Article

Playing for a Virtual Audience: The Impact of a Social Factor on Gestures, Sounds and Expressive Intents

Simon Schaerlaeken *, Didier Grandjean † and Donald Glowinski †

Neuroscience of Emotion and Affective Dynamics Lab, Faculty of Psychology and Educational Sciences and Swiss Center for Affective Sciences, University of Geneva, 1205 Geneva, Switzerland; didier.grandjean@unige.ch (D.G.); donald.glowinski@unige.ch (D.G.)

* Correspondence: simon.schaerlaeken@unige.ch; Tel.: +41-78-943-7215

† Co-senior authors.

Abstract: Can we measure the impact of the presence of an audience on musicians' performances? By exploring both acoustic and motion features for performances in Immersive Virtual Environments (IVEs), this study highlights the impact of the presence of a virtual audience on both the performance and the perception of authenticity and emotional intensity by listeners. Gestures and sounds produced were impacted differently when musicians performed at different expressive intents. The social factor made features converge towards values related to a habitual way of playing regardless of the expressive intent. This could be due to musicians' habits to perform in a certain way in front of a crowd. On the listeners' side, when comparing different expressive conditions, only one congruent condition (projected expressive intent in front of an audience) boosted the participants' ratings for both authenticity and emotional intensity. At different values for kinetic energy and metrical centroid, stimuli recorded with an audience showed a different distribution of ratings, challenging the ecological validity of artificially created expressive intents. Finally, this study highlights the use of IVEs as a research tool and a training assistant for musicians who are eager to learn how to cope with their anxiety in front of an audience.

Keywords: immersive virtual environment; music; performance; expressiveness; authenticity; emotions; social

1. Introduction

Musical performances are the result of a complex interactive phenomenon between the musicians and the audience who attends and appreciates them. However, the understanding of this complex phenomenon could benefit from a scientific approach. First, experiments on musical performances could bring insights into how to deal with musicians' performance anxiety, as musicians' training involves learning to cope with stress during performances. However, opportunities to train musicians on how to regulate their emotions during concerts are limited. The most effective method seems to combine relaxation training with exposure to stressful events (to build up realistic expectations of what will be felt during performances) and cognitive restructuring (to counteract on self-handicapping habitual thoughts and attitudes) [1]. In fact, musicians' experience and ability to resist stress mainly depends on the opportunities they have during their career to perform during live performances and having repeated peer sessions. Second, researchers interested in the effect of audiences on musicians and their body language are usually left with uncontrolled and highly variable situations when they study concerts. On a practical side, recording physiological and motion capture data during concert can hinder musicians during concerts, thus impacting the quality of their performance, making concerts a challenging environment for scientific research.

Immersive Virtual Environments (IVEs) have been used by researchers to control for these complex parameters [2]. IVEs allow researchers to create realistic virtual environments with unlimited configurations that can adapt in real time to users' behavior. It allows for researchers to control for environmental parameters such as the audience, the space between different objects, and the lighting. IVEs offer researchers the opportunity to compute perceptual analyses and create new roads for computational development. This environment combined with a motion capture set-up was used in this study to precisely record expressive musical gestures and explore the possible underlying behavioral mechanisms impacted by the presence of an audience in the context of different levels of expressive intents.

1.1. Controlling Environmental Variables in Virtual Reality

Musicians develop unique abilities allowing them to adapt their behavior to different social and environmental contexts [2]. Consequently, it is necessary to control for many parameters when recording musical performances, e.g., sounds, lights, the presence of other musicians or audience. Thus IVEs represent a key methodological tool for psychological research as it can provide greater experimental control, more precise measurements, ease of replication across participants, and high ecological validity, making it extremely attractive for researchers [3,4]. They can also provide live feedbacks to participants. Virtual reality and IVEs have been used in research for patients suffering from post-traumatic stress disorder [5–7] and for treating phobias such as fear of flying [8] and arachnophobia [9]. The use of such technology has also been proven to be efficient for treating social phobia and reducing the fear of the public speaking [10,11]. Few studies have considered using virtual environments in music to study performance anxiety [12,13]. In a recent study by Williamon et al. (2014), musicians were invited to enter IVEs to train their ability to cope with the pressure of performing live and rated such tool useful for developing their performance skills and very realistic. It demonstrated that simulated environments are able to offer a realistic experience of performance contexts.

1.2. Music Performance: From Sound to Gesture

Communicating and expressing emotions through music is the main reason why people engage in this activity [14]. Evidence points at a general ability to accurately recognize emotions expressed with music (e.g., happiness, sadness, and nostalgia) [15–21]. Regardless of cultural background or musical training, people are generally able to name the intended emotion, providing evidence for an universal recognition not only of the expression of basic emotion but also of more complex feelings [22–24]. Moreover, many studies have tried to capture the acoustic cues that musicians use to convey specific emotions (e.g., [19,20,25,26]). These cues involve changes in tempo, sound level, articulation, timbre, timing, tone attack and decay, intonation, vibrato extent and frequency, accents on particular notes, etc. On the listener's side, judgments of intended emotions have been related to specific musical features, including tempo, articulation, intensity, and timbre [17,27–29]. In comparison, the same observations can be made for vocal expression and emotional prosody, as specific acoustic cues are predominant in accurately recognizing the intended emotions [30].

While auditory information plays a crucial role in music communication models, Finnäs (2001) noticed an increasing interest in the visual component influencing the perception of the musical performance [31]. While the auditory stream convey emotions in music, its associated movements also contain significant information. Many musical traditions have included a combination of both audio and visual stimulations during the experience of music performance [32,33]. Such practice still remains in our mediatized society [34]. The visual component of the live music performance contributes significantly to the appreciation of music performance [35,36]. Malin (2008) even concludes that a "variety of musical properties and types of evaluations can be affected by the visual information" [37]. Although both auditory and visual kinematic cues contribute significantly to the perception of overall expressiveness, the effect of visual kinematic cues appears to be somewhat stronger [38]. It also

provides preliminary evidence of cross-modal interactions in the perception of auditory and visual expressiveness in music performance. The visual component should not be categorized as a marginal phenomenon in music perception, but as an important factor in the communication of meaning. This process of cross-modal integration exists for many genre of music, from classical to pop and rock music [33,36]. All in all, visual kinematic cues have been found to influence the perception of phrasing and musical tension [33], felt emotion [39] the perception of emotional expression [40], and the overall appreciation of the performance (for a meta-analysis, see [38]).

　　A crucial visual cue is related to musicians' gestures, and how their bodies move during performances. Two types of movements can be here distinguished: instrumental actions and ancillary/expressive movements [41,42]. The former are creating sound while the latter have an intrinsic relationship with the music, representing a link between the music and the expressive intention of the musician [43]. Musical gestures are mainly made to produce sounds but are also used by the musician as means to convey or express emotions (see review Expressive Gesture [44,45]). Musicians' expressive gestures fall into two categories: (i) communicating their expressive intentions; and (ii) expressing their feelings without intending to communicate them [46]. Gestures contribute to communicate information to the audience as well as the other musicians. Expressive movements occur frequently in musical performances, even though these movements are not mandatory for musical performance such as during training [47]. Furthermore, across performers, these idiosyncratic expressive movements appear to have some consistencies [42,48]. Finally, Vines, Krumhansl, Wanderley, and Levitin (2006) concluded that these movements are not randomly performed, but rather are used to communicate a holistic, musical, expressive unit [33]. Understanding how this unit works is the primary goal of researchers interested in musical gestures.

