [44] and sense of presence [45].

In light of our findings, the experimental setup may ultimately result in a training tool for music schools. The system developed for this study can actually be easily adapted to study context-sensitivity in real-time and for systematic manipulation of the different concert conditions. However, finding unique ways in which to utilize the simulations to create structured immersive experiences that vary for people with different skill levels, types of performance anxiety, and levels of exposure to performances can be difficult. To cultivate broad and specific usages for this type of simulation training, it is vital that investigations continue in this area.

From the musician's point of view, a better understanding of the factors influencing the irrational apprehension felt before performing on stage is key. As it allows to reduce the impacts of anxiety on their playing technique and focus. An innovative field of research can be expanded and lead to significant developments in musical training by helping participants to face their reactions confidently and to elaborate novel mental and behavioral strategies for coping with audience.

Acknowledgements.

We thank the musicians from the Geneva University of Music (HEM) for their participation to the experiment. This study has been achieved with the partial support of the NCCR Affective Sciences.

References

[1] A. R. Jensenius, "Some video abstraction techniques for displaying body movement in analysis and performance," *Leonardo*, vol. 46, no. 1, pp. 53–60, 2013.

[2] J. W. Davidson, *Bodily communication in musical performance*. Oxford University Press, USA, 2005.

[3] A. Williamon, L. Aufegger, and H. Eiholzer, "Simulating and stimulating performance: introducing distributed simulation to enhance musical learning and performance," *Front. Psychol.*, vol. 5, 2014.

[4] D. Glowinski, M. Mancini, R. Cowie, A. Camurri, C. Chiorri, and C. Doherty, "The movements made by performers in a skilled quartet: a distinctive pattern, and the function that it serves," *Front. Psychol.*, vol. 4, p. 841, 2013.

[5] V. M. A. Peutz, "The Variable Acoustics of the Espace de Projection of Ircam (Paris)," in *Audio Engineering Society Convention 59*, 1978.

[6] S. Bolzinger, O. Warusfel, and E. Kahle, "A study of the influence of room acoustics on piano performance," *J. Phys. IV*, vol. 4, no. C5, pp. C5–617, 1994.

[7] K. Carr and R. England, *Simulated and virtual realities: Elements of perception*. CRC Press, 1995.

[8] W. Barfield and T. A. Furness, *Virtual environments and advanced interface design*. Oxford University Press, 1995.

[9] C. D. Wickens and P. Baker, "Cognitive issues in virtual reality.," 1995.

[10] J. Blascovich, J. Loomis, A. C. Beall, K. R. Swinth, C. L. Hoyt, and J. N. Bailenson, "Immersive virtual environment technology as a methodological tool for social psychology," *Psychol. Inq.*, vol. 13, no. 2, pp. 103–124, 2002.

[11] C. Cruz-Neira, D. J. Sandin, and T. A. DeFanti, "Surround-screen projection-based virtual reality: the design and implementation of the CAVE," in *Proceedings of the 20th annual conference on Computer graphics and interactive techniques*, 1993, pp. 135–142.

[12] M. V. Sanchez-Vives and M. Slater, "From presence to consciousness through virtual reality," *Nat. Rev. Neurosci.*, vol. 6, no. 4, pp. 332–339, 2005.

[13] J. Difede, J. Cukor, N. Jayasinghe, I. Patt, S. Jedel, L. Spielman, C. Giosan, and H. G. Hoffman, "Virtual reality exposure therapy for the treatment of posttraumatic stress disorder following September 11, 2001.," *J. Clin. Psychiatry*, vol. 68, no. 11, pp. 1639–1647, 2007.

[14] B. O. Rothbaum, L. Hodges, S. Smith, J. H. Lee, and L. Price, "A controlled study of virtual reality exposure therapy for the fear of flying.," *J. Consult. Clin. Psychol.*, vol. 68, no. 6, p. 1020, 2000.

[15] S. Bouchard, S. Côté, J. St-Jacques, G. Robillard, and P. Renaud, "Effectiveness of virtual reality exposure in the treatment of arachnophobia using 3D games," *Technol. Health Care*, vol. 14, no. 1, pp. 19–27, 2006.

[16] M. Schmitt and T. Rist, "Avatar arena: Virtual group-dynamics in multi-character negotiation scenarios," *Lect. Notes Comput. Sci.*, pp. 358–358, 2003.

[17] O. Gillath, C. McCall, P. R. Shaver, and J. Blascovich, "What can virtual reality teach us about prosocial tendencies in real and virtual environments?," *Media Psychol.*, vol. 11, no. 2, pp. 259–282, 2008.

[18] M. Slater, D.-P. Pertaub, and A. Steed, "Public speaking in virtual reality: Facing an audience of avatars," *Comput. Graph. Appl. IEEE*, vol. 19, no. 2, pp. 6–9, 1999.

[19] M. Mancas, R. B. Madhkour, D. De Beul, J. Leroy, N. Riche, Y. P. Rybarczyk, and F. Zajéga, "Kinact: a saliency-based social game," in *Proceedings of the 7th International Summer Workshop on Multimodal Interfaces eNTERFACE11*, 2011, vol. 8.

[20] Y. Rybarczyk, G. Carrasco, T. Cardoso, and I. P. Martins, "A serious game for multimodal training of physician novices," *ICERI2013 Proc.*, pp. 4944–4949, 2013.

[21] R. Hamilton, J.-P. Caceres, C. Nanou, and C. Platz, "Multi-modal musical environments for mixed-reality performance," *J. Multimodal User Interfaces*, vol. 4, no. 3–4, pp. 147–156, 2011.

