Can naïve observers distinguish a violinist's solo from an ensemble performance? A pilot study

Donald Glowinski, Antonio Camurri
InfoMus Lab – Casa Paganini
Genoa, Italy
name.lastname@unige.it

Kim Torres-Eliard, Didier Grandjean
NEAD
Geneva, Switzerland
name.lastname@unige.ch

Carlo Chiorri
DISFOR
University of Genoa, Italy
name.lastname@unige.it

ABSTRACT
This paper investigates whether specific non-verbal behavioral variables may enable to distinguish between performing an action alone or jointly in a group. We consider the test case of a first violinist in a string quartet. Starting from the observation of audio-video recordings, non-expert participants were instructed to report whether they reckoned the performance being a solo or an ensemble one. The role of the musician's expressivity and expressed emotions on the participants' perception was also investigated.

Categories and Subject Descriptors
H.5.5 [**Information Interfaces and Presentation**]: Sound and Music Computing

General Terms
Experimentation.

Keywords
Ensemble Music Performance, Emotion, Non-verbal expressive behavior.

1. INTRODUCTION
Playing music jointly with others may affect individual behavior. Joint performance requires strategies to cope with others' intentions and actions and to adapt one's behavior accordingly. The success of the interaction between musicians may depend upon one's ability to anticipate and manage others' actions and ensure efficient group coordination. This paper aims at addressing a few issues about the interaction between musicians: to what extent external observers, e.g. the audience, can identify such attitude and sensitivity to others' behavior, including emotional processing among the musicians? Are there any specific auditory and visual non-verbal cues that may help in distinguishing between a solo and a group performance? To answer these questions, we considered a music ensemble scenario and, specifically, the audio-video recordings of the first violinist of a string quartet performing alone and with the other musicians of his ensemble.

SBM'12, October 26, 2012, Santa Monica, California, USA.

Music ensemble analysis can stand as an original test-bed to analyze social interaction and to investigate the social behavior of an individual, such as how she adjusts his/her own behavior to reach a successful interaction with others. Lately, an increasing number of studies adopted orchestra and small music ensemble scenarios to study interpersonal interaction between musicians themselves and with the audience. The EU ICT-FET three-year Project SIEMPRE (May 2010 - April 2013) has undertaken cross-disciplinary research to investigate novel paradigms and computational models of non-verbal creative group communication also adopting music scenarios (www.siempre.infomus.org). Within this project, new techniques were developed for the automated analysis of multimodal recordings of the musicians' performance in the two conditions used in the present study: *solo* vs *ensemble* performance. The main aim is to identify a set of non-verbal cues, potentially used by the audience, that characterize the social behavior and the emotional reactions of the musician, including, e.g., communicative gestures to regulate the ensemble performance or expressive emotional behaviors that may be used by an individual to distinguish between the two performance conditions (solo vs ensemble). Specifically, this study aimed at (i) investigating whether external observers can distinguish between the two modalities of the performance (solo vs ensemble) using audio-video recordings, and (ii) investigating the role of emotional cues expressed by the musician in the strategies adopted by the participants. The ultimate goal will be to correlate the results of the perceptual experiments (participants' ratings) with the results from the automated behavioral analysis of musicians.

The paper is organized as follows: Section 2 gives an overview of the related work on the perception of music performance; Section 3 provides the details of the experiment we designed to investigate whether and how one can distinguish between a solo *Vs*. an ensemble performance; Section 4 presents and discuss the results obtained. Conclusions and implications are reported in Section 5.

2. RELATED WORK
2.1 Behavioral cues in music performance
An increasing number of studies investigated non-verbal behavior, including those related to emotional processes, in music performance. In this context, two main types of cues have been pointed out: first, key gestures using upper-body parts such as head, hand to capture others' attention and to coordinate the ensemble. This first set of cues includes for example impulsive head's nod to indicate a synchronous start [6] or gaze interaction to capture co-performers' attention [14]. The temporal dynamics of human behavior can be decisive in distinguishing between observed behavioral expressions. The pilot study by Castellano et

al. on pianists' performance showed that dynamic aspects of motion features are complementary to postural and gesture shape-related information [3]. As observed by Davidson et al., some can be self-explanatory gestures (e.g., nods, thumbs up); whereas other specific gestures can be developed within an ensemble through rehearsal, and they may acquire a specific meaning only for the members of the ensemble [6].