1.2.1. Emotional Intensity and Expressive Intents in Musical Performances

　　With the recognition of emotions in music, the emotional intensity and expressive intents have been shown to be dependent on auditory cues. In their study on such information, researchers asked participants to rate the emotional expressiveness of music performances in which timing and intensity were parametrically manipulated [49]. Emotion judgments monotonically increased with performance variability, and timing changes were reported to explain more variance in emotional expressiveness than sound intensity. Changes in tempo and sound intensity in a music performance were also shown to be correlated with one another, and with real-time ratings of emotional arousal [50]. A systematic relationship between emotionality ratings, timing, and loudness was highlighted when listeners rated their moment-to-moment level of perceived emotionality while listening to music performances. Therefore, the variation of acoustic features associated with the expressive intent of the musician during a performance appears to have a crucial impact on the emotionality perceived. On the visual side, Davidson (2005) demonstrated that certain perceptual elements of a musician's gestures are sufficient for the audience to identify a musician's expressive intent [43]. She suggested the use of three level of expressiveness to be able to study the link between expressive gestures and musical performance: (1) without expression, labeled as "deadpan"; (2) with normal expression/concert-like, labeled as "projected"; and (3) with exaggerated expression, labeled as "exaggerated". Some body parts have been reported to convey more expressive information, specifically head, shoulders, arms, and torso [46,48,51–53]. Some motion features have also been associated with expressive motion such as the quantity of motion [54]. The use of motion features helps understand broad, unrefined body reaction and gives a first glimpse of the behavioral components of expressive gestures. It might help understand how musicians cope with the audience [28,55,56]. For example, musician facing an audience and playing in exaggerated expressive manners could be affected by the amount of supplementary stress caused by the difficulty of the task. All in all, as mentioned by Shaffer (1992) [57], "a performer can be faithful to its structure and at the same time have the freedom to shape its moods" (p. 265). This corresponds to a phenomenon called *performance expression*. It refers to "the small and

large variations in timing, dynamics, timbre, and pitch that form the micro-structure of a performance and differentiate it from another performance of the same music" ([58], p. 118).

1.2.2. Authenticity in Musical Performances

The importance of authenticity is undervalued in emotion research and musical performances. Authenticity could be an underlying factor of emotion communication through music. For example, in popular music culture, audio-visual performances convey markers of authenticity, which are essential for the creation of credibility and emotions [34]. In popular music, as the saying goes, "seeing is believing" ([34], p. 85). In everyday life, the anthropologist Erving Goffman, one of the great pioneers of social science research on emotions, affirmed that "We all play emotion theater most of the time". Goffman (1982) demonstrated that human beings mostly try to present themselves in the best light and always stage their daily lives to protect themselves [59]. Faked emotions or the modulation of the expression of emotions play a central role in self-preservation by keeping inappropriate emotional expressions to damage self-presentation. Scherer et al. (2013) argued that one should abandon the idea that, for the sake of complete authenticity, actors should live through "real emotions" on the stage [60]. Specific emotional expressions are only credible, i.e., appear authentic, when they can be perceived as generated by appraisals that fit the respective circumstances. This means that in order to succeed in appearing credible, the artist must: (1) pick the most appropriate set of vocal, facial and body expression elements for the respective emotions and combine them dynamically, in a psycho-biologically valid fashion; (2) achieve precise synchronization of the respective processes, letting the expression unfold in an appropriate fashion; and (3) handle the situational development appropriately in terms of its dynamic flow [60]. These requirements demand the highest amount of professionalism when trying to voluntary display a certain dynamic forms of expression. In music performance, one could also argue that self-awareness is a key feature in the perception of authentic emotion. Musicians exhibit this awareness at different times with both their technique and their emotional expressiveness. For example, after sight-reading a new music for some time, as the musician builds up the motor repertoire required, he or she can focus more on putting more expressive intent into his/her movements. Once these abilities become automatized, i.e., habits for a specific performance, they are no longer at the forefront of the individual's consciousness, the musician will then begin to bring components of their own personal performance style to the music. This, in turn, contributes to the perceived authenticity of the ultimate performance.

Even though the view of what is an authentic musical performance is subjective and based on individual bias, listeners tend to agree that authentic musical performance styles all have a sense of uniqueness. For example, celebrated pianist Glenn Gould is often noted as an exemplary expressive musician with an extremely particular performance style. This individuality is one of the hallmarks of authenticity in performance. Wöllner (2013) suggests that individual artistic expression can be quantified as such when the performance matches the listener's "mental prototype" of what a unique and authentic performance would look and sound like [61]. Overall, it is important to note that, as implied in the BRECVEMA model regarding "appreciation emotions" in aesthetic judgment [62], while the "mental prototype" we all use when making judgments about the authenticity of a performance are socially and culturally driven, the gestures that characterize "authenticity" in music are extremely useful in analyzing how skilled musicians play "emotion theater" to create moving and expressively credible performances.

1.3. Goal of This Study

This study aims to investigate the impact of the audience presence on both aspects of a music performance, from both performers and observers' views. By studying the difference in acoustic and motion features at different levels of expressive intent, we want to demonstrate the impact of the audience presence on the link between the expressive intents performed on the musician's side and the emotionality perceived on the observer's side. We hypothesized that the presence of a virtual audience

would hinder the movements of the musicians due to the stress generated by the act of performing live. More specifically, it would reduce the differences in acoustic and motion features between the different expressive conditions. Consequently, this would also impact the emotional intensity and authenticity perceived by the audience. The participants should therefore report similar values across expressive conditions.

To understand such complex phenomenon, we recorded musicians playing with different expressive manners in front of a virtual crowd or an empty room. We analyzed motion and acoustic features and measured the impact of our social factor, i.e., the audience. Afterwards, we presented video clips of musicians playing and asked participants to rate both the emotional intensity and authenticity perceived. We performed a series of analysis to link these values to the audio and visual cues explored.

2. Materials and Methods

This experiment was divided into two phases: the recording sessions and the rating experiment. Both phases were approved by the local ethical committee of the department of Psychology, University of Geneva. These two phases aim to emphasize, respectively, on the proximal and distal cues of a Brunswik lens model [63]. This type of model has been shown to be highly representative in the case of emotional prosody [64] and in music [27].

2.1. Recording Session

Four violinists (3 females, $M_{age} = 22$) took part in the recordings. They were paid according to the ethical protocol. They agreed with the use of the material recorded as stimuli for this study. Musicians performed inside an Immersive Virtual Environment (IVE) with the use of a system of three screens, seven TITAN QUAD 3D projectors (Digital Projection Limited, Manchester, UK), and stereo glasses presenting seamless and perceptively coherent 3D images. Two different virtual environments were created for this experiment: a room filled with an audience behaving naturally and attending the concert, and the same room without the audience (Appendix Figure A1). The virtual audience was composed of high quality agents with realistic facial expressions and behaviors. An agent is created using four different components: (1) realistic body from a 3D scan of real actors; (2) realistic body animations created from motion capture footage; (3) accurate and controllable facial expressions based on FACSGen [65,66]; and (4) expressive behavior and social interaction modeling. The audience behavior could smoothly change from engaged to disengaged behaviors. Specifically, in the engaged behavior, the audience attention increases through the convergence of individual gazes focusing on the musician [67]. On the other hand, a disengaged audience is rendered by allowing the gaze of avatar to wander around as generally observed in distracted people. Furthermore, for realism purposes, each avatar had random idle animations of their body while looking at the musician in order to approximate usually seen fluent behaviors at concerts. To avoid unnatural uniformity, part of the audience (5–10%) is always modeled in the disengaged condition when the majority is engaged, and vice versa [68].