[22] W. Krebber, H.-W. Gierlich, and K. Genuit, "Auditory virtual environments: basics and applications for interactive simulations," *Signal Process.*, vol. 80, no. 11, pp. 2307–2322, 2000.

[23] E. K. Orman, "Effect of virtual reality graded exposure on anxiety levels of performing musicians: A case study," *J. Music Ther.*, vol. 41, no. 1, pp. 70–78, 2004.

[24] Bissonette, J., Dube, F., Provencher, M. D., and Moreno Sala, M. T., "The effect of virtual training on music performance anxiet," in *Proceedings of the International Symposium on Performance Science 2011*, 2011, pp. 585–590.

[25] S. Dahl and A. Friberg, "Visual Perception of Expressiveness in Musicians' Body Movements," *Music Percept.*, vol. 24, no. 5, pp. 433–454, 2007.

[26] M. Leman, *Embodied Music Cognition and Mediation Technology*. Mit Press, Cambridge, 2007.

[27] C. Palmer and D. Deutsch, *Music performance: Movement and coordination*. Elsevier, 2012.

[28] P. N. Juslin and P. Laukka, "Communication of Emotions in Vocal Expression and Music Performance : Different Channels, Same Code?," *Psychol. Bull.*, vol. 129, no. 5, pp. 770–814, 2003.

[29] M. M. Wanderley, B. W. Vines, N. Middleton, C. McKay, and W. Hatch, "The Musical Significance of Clarinetists' Ancillary Gestures : An Exploration of the Field," *J. New Music Res.*, vol. 34, no. 1, pp. 97–113, 2005.

[30] J. W. Davidson, "Visual perception of performance manner in the movements of solo musicians," *Psychol. Music*, vol. 21, no. 2, pp. 103–113, 1993.

[31] M. R. Thompson and G. Luck, "Exploring relationships between pianists' body movements, their expressive intentions, and structural elements of the music," *Music. Sci.*, vol. 16, no. 1, pp. 19–40, 2012.

[32] B. Burger, S. Saarikallio, G. Luck, M. R. Thompson, and P. Toiviainen, "Relationships between perceived emotions in music and music-induced movement," *Music Percept. Interdiscip. J.*, vol. 30, no. 5, pp. 517–533, 2013.

[33] M. Sakata and S. Wakamiya, "An Analysis of Body Movement on Music Expressivity Using Motion Capture," in *Intelligent Information Hiding and Multimedia Signal Processing, 2009. IIH-MSP'09. Fifth International Conference on*, 2009, pp. 1172–1176.

[34] M. Nusseck and M. M. Wanderley, "Music and motion—How music-related ancillary body movements contribute to the experience of music," 2009.

[35] A. G. Van Zijl and G. Luck, "Moved through music: The effect of experienced emotions on performers' movement characteristics," *Psychol. Music*, vol. 41, no. 2, pp. 175–197, 2013.

[36] R. E. Guadagno, J. Blascovich, J. N. Bailenson, and C. Mccall, "Virtual humans and persuasion: The effects of agency and behavioral realism," *Media Psychol.*, vol. 10, no. 1, pp. 1–22, 2007.

[37] M. Garau, M. Slater, V. Vinayagamoorthy, A. Brogni, A. Steed, and M. A. Sasse, "The impact of avatar realism and eye gaze control on perceived quality of communication in a shared immersive virtual environment," in *Proceedings of the SIGCHI conference on Human factors in computing systems*, 2003, pp. 529–536.

[38] R. A. McLean, W. L. Sanders, and W. W. Stroup, "A unified approach to mixed linear models," *Am. Stat.*, pp. 54–64, 1991.

[39] A. Camurri, I. Lagerlöf, and G. Volpe, "Recognizing Emotion from Dance Movement: Comparison of Spectator Recognition and Automated Techniques," *Int. J. Hum.-Comput. Stud. Elsevier Sci.*, vol. 59, pp. 213–225, juillet 2003.

[40] P. E. Keller, S. Dalla Bella, and I. Koch, "Auditory imagery shapes movement timing and kinematics: Evidence from a musical task.," *J. Exp. Psychol. Hum. Percept. Perform.*, vol. 36, no. 2, p. 508, 2010.

[41] G. Novembre, L. F. Ticini, S. Schütz-Bosbach, and P. E. Keller, "Motor Simulation and the Coordination of Self and Other in Real-Time Joint Action.," *Soc. Cogn. Affect. Neurosci.*, p. nst086, 2013.

[42] M. W. Eysenck, N. Derakshan, R. Santos, and M. G. Calvo, "Anxiety and cognitive performance: attentional control theory.," *Emotion*, vol. 7, no. 2, p. 336, 2007.

[43] O. P. John, E. M. Donahue, and R. L. Kentle, "The big five inventory," *Berkeley Univ. Calif. Berkeley Inst. Personal. Soc. Res.*, 1991.

[44] C. D. Spielberger, *State-Trait Anxiety Inventory*. Wiley Online Library, 2010.

[45] J. Lessiter, J. Freeman, E. Keogh, and J. Davidoff, "A cross-media presence questionnaire: The ITC-Sense of Presence Inventory," *Presence*, vol. 10, no. 3, pp. 282–297, 2001.

[46] R. N. Spreng, M. C. McKinnon, R. A. Mar, and B. Levine, "The Toronto Empathy Questionnaire: Scale development and initial validation of a factor-analytic solution to multiple empathy measures," *J. Pers. Assess.*, vol. 91, no. 1, pp. 62–71, 2009.