The second set of non-verbal cues includes long-range behavioral variations, which are gradual and may not be as salient as the gestures described above. These behavioral cues may refer to implicit adaptation, emotional communication, and co-ordination processes of musicians during the performance [7]. Studies show that communicative gestures may typically cover expressivity aspects as well: the intensity of body sway or amplitude of arm movement, for example, may aim at regulating the whole ensemble performance. These gestures may also be a part of an emotional expression [4]. Most studies used an observational approach that focused on collective rehearsal of music ensembles (e.g., quartet). They considered specific attitudes that may support communication and coordination between musicians, which naturally include emotional expressive aspects. On the opposite, other studies have investigated performance of the same piece with different levels of expressivity and compared individual performances [13]. To our knowledge, no experimental design has been setup so far to disentangle what may directly relate communicative with expressive features.

2.2 Perceptual experiments

Several studies have been carried out to study how an external observer perceives music performance. A main focus has been on the communication of emotional expression between the performers and the observers. Juslin et al. [9,10] devised a framework of analysis based on the lens model (see below) to investigate how the cues related to the performance (e.g. tempo, loudness) map onto the perceptual cues of the observers. By analyzing the matching between performers' and listeners' cue utilization, one can predict the efficacy of the communication process. The Juslin's framework of analysis has been extended to include other modalities than audio, such as video. In particular, visual aspects of the performance have been analyzed to show how musicians' movement may provide expressive information that supplements the one conveyed by audio features [15]. Related research investigated the specific contribution of body parts in the evaluation of an expressive content. Several video-based processing have been applied: Dahl et al. [5], for example, presented participants with videos of playing musicians (marimba, clarinet) where selected body parts such as head or arm were masked. This experiment aimed at understanding how the evaluation of the performance may depend upon the available information on body during the performance. Motion capture data acquired through motion capture systems (e.g., Qualysis) or advanced RGB-D cameras (e.g., Kinect) have been also used to create point light displays of musicians to put in evidence the role of kinematic features in the communication of expressivity [11,12].

3. METHOD

3.1 Rationale

The humans are constantly involved in decoding and inferring the mental states, especially emotional ones, of others grounding on visual and auditory information. In the context of social communication, Brunswik defined a model of the perception and attribution of the mental states of others [1], namely the Lens Model. This model has been adapted in the context of musical communication in general [9] and between the performer and the listener in particular [10], see Figure 1). In the original version of Brunswik's model, the possibility for interactions between cues expressed by different musicians allowing the listener to infer a solo vs ensemble performance (i.e. the social context) were not specified. In the new version presented in Figure 1 we propose, according to the main aim of this study, a differentiation between the cues expressed by each musician and on how the social context, i.e. the relationships between the musicians, impacts specific cues allowing the listeners to infer the social context of the musical performance. Of course these kinds of information are usually implicitly processed by the listener. In the present experiment and future studies we propose to systematically investigate how the social context (represented by the XZ_{1n} interaction in Figure 1 as an example) is inferred and which kinds of cues are essential to be able to distinguish solo versus ensemble musical performances in an explicit judgment. The present study focuses on the ability of listeners to distinguish a solo from an ensemble performance, a first necessary step in the understanding of the ability to decode social context in musical performances. Future studies will investigate the nature and dynamics of the cues predicting explicit judgments differentiating solo vs ensemble musical performances. This study grounded on such model to investigate the cues used by the listeners to infer not only social aspects (i.e., solo vs ensemble performance) but also the role of emotional expressivity in the perception of musical performances.

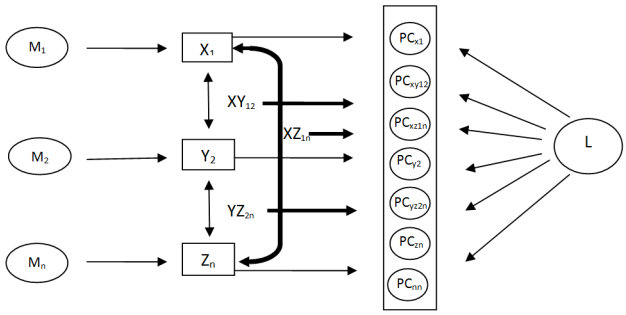

Figure 1. A social and emotional perspective of the Lens Model during a musical performance (adapted form Juslin & Lindstrom [9]). M: musician(s); X, Y, Z: social and/or emotional communication cues produced by musicians; XY, YZ, XZ: interactions between social and/or emotional cues; PC: perceived cues by listener (L) on which are based the attribution of social and/or emotional communication in a musical performance