Each musician was instructed to play 30-second-long interpretations of Bach's Partita No. 2 in D minor, BWV 1004: Sarabande. The part of the musical score to be played was carefully selected to correspond to complete musical phrases. The excerpts were interpreted according to three selected expressive intents: deadpan, projected, and exaggerated [43]. The different combinations of conditions were performed in a pseudo-randomized order by each musician. Six excerpts were recorded for each of the 4 musicians, adding up to a total of 24 excerpts. The sound was captured using an Olympus LS-10 (Olympus, Tokyo, Japan). Motion capture data was also recorded for every piece using a VICON optical motion tracking system (Vicon Motion Systems Ltd., Oxford, UK) composed of eight Bonita 3 cameras (Vicon Motion Systems Ltd., Oxford, UK). A total of 26 markers were used, covering selected body parts based on recent literature on the analysis of music performance, i.e., the head, arms, and torso [48,51–53].

2.2. Rating Experiment

Forty participants took part in the rating study (19 females). All of them spoke French as first language. The average age was 23 years ($SD = 7.19$) and most participants were psychology students. Participants completed the experiment on a computer. The experiment itself was programmed with Limesurvey [69] and ran on computers with a screen resolution of 1280×1024 pixels. Loudness was set to 50% and could be adjusted by the participant. Headphones were provided. The experiment lasted ~30 min. The participants had to complete a musical habits questionnaire before starting the experiment. Stimuli were fully mixed together and presented in a unique random order for each participant. The participant listened to each stimulus while watching the point light display (PLD) of the musician's movements and then answered multiple questions. They rated the emotional intensity of the stimuli, the authenticity, each of the 9 emotions from the Geneva Emotional Music Scales (GEMS) (Appendix Figure A3) [24]. All ratings were done on sliders from 0 to 100. The participants were also asked to rate the importance of each body part in evaluating the general emotional intensity. This process was repeated for the 24 stimuli per participant.

2.3. Multi-Modal Expressive Behavior Analysis

Drawing upon the recent studies [70,71], we considered two types of expressive body features: the kinetic energy and the Body Twist Index (BTI). The former helps understand broad, unrefined body reactions and gives a glimpse of the behavioral components of expressive gestures. The latter captures body shape related information, i.e., the relative displacement of body parts with respect to other ones. In the case of violin player, this second feature is critical since the upper and lower parts of the body are dissociated. Violinists tend to twist their body more while playing compared to cello players for example.

Listening to music does not involved watching or performing gesture in around 80% of the time [72]. We therefore computed acoustic features and focused essentially on the metrical centroid [73]. This feature offers a very detailed description of the metrical structure of a musical piece. Time-related aspect of music is thought to have an impact on the emotional arousal and linked to the notion of musical entrainment [62]. It is also the preferred acoustic cue used by listener to perceived different emotions in a piece of music [27]. The metrical centroid is expressed in beat per minutes (BPM). Low BPM values indicate a prevalence of high metrical level (i.e., slow pulsations corresponding to whole notes, bars, etc.). High BPM values indicate on the contrary that more elementary metrical levels predominate (i.e., very fast levels corresponding to very fast rhythmical values). We hypothesized that these three features could help modeling critical changes in musician's expressive responses to the presence of an audience.

Both motion features were calculated using authors' MATLAB (v2016b, MathWorks, Inc., Natick, MA, USA, 2016) toolbox (build upon the MoCap Toolbox [74]). The kinetic energy was computed for every marker of the motion capture data and then averaged across markers and over time. The Body Twist Index was represented by the average angle between the pelvis and a perpendicular line to the shoulders, considering only the top quantile (above quantile 75%) of the data for each excerpt. Both features were z-scored per musician (Figure 1). The acoustic feature was computed with the MIRToolbox 1.6.1 [75]. It was computed on overlapping frames of the musical excerpts (duration: 1 s, hop: 0.25 s). The average value of the dynamic acoustic feature was also z-scored per musician (Figure 1).

Linear models were used in this study for modeling the performance of the musicians while linear mixed models were used to estimate the participants' ratings. Mixed models offer two advantages: they incorporate random effects and they allow handling correlated data and unequal variance [76]. When using features as fixed effect in the modeling of participants' perception of authenticity and emotional intensity, we divided them into four bins ("0–25","25–50","50–75", and "75–100"). This allowed contrasts to be computed between bins. Comparing model was done using Chi-squared testing. All *p*-values were corrected using False Discovery Rate (FDR).
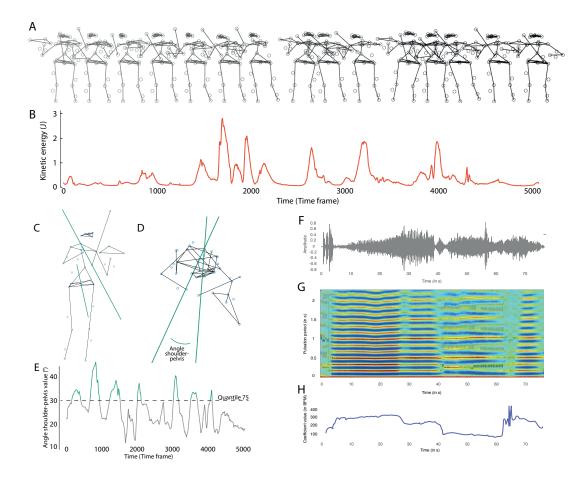
Figure 1. Motion and acoustic features computed: (**A**) Chronograph of the motion capture data for one excerpt played in front of an audience while exaggerating expressive intent. (**B**) Kinetic energy associated with movements of all the markers for the excerpt depicted in (**A**). (**C**) 3D view of the motion capture and the line associated with the pelvis and shoulder. (**D**) Transversal view of the motion capture data and the angle computed between the line from the pelvis and shoulders. (**E**) Computation of Body Twist Index. The angle between the aforementioned lines is computed over the duration of the excerpt. The Body Twist Index consists of the average of all values comprised in the top quantile (above the quantile 75 value). (**F**) Sound profile of the performance of the Scherzo of L. van Beethoven's Symphony No.9 in D minor, op.125. (**G**) Corresponding autocorrelogram with tracking of the metrical structure. (**H**) Corresponding metrical centroid curve (Copyright Grandjean, D. et al., 2013 [73]).

3. Results

In this section, we present the analysis of both the proximal and distal components of the performance, i.e., from both performers' and observers' side.

3.1. Proximal Performance Data Analysis

In order to characterize the musical gestures and motion, we computed the z-scored values of all features (kinetic energy, body twist index, and metrical centroid) recorded during the performance at different expressive intents with the presence or not of a virtual audience. Based on the marginality principle, we considered only a linear model containing main effects of both the expressive intents and the social factor as well as the interaction between these factors ($F_{kinetic energy}(5, 18) = 39.78, p = 4.1 \times 10^{-9}$, $R^2_{adjusted} = 0.89$ & $F_{BTI}(5, 18) = 6.77, p = 0.001, R^2_{adjusted} = 0.55, F_{metrical centroid}(5, 17) = 16.58$, $p = 5.17 \times 10^{-6}, R^2_{adjusted} = 0.78$).

Both motion features increased with stronger expressive intents in the absence of the social factor (Figure 2). When the audience was present, while the kinetic energy still increased with the expressive intent, the difference with the other social condition was not significant. When comparing the presence and absence of the audience, the impact of the social factor only appeared for the deadpan expressive condition for the BTI. In this case, the presence of the audience was characterized by a significant increase in the "twist" angle ($F_{BTI,DP,Social}(1,18) = 5.41$, $p = 0.032$). Noteworthy was the significant difference for the BTI in the deadpan and exaggerated condition when the audience was present or not. The BTI value increased in the deadpan condition with the presence of an audience while it was diminished in the exaggerated condition ($F_{BTI,DP/EXAG,Social}(1,18) = 5.64$, $p = 0.028$).
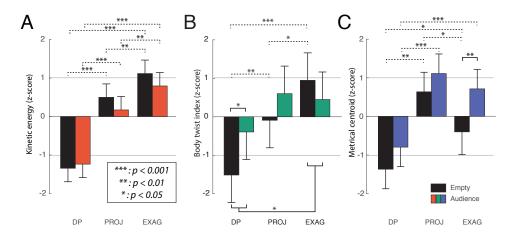

Figure 2. Impact of the interaction of the expressiveness and the presence of an audience on body features (deadpan: DP, projected: PROJ, and exaggerated: EXAG): (**A**) Kinetic energy (red); (**B**) Body Twist Index (green); and (**C**) Metrical centroid (blue). (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$).

In the case of the acoustic feature, metrical centroid, we observed a different pattern when the virtual audience was absent (Figure 2). When asked to play with a "projected" expressive intent, the metrical centroid of the performance was significantly greater, meaning that the more elementary metrical levels predominated (i.e., very fast levels corresponding to very fast rhythmical values). It was however lower for both the "deadpan" and "exaggerated" conditions, contrarily to the linear increase observed in motion features. The effect of the social factor was highlighted in the increase of the beats per minute of the metrical centroid for every expressive intent, especially for the "exaggerated" condition where this increase was significant ($F_{MetCent,EXAG,Social}(1,17) = 9.24$, $p = 0.007$).

3.2. Distal Participant Data Analysis

The second part of our analyses focused on the ratings given by the participants on both authenticity and emotional intensity. Across all stimuli, the reliability of our participants' ratings was high, $\alpha = 0.92$ for authenticity and $\alpha = 0.93$ for intensity. Participants were grouped together into two categories, music-lovers and musicians, based on their responses for the musical habits questionnaires [77]. No significant difference in means was observed between groups for both authenticity ($M_{musicians} = 43.155$, $SD_{musicians} = 25.28$, $M_{music-lovers} = 46.217$, $SD_{music-lovers} = 23.98$, $t(819) = -1.829$, $p = 0.067$, $d = -0.12$) and intensity ($M_{musicians} = 42.308$, $SD_{musicians} = 24.598$, $M_{music-lovers} = 45.314$, $SD_{music-lovers} = 22.059$, $t(819) = -1.89$, $p = 0.059$, $d = -0.13$). The difference between both groups is marginal. Moreover, this difference is trivial due to the very small effect size associated with the p-value being influenced by the large number of trials (see Cohen's guideline [78]). We thus concluded that an analysis could be performed on the dataset as a whole. Responses for authenticity and emotional intensity were also highly correlated ($r = 0.727$). This is noticeable with the relatively similar outcomes of the evaluated models. This next section focuses on separate model

estimations for both authenticity and emotional intensity. Two different models were implemented. First, we estimated both dependent variables using both the expressive intent and the social factor. Second, we explored the impact of the variation of both motion and acoustic features on the perceived emotional intensity and authenticity.

3.2.1. Interaction Effect of the Expressive Intents and the Presence of an Audience on Perceived Authenticity and Emotional Intensity

Two models were computed to estimate the influence of the expressive intents and the presence of an audience, respectively, on the perceived authenticity and the emotional intensity. The first model estimated the perceived emotional intensity using the different categories of expressive intent and the presence of an audience, as well as the interaction, as fixed effects, and with the participants and the musician at play as random effects (a model computing only the main effect and not the interaction is not presented in this article based on the principle of marginality; however, graphs related to such model can be found in the Supplementary Materials (Appendix Figure A2)). This model was significantly better than a model using only the main fixed effects, no interaction, and the same random effects (intensity: $\chi^2(3, N_{trials} = 875) = 12.036$, $p = 0.01$, $AIC = 7640.6$, $BIC = 7688.4$, $R_m^2 = 0.08$, $R_c^2 = 0.42$). The second model estimated the perceived authenticity using the same fixed and random effects. This model was significantly better than a model using only the main fixed effects, no interaction, and the same random effects (authenticity: $\chi^2(3, N_{trials} = 875) = 18.026$, $p = 8.68 \times 10^{-4}$, $AIC = 7738.7$, $BIC = 7786.4$, $R_m^2 = 0.11$, $R_c^2 = 0.42$).

Both perceived authenticity and emotional intensity increased significantly with every increment of the musicians' expressiveness in the case of the absence of audience (intensity: $\chi^2_{EmoInt,DP/Proj,Absence}(1, N_{trials} = 875) = 22.909$, $p = 1.69 \times 10^{-6}$, $\chi^2_{EmoInt,Proj/Exag,Absence}(1, N_{trials} = 875) = 13.765$, $p = 0.0002$, and authenticity: $\chi^2_{Auth,DP/Proj,Absence}(1, N_{trials} = 875) = 33.624$, $p = 6.6 \times 10^{-9}$, $\chi^2_{Auth,Proj/Exag,Absence}(1, N_{trials} = 875) = 11.009$, $p = 0.0009$) (Figure 3). When the audience was present, the ratings associated with such stimuli were only significantly different between the deadpan and projected condition for both dependent variables (intensity: $\chi^2_{EmoInt,DP/Proj,Absence}(1, N_{trials} = 875) = 40.608$, $p = 1.8e^{-10}$, and authenticity: $\chi^2_{Auth,DP/Proj,Presence}(1, N_{trials} = 875) = 74.007$, $p < 2.2 \times 10^{-16}$). The comparison between the presence and the absence of a virtual audience highlighted significantly different results only for the projected condition ($\chi^2_{EmoInt,Proj,Social}(1, N_{trials} = 875) = 9.327$, $p = 0.002$ and ($\chi^2_{Auth,Proj,Social}(1, N_{trials} = 875) = 15.896$, $p = 6.69 \times 10^{-5}$).
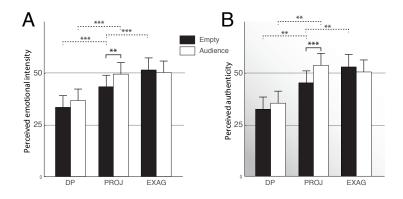

Figure 3. Interaction of the expressiveness and the presence of an audience on: (**A**) the perceived emotional intensity; and (**B**) the perceived authenticity (deadpan: DP, projected: PROJ, and exaggerated: EXAG) (** $p < 0.01$, *** $p < 0.001$).

3.2.2. Effect of the Motion and Acoustic Features and the Presence of an Audience

Perceived authenticity and emotional intensity were also modeled using the motion and acoustic features computed on the material recorded. Three models were computed using, respectively, one of the three features, energy kinetic, body twist index, and metrical centroid. The interaction of the features, used here as continuous predictors, and the fixed effect representing the presence of an audience were used in these models. The participants and the musicians were used as random effects. When comparing such model with models with no interaction, only the models for kinetic energy and metrical centroid were significantly improving the model accuracy for both emotional intensity (kinetic energy: $\chi^2(4, N_{trials} = 875) = 13.572$, $p = 0.011$, $AIC = 7649.9$, $BIC = 7707.2$, $R^2_m = 0.8$, $R^2_c = 0.42$, metrical centroid: $\chi^2(4, N_{trials} = 875) = 32.133$, $p = 1.8 \times 10^{-6}$, $AIC = 7653.7$, $BIC = 7711$, $R^2_m = 0.08$, $R^2_c = 0.43$) and authenticity (kinetic energy: $\chi^2(4, N_{trials} = 875) = 12.339$, $p = 0.018$, $AIC = 7750.3$, $BIC = 7817.6$, $R^2_m = 0.105$, $R^2_c = 0.42$, metrical centroid: $\chi^2(4, N_{trials} = 875) = 35.714$, $p = 3.31 \times 10^{-7}$, $AIC = 7760.5$, $BIC = 7817.8$, $R^2_m = 0.10$, $R^2_c = 0.42$).