3.2 Material

The material used for the experiment is a set of synchronized Audio/Video recordings of the Schubert *The Death and The Maiden piece,* interpreted by the first violinist of the *Quartetto di Cremona*. The selection of the video was based on the annotations made by the first violinist after each of his performances. We ensured that a broad range of expressive performance qualities could be represented in our sample recordings by considering the annotation given by the musician himself (e.g., worst and best interpretations). Videos focused on the upper-body part of the first violinist (720x576, 25 fps, mpeg2, Figure 3) after being has been cropped and centered on the position of the musician. Audio from piezoelectric-microphones attached to the instrument has been synchronized with the video (stereo, 48Khz, mp3) so that other instruments of the quartet cannot be heard in the ensemble

condition. Each audio/video recording of the Schubert music fragment has further been spotted into five short videos corresponding to five musical segments that have their own writing style (e.g., harmonic texture *Vs.* fugato writing style).

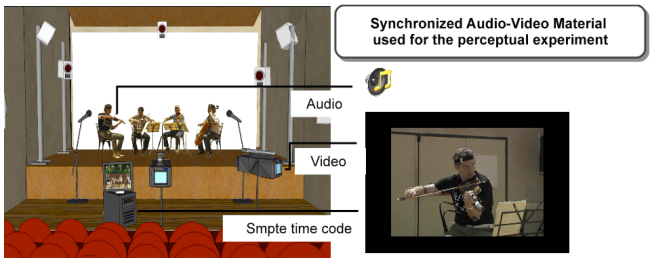

Figure 2. Illustration of the synchronized audio-video material used for the perceptual experiment. On left side, a view of the Multimodal Setup for the recording of the Quartetto di Cremona at Casa Paganini (University of Genoa). The selected material consisted in (i) audio from piezo-electric microphone attached to the first violinist instrument (ii) High-quality video centered on the first violinist (iii) smtpe time code managed through the EyesWeb software platform to synchronize the audio-video streams.

3.3 Participants

Twenty participants (5 males) took part to the experiment (Mean age 23.3±2.9 years, range 18–31). They received 15 CHF for their participation

3.4 Participants

Each participant was presented with a set of 60 samples selected from the full set of audio-video recordings of the first violin's performance. A procedure based on random permutation of pre-established lists of samples ensured that the Solo and Ensemble conditions as well as the five musical segments be presented with the same frequency. Audio-Video recordings were displayed through a dedicated Flash application on a flat screen (17") and headphones (Sennheiser) (see snapshot of the interface at Figure 3). The whole procedure consisted in three main phases:

1) After each audio-video sequence, the participants had to report whether they reckoned the performance being a solo or an ensemble one and then to rate their level of confidence in the correctness of their answer using a visual analogic continuous scale (from 1 to 100).

2) The second part of the questionnaire investigated the participants' perception of the musician's expressivity and expressed emotions. They were asked to assess the level of expressivity and the level of expressed emotions of the performance by rating the 9 GEMS dimensions [16].

3) At the end of the session, the participant was asked to report which musician's body features (e.g., head, arm, instrument movement) she/he most focused on to assess the performance.

The experiment duration was about 1h20.

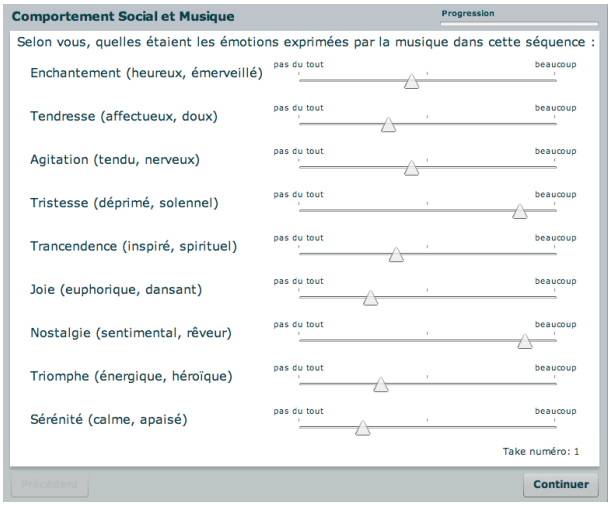

Figure 3. A snapshot of the French Flash interface developed for the perceptual experiment. This window deals with the ratings