When the musician was playing in front of an empty room, the recorded material was rated as more emotionally intense and authentic as the kinetic energy was increasing. The stimuli associated with higher value for kinetic energy were significantly rated higher for both dependent variables (intensity: $\chi^2_{EmoInt,Low/MidKinEn,Absence}(1, N_{trials} = 875) = 4.368$, $p = 0.036$, $\chi^2_{EmoInt,Mid/HighKinEn,Absence}(1, N_{trials} = 875) = 4.948$, $p = 0.026$; and authenticity: $\chi^2_{Auth,Low/MidKinEn,Absence}(1, N_{trials} = 875) = 6.66$, $p = 0.009$, $\chi^2_{Auth,Mid/HighKinEn,Absence}(1, N_{trials} = 875) = 4.47$, $p = 0.034$) In the case of metrical centroid, mid-range values were rated as more intense and authentic compared to extreme values (Figure 4). The rating associated with lower values were significantly different from the high values, while the high values were significantly (and marginally in the case of authenticity) different from the mid values (intensity: $\chi^2_{EmoInt,Low/HighMetCent,Absence}(1, N_{trials} = 875) = 16.473$, $p = 4.9 \times 10^{-5}$, $\chi^2_{EmoInt,Mid/HighMetCent,Absence}(1, N_{trials} = 875) = 5.288$, $p = 0.021$; and authenticity: $\chi^2_{Auth,Low/HighMetCent,Absence}(1, N_{trials} = 875) = 21.277$, $p = 3.9 \times 10^{-6}$, $\chi^2_{Auth,Mid/HighMetCent,Absence}(1, N_{trials} = 875) = 3.773$, $p = 0.052$).

The introduction of an audience in front of the musicians influenced how the musicians performed (Figure 2) but also brought significant changes in how intense and authentic the performance was perceived. Those changes highlighted the emergence of a bipartite distribution of our data instead of the tripartite grouping of the expressive intents (Figure 4). For both the authenticity and the emotional intensity, the two levels of the bipartite distribution were significantly different from each another (intensity: $\chi^2_{EmoInt,KinEn,Low/HighKinEn,Audience}(1, N_{trials} = 875) = 59.625$, $p = 1.14 \times 10^{-14}$, $\chi^2_{EmoInt,Low/HighMetCent,Audience}(1, N_{trials} = 875) = 36.056$, $p = 1.91 \times 10^{-9}$, authenticity: $\chi^2_{Auth,Low/HighFeatKinEn,Audience}(1, N_{trials} = 875) = 81.819$, $p < 2.2 \times 10^{-16}$, $\chi^2_{Auth,Low/HighMetCent,Audience}(1, N_{trials} = 875) = 48.918$, $p = 2.6 \times 10^{-12}$). The values of the features at which the separation occurred was calculated. It corresponds at 6.73×10^{-3} W for the kinetic energy ($kineticenergy_{min} = 2.207 \times 10^{-3}$ W and $kineticenergy_{max} = 0.0304$ W) and 308.18 BPM for the metrical centroid ($metricalcentroid_{min} = 260.48$ BPM and $metricalcentroid_{max} = 335.57$ BPM).
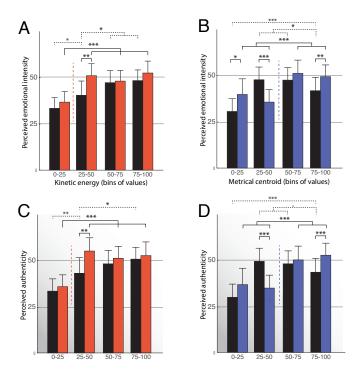
Figure 4. Impact of the interaction of the computed features: (**A**,**C**) kinetic energy (red); and (**B**,**D**) metrical centroid (blue); and the presence of an audience on: (**A**,**B**) the perceived emotional intensity; and (**C**,**D**) the perceived authenticity (deadpan: DP, projected: PROJ, and exaggerated: EXAG) ($p < 0.01$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$).

4. Discussion

In this study, we highlighted the changes in motion and acoustic features associated with different expressive manners. We measured the impact of the presence of virtual audience on those features. We also modeled the perceived emotional intensity and authenticity associated with different expressive manners and feature values. We provided scientific evidence about how emotional and authentic musical performances could be perceived.

4.1. Absence of an audience

We observed that different expressive intents were characterized by different quantities of body movements, as already demonstrated in the literature, in the absence of an audience [43,54]. Specifically, this study explored two features, the kinetic energy and newly developed body twist index, highlighting that energetic movements and wider body twists were associated with the magnitude of the expressive intent. Both features were increasing with the different expressive intents. Generally, it was revealed that musicians tended to make more movements when playing expressively [54]. This absence of the audience allowed them to fully twist and put more energy into their gestures. When analyzing the sound produced, the metrical centroid representing the metrical structure of the piece was also impacted by the different expressive manners. However, its value did not linearly increase between deadpan, projected and exaggerated expressive intents. The projected expressive intent was characterized by the highest value for the metrical centroid feature. This highlighted the predominance of faster notes within the musical piece. The other two expressive intents could be symbolized by longer notes or slow pulsations. In the case of the deadpan style, this suggested a more controlled way of playing, while emphasizing on a regular and slower beat. When musicians were exaggerating, the excessive expressive intent was marked on both slower and faster pulsations at different timing driving the decrease in the centroid value. As musicians dedicated more attention to

their expressive intent, they change the way they played a certain piece. This confirms the phenomenon of performance expression [57,58].

When studying participants' perception of emotional intensity and authenticity, both ratings were affected significantly by the different expressive intents in the absence of an audience. Ratings were significantly increasing with the expressive conditions showing greater emotional intensity and authenticity in the exaggerated condition. The study of the impact of the physical attributes of the performance on the perception of emotional intensity and authenticity was conducted using feature values instead of the well-documented expressive conditions. The perception associated to stimuli with no virtual audience displayed was in line with the feature values obtained for different expressive intent. The deadpan (low kinetic energy and low metrical centroid), projected (medium kinetic energy and high metrical centroid), and exaggerated (high kinetic energy and medium metrical centroid) conditions are perceived, respectively, as relatively low, medium and high emotional intensity and authenticity. They showed a tripartite distribution of values which would fit with the three expressive conditions [43].

Before addressing the impact of the social factor on participants' ratings, we also noted the high correlation between authenticity and emotional intensity. Two conclusions can be drawn from such results. Firstly, the difficulty to discriminate between both ratings might be due to an underlying link between them. In music, authenticity represents an important part of the performance to make the listener feel the desired emotions [60]. Secondly, such link should be further untangled with a different experimental setup. We propose the use of "fake" stimuli, where musicians would fake the emotion felt and performed. This could for example be conducted by using mood induction procedures for "real" emotion stimuli while asking musicians to fake the rest of the stimuli. The efficiency of mood induction procedures have been already proven, especially for negative emotions [79].

4.2. Impact of the Presence of a Virtual Audience

When considering the impact of the presence of an audience on the body features, the data showed an interaction effect between the social factor and the expressive conditions on the features values. The only non-significant model is associated to kinetic energy. This expressive cue is well-known and extensively used by musicians to impact their communication of expressive intents. The presence of an audience does not significantly impact the well-regulated quantity of movement during the performance. The use of two other features, body twist index and metrical centroid, brings complementary information on the modification of such expressive performances when playing in front of an audience. In both features, the differences between the three expressive conditions tend to fade away with the presence of an audience. Values for each features converge towards ones for the projected expressive intent. This phenomenon could be linked to habits [80]. Musicians that are used to play in front of an audience could be tuning their movements in a certain way to make them comfortable yet expressive. In this study, the presence of an audience seems to push musicians to use this set of usual movements. When playing with no expressive intent, the musician could feel the need to still express some emotions for the audience to enjoy the performance more. Deadpan expressive intents are usually performed in a controlled environment, e.g., at home, alone while rehearsing, when the musician's locus on control is self-oriented. Internals' performance and stress are proven to be better controlled, for instance in a work environment [81]. The presence of an audience produces considerable effect, shifting the locus of control, disrupting the habits associated with this expressive condition and putting musicians in a more complex, stressful and less controlled situation. To counterbalance this effect, musicians tend to reach back to a controlled and habitual situation mimicking projected expressive movements. Similarly, exaggerating expressive intents in front of a crowd could put musicians in an uncomfortable position where their play seems a bit less authentic. To counterbalance this effect, musicians will then naturally try to be less expressive and go back to a projected expressive condition. Consequently, our first hypothesis is therefore validated showing that gestures become less differentiable across conditions.

Similar effects could be observed on the emotional intensity and authenticity. Both deadpan and exaggerated stimuli produce responses similar to the projected conditions when an audience is present during the musicians' performances. As the values of the features tend to converge towards the likes of the projected expressive intent, perception converges towards a uniform value represented by the projected scenario. This phenomenon is most likely due to the unease felt by the musician when playing with an unfamiliar expressive style in a habitual situation. The difference in perceptual intensity and authenticity is the strongest in the case of musicians playing as they would in a concert and facing a virtual audience. Contrarily to deadpan and exaggerated conditions, the ease coming with the congruence of the common context—playing in a concert-like fashion and facing an audience—allows musicians to appear more authentic and communicate their expressive intent better in the process.

When playing in front of an audience, the previously observed tripartite distribution of ratings disappears. For both emotional intensity and authenticity, the distribution based on the both kinetic energy and metrical centroid becomes dichotomist. The emergence of such threshold between perception of high and low emotional intensity/authenticity is a repercussion of the convergence of exaggerated performances toward a concert-like situation. In front of an audience, both projected and exaggerated context are evaluated as highly authentic and emotionally intense. The second hypothesis stating that different expressive conditions would be rated similarly for emotional intensity and authenticity in the social condition is therefore verified. Furthermore, the calculated threshold could consequently be used for automatic detection of emotional intensity.

Both models using the expressive conditions and the features values are converging towards one conclusion: the use of the methodological framework designed by Davidson, 2005, is here showing its limits in its ecological validity [43]. While the three expressive intents can be communicated accurately by the musicians to the listeners, the presence of an audience, even a virtual one, during the recordings reduces the ability to differentiate between such categories of expressive intents. The distinction between the projected and exaggerated conditions is blurred and we suggest to take into account such modifications when recording musicians in front of a public, virtual or real. This audience effect on the expressive intents might here be due to an emergence of non-explicit and involuntary regulation processes in social context, driving a strong impact of such conditions. Such processes should be further explored.

4.3. Interactive Virtual Environments as a Tool to Study Musical Performances

Our findings supports the use of tools such as IVEs in music research . Previous studies in the domain of social phobias already recommended the use of IVE for coping with stress generated by the fear of public speaking [10,11]. Similarly, IVEs could become crucial tools to train musicians to cope with stress related to live performances and aid learning [2]. As shown in this study, the presence of an audience impacts both the movements and sounds related to the performance and, consequently, the perception of listeners. The system developed in this study could also be easily adapted to context-sensitivity in real-time and could provide feedbacks to the musicians, e.g., when their movements are radically diverging from a previous recorded performance. It could help musicians understand the impact of stress on their performance allowing them to develop coping mechanisms for musical performance anxiety. Finally, research on communication of emotions, alongside with this study and the IVE, might be used by music teachers to enhance performers' expressiveness [82,83].

5. Conclusions

To conclude, the presence of an audience generated important variations in both acoustic and motion features related to music performance. This influence is to be taken into account when approaching music research during concerts. Immersive Virtual Environments could therefore be utilized both for research and as a tool for training musicians to cope with audience anxiety.

Appendix A

Figure A1. Interactive Virtual Environment: (**A**) Example views of both social conditions (left, empty; right, audience); (**B**) details of a disengaged audience (deadpan: DP, projected: PROJ, and exaggerated: EXAG).
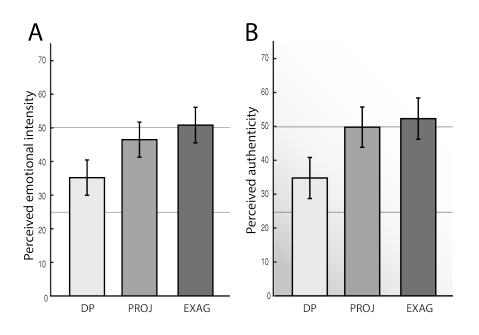
Appendix B

Figure A2. Impact of the expressiveness on (**A**) the perceived emotional intensity; (**B**) the perceived authenticity (deadpan: DP, projected: PROJ, and exaggerated: EXAG).
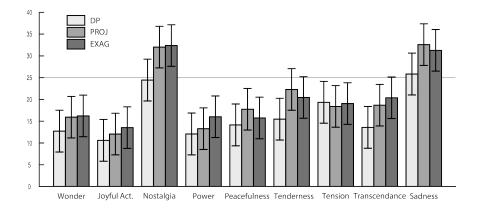
Appendix C

Figure A3. Impact of the interaction between the Geneva emotional scale and the expressiveness (deadpan: DP, projected: PROJ, and exaggerated: EXAG).

References

1. Wilson, G.D.; Roland, D. Performance anxiety. In *The Science and Psychology of Music Performance: Creative Strategies for Teaching and Learning*; Oxford University Press: Oxford, UK, 2002; pp. 47–61.
2. Williamon, A.; Aufegger, L.; Eiholzer, H. Simulating and stimulating performance: Introducing distributed simulation to enhance musical learning and performance. *Front. Psychol.* **2014**, *5*, 1–9.
3. Blascovich, J.; Loomis, J.; Beall, A.C.; Swinth, K.R.; Hoyt, C.L.; Bailenson, N.; Bailenson, J.N. Immersive virtual environment technology as a methodological tool for social psychology. *Psychol. Inq.* **2002**, *13*, 103–124.
4. Sanchez-Vives, M.V.; Slater, M. From presence to consciousness through virtual reality. *Nat. Rev. Neurosci.* **2005**, *6*, 332–339.
5. Difede, J.; Cukor, J.; Hoffman, H.G. Virtual Reality Exposure Therapy for the Treatment of Posttraumatic Stress Disorder Following September 11, 2001. *J. Clin. Psychiatry* **2007**, *68*, 1639–1647.
6. Rizzo, A.; Reger, G.; Gahm, G.; Difede, J.; Rothbaum, B.O. Virtual reality exposure therapy for combat-related PTSD. In *Post-Traumatic Stress Disorder*; Springer: New York, NY, USA, 2009; pp. 375–399.
7. Rizzo, A.S.; Buckwalter, J.G.; Forbell, E.; Reist, C.; Difede, J.; Rothbaum, B.O.; Lange, B.; Koenig, S.; Talbot, T. Virtual Reality Applications to Address the Wounds of War. *Psychiatr. Ann.* **2013**, *43*, 123–138.
8. Rothbaum, B.O.; Hodges, L.; Smith, S.; Lee, J.H.; Price, L. A controlled study of virtual reality exposure therapy for the fear of flying. *J. Consult. Clin. Psychol.* **2000**, *68*, 1020–1026.
9. Bouchard, S.; Côté, S.; St-Jacques, J.; Robillard, G.E.; Renaud, P. Effectiveness of virtual reality exposure in the treatment of arachnophobia using 3D games. *Technol. Health Care* **2006**, *14*, 19–27.
10. Pertaub, D.; Slater, M.; Barker, C. An experiment on fear of public speaking in virtual reality. In *Studies in Health Technology and Informatics*; IOS Press: Amsterdam, The Netherlands, 2001; pp. 372–378.
11. North, M.M.; North, S.M.; Coble, J.R. Virtual reality therapy: An effective treatment for the fear of public speaking. *Int. J. Virtual Real. IJVR* **2015**, *3*, 1–6.
12. Orman, E.K. Effect of virtual reality graded exposure on anxiety levels of performing musicians: A case study. *J. Music Ther.* **2004**, *41*, 70–78.
13. Bissonnette, J.; Dubé, F.; Provencher, M.D.; Moreno Sala, M.T. Evolution of music performance anxiety and quality of performance during virtual reality exposure training. *Virtual Real.* **2016**, *20*, 71–81.
14. Juslin, P.N.; Laukka, P. Expression, Perception, and Induction of Musical Emotions: A Review and a Questionnaire Study of Everyday Listening. *J. New Music Res.* **2004**, *33*, 217–238.
15. Behrens, G.A.; Green, S.B. The ability to identify emotional content of solo improvisations performed vocally and on three different instruments. *Psychol. Music* **1993**, *21*, 20–33.
16. Gabrielsson, A. Expressive intention and performance. In *Music and the mind machine*; Springer: Berlin/Heidelberg, Germany, 1995; pp. 35–47.

17.　Gabrielsson, A.; Juslin, P.N. Emotional Expression in Music Performance: Between the Performer's Intention and the Listener's Experience. *Psychol. Music* **1996**, *24*, 68–91.

18.　Juslin, P.N. Emotional Communication in Music Performance: A Functionalist Perspective and Some Data. *Music Percept. Interdiscip. J.* **1997**, *14*, 383–418.

19.　Juslin, P.N. Perceived Emotional Expression in Synthesized Performances of a Short Melody: Capturing the Listener's Judgment Policy. *Music. Sci.* **1997**, *1*, 225–256.

20.　Juslin, P.N.; Madison, G. The Role of Timing Patterns in Recognition of Emotional Expression from Musical Performance. *Music Percept. Interdiscip. J.* **1999**, *17*, 197–221.

21.　Laukka, P.; Juslin, P.N. Improving emotional communication in music performance through cognitive feedback. *Music. Sci. J. Eur. Soc. Cognit. Sci. Music* **2000**, *4*, 151–183.

22.　Balkwill, L.l.; Thompson, W.F. A Cross-Cultural Investigation of the Perception of Emotion in Music: Psychophysical and Cultural Cues. *Music Percept. Interdiscip. J.* **1999**, *17*, 43–64.

23.　Fritz, T.; Jentschke, S.; Gosselin, N.; Sammler, D.; Peretz, I.; Turner, R.; Friederici, A.D.; Koelsch, S. Universal Recognition of Three Basic Emotions in Music. *Curr. Biol.* **2009**, *19*, 573–576.

24.　Zentner, M.; Grandjean, D.; Scherer, K.R. Emotions evoked by the sound of music: Characterization, classification, and measurement. *Emotion* **2008**, *8*, 494–521.

25.　Jansens, S.; Bloothooft, G.; de Krom, G. Perception And Acoustics Of Emotions In Singing. In Proceedings of the Fifth European Conference on Speech Communication and Technology, Rhodes, Greece, 22–25 September 1997.

26.　Mergl, R.; Piesbergen, C.; Tunner, W. *Musikalisch-Improvisatorischer Ausdruck und Erkennen von Gefühlsqualitäten*; Hogrefe: Göttingen, Germany, 2009; pp. 1–11.

27.　Juslin, P.N. Cue utilization in communication of emotion in music performance: relating performance to perception. *J. Exp. Psychol. Hum. Percept. Perform.* **2000**, *26*, 1797–1813.

28.　Juslin, P.N.; Laukka, P. Communication of emotions in vocal expression and music performance: Different channels, same code? *Psychol. Bull.* **2003**, *129*, 770–814.

29.　Juslin, P.N.; Sloboda, J.A. *Music and Emotion: Theory and Research.*; Oxford University Press: Oxford, UK, 2001.

30.　Banse, R.; Scherer, K.R. Acoustic profiles in vocal emotion expression. *J. Personal. Soc. Psychol.* **1996**, *70*, 614–636.

31.　Finnäs, L. Presenting music live, audio-visually or aurally—does it affect listeners' experiences differently? *Br. J. Music Educ.* **2001**, *18*, 55–78.

32.　Frith, S. *Performing Rites: On the Value of Popular Music*; Harvard University Press: Cambridge, MA, USA, 1998.

33.　Vines, B.W.; Krumhansl, C.L.; Wanderley, M.M.; Levitin, D.J. Cross-modal interactions in the perception of musical performance. *Cognition* **2006**, *101*, 80–113.

34.　Auslander, P. *Liveness: Performance in a Mediatized Culture*; Routledge: Abingdon, UK, 2008.

35.　Bergeron, V.; Lopes, D.M. Hearing and seeing musical expression. *Philos. Phenomenol. Res.* **2009**, *78*, 1–16.

36.　Cook, N. Beyond the notes. *Nature* **2008**, *453*, 1186–1187.

37.　Malin, Y. Metric Analysis and the Metaphor of Energy: A Way into Selected Songs by Wolf and Schoenberg. *Music Theory Spectr.* **2008**, *30*, 61–87.

38.　Platz, F.; Kopiez, R. When the eye listens: A meta-analysis of how audio-visual presentation enhances the appreciation of music performance. *Music Percept.* **2012**, *30*, 71–83.

39.　Chapados, C.; Levitin, D.J. Cross-modal interactions in the experience of musical performances: Physiological correlates. *Cognition* **2008**, *108*, 639–651.

40.　Vines, B.W.; Krumhansl, C.L.; Wanderley, M.M.; Dalca, I.M.; Levitin, D.J. Music to my eyes: Cross-modal interactions in the perception of emotions in musical performance. *Cognition* **2011**, *118*, 157–170.

41.　Cadoz, C.; Wanderley, M.M.; Cadoz, C.; Wanderley, M.M.; Music, G.; Wanderley, M. Gesture-Music. In *Trends Gestural Control Music*; IRCAM: Paris, France, 2000.

42.　Wanderley, M.M.; Vines, B.W.; Middleton, N.; McKay, C.; Hatch, W. The Musical Significance of Clarinetists' Ancillary Gestures: An Exploration of the Field. *J. New Music Res.* **2005**, *34*, 97–113.

43.　Davidson, J.W. Bodily communication in musical performance. In *Musical Communication*; OUP Oxford: Oxford, UK, 2005; pp. 215–238.

44.　Glowinski, D.; Mancini, M.; Cowie, R.; Camurri, A.; Chiorri, C.; Doherty, C. The movements made by performers in a skilled quartet: A distinctive pattern, and the function that it serves. *Front. Psychol.* **2013**, *4*, 1–9.

45. Palmer, C. *Music Performance: Movement and Coordination*; Elsevier: Amsterdam, The Netherlands, 2012.

46. Dahl, S.; Friberg, A. Visual Perception of Expressiveness in Musicians' Body Movements. *Music Percept. Interdiscip. J.* **2007**, *24*, 433–454.

47. Wanderley, M.M. *Quantitative Analysis of Non-Obvious Performer Gestures*; Springer: Berlin/Heidelberg, Germany, 2002; pp. 241–253.

48. Nusseck, M.; Wanderley, M.M. Music and Motion—How Music-Related Ancillary Body Movements Contribute to the Experience of Music. *Music Percept. Interdiscip. J.* **2009**, *26*, 335–353.

49. Bhatara, A.K.; Duan, L.M.; Tirovolas, A.; Levitin, D.J. Musical expression and emotion: Influences of temporal and dynamic variation. 2009, manuscript submitted for publication.

50. Sloboda, J.A.; Lehmann, A.C. Tracking Performance Correlates of Changes in Perceived Intensity of Emotion During Different Interpretations of a Chopin Piano Prelude. *Music Percept.* **2001**, *19*, 87–120.

51. Sakata, M.; Wakamiya, S.; Odaka, N.; Hachimura, K. Effect of body movement on music expressivity in jazz performances. In *International Conference on Human-Computer Interaction*; Springer: Berlin/Heidelberg, Germany, 2009; pp. 159–168.

52. Thompson, M.R.; Luck, G. Exploring relationships between pianists' body movements, their expressive intentions, and structural elements of the music. *Musicae Scientiae* **2012**, *16*, 19–40.

53. Van Zijl, A.G.W.; Luck, G. Moved through music: The effect of experienced emotions on performers' movement characteristics. *Psychol. Music* **2013**, *41*, 175–197.

54. Camurri, A.; Lagerlöf, I.; Volpe, G. Recognizing emotion from dance movement: Comparison of spectator recognition and automated techniques. *Int. J. Hum. Comput. Stud.* **2003**, *59*, 213–225.

55. Leman, M. *Embodied Music Cognition and Mediation Technology*; Mit Press: Cambridge, CA, USA, 2008.

56. Mancas, M.; Madhkour, R.B.; Beul, D.D. Kinact: A saliency-based social game. In Proceedings of the 7th International Summer Workshop on Multimodal Interfaces, Plzen, Czech Republic, 1–26 August 2011; Volume 4, pp. 65–71.

57. Shaffer, L.H. How to interpret music. In *Cognitive Bases of Musical Communication*; American Psychological Association: Washington, DC, USA, 1992; pp. 263–278.

58. Palmer, C. Music Performance. *Ann. Rev. Psychol.* **1997**, *48*, 115–138.

59. Goffman, E. *The Presentation of Self in Everyday Life*; Penguin Books: London, UK, 1959.

60. Scherer, K.R.; Keith, G.; Schaufer, L.; Taddia, B.; Pregardien, C. The singer's paradox: on authenticity in emotional expression on the opera stage. In *Emotion*; OUP Oxford: Oxford, UK, 2013; pp. 55–73.

61. Wöllner, C. How to quantify individuality in music performance? Studying artistic expression with averaging procedures. *Front. Psychol.* **2013**, *4*, 1–3.

62. Juslin, P.N. From Everyday Emotions to Aesthetic Emotions: Towards a Unified Theory of Musical Emotions. *J. Psychol.* **2013**, *10*, 235–266.

63. Brunswik, E. *Perception and the Representative Design of Experiments*; Univer: Berkeley, CA, USA, 1956.

64. Grandjean, D.; Bänziger, T.; Scherer, K.R. Intonation as an interface between language and affect. *Prog. Brain Res.* **2006**, *156*, 235–247.

65. Krumhuber, E.G.; Tamarit, L.; Roesch, E.B.; Scherer, K.R. FACSGen 2.0 animation software: Generating three-dimensional FACS-valid facial expressions for emotion research. *Emotion* **2012**, *12*, 351.

66. Roesch, E.; Tamarit, L.; Reveret, L.; Grandjean, D.; Sander, D.; Scherer, K. FACSGen: A tool to synthesize emotional facial expressions through systematic manipulation of facial action units. *J. Nonverbal Behav.* **2011**, *35*, 1–16.

67. Guadagno, R.E.; Blascovich, J.; Bailenson, J.N.; Mccall, C. Virtual humans and persuasion: The effects of agency and behavioral realism. *Media Psychol.* **2007**, *10*, 1–22.

68. Garau, M.; Slater, M.; Vinayagamoorthy, V.; Brogni, A.; Steed, A.; Sasse, M.A. 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, Ft. Lauderdale, FL, USA, 5–10 April 2003; pp. 529–536.

69. Schmitz, C. LimeSurvey: An Open Source Survey Tool. 2012. Available online: http://www.limesurvey.org (accessed on 5 November 2016).

70. Dardard, F.; Gnecco, G.; Glowinski, D. Automatic classification of leading interactions in a string quartet. *ACM Trans. Interact. Intell. Syst.* **2016**, *6*, 1–27.

71. Glowinski, D.; Baron, N.; Shirole, K.; Coll, S.Y.; Chaabi, L.; Ott, T.; Rappaz, M.A.; Grandjean, D. Evaluating music performance and context-sensitivity with Immersive Virtual Environments. *EAI Endors. Trans. Creat. Technol.* **2015**, *2*, 1–10.

72. Juslin, P.N.; Liljestrom, S.; Laukka, P.; Vastfjall, D.; Lundqvist, L.O. Emotional reactions to music in a nationally representative sample of Swedish adults: Prevalence and causal influences. *Music. Sci.* **2011**, *15*, 174–207.

73. Lartillot, O.; Cereghetti, D.; Eliard, K.; Trost, W.J.; Rappaz, M.-A.; Grandjean, D. Estimating Tempo and Metrical Features by Tracking the Whole Metrical Hierarchy. In Proceedings of the 3rd International Conference on Music & Emotion (ICME3), Jyväskylä, Finland, 11–15 June 2013; pp. 11–15.

74. Burger, P. *MoCap Toolbox—A Matlab Toolbox for Computational Analysis of Movement Data*; Logos Verlag Berlin: Berlin, Germany, 2013; pp. 172–178.

75. Lartillot, O.; Toiviainen, P. A matlab toolbox for musical feature extraction from audio. In Proceedings of the 10th Int Conference on Digital Audio Effects DAFx07, Bordeaux, France, 10–15 September 2007; pp. 1–8.

76. McLean, R.A.; Sanders, W.L.; Stroup, W.W. A unified approach to mixed linear models. *Am. Stat.* **1991**, *45*, 54.

77. Zentner, M. (University of Innsbruck, Innsbruck, Austria). Personal Communication, 2004.

78. Cohen, J. A power primer. *Psychol. Bull.* **1992**, *112*, 155–159.

79. Westermann, R.; Stahl, G.; Hesse, F. Relative effectiveness and validity of mood induction procedures: analysis. *Eur. J. Soc. Psychol.* **1996**, *26*, 557–580.

80. Graybiel, A.M. Habits, rituals, and the evaluative brain. *Annu. Rev. Neurosci.* **2008**, *31*, 359–387.

81. Anderson, C.R. Locus of control, coping behaviors, and performance in a stress setting: A longitudinal study. *J. Appl. Psychol.* **1977**, *62*, 446.

82. Juslin, P.N.; Karlsson, J.; Lindström, E.; Friberg, A.; Schoonderwaldt, E. Play it again with feeling: Computer feedback in musical communication of emotions. *J. Exp. Psychol. Appl.* **2006**, *12*, 79.

83. Juslin, P.N.; Persson, R.S. Emotional communication. In *The Science and Psychology of Music Performance: Creative Strategies for Teaching and Learning*; Oxford University Press: Oxford, UK, 2002; pp. 219–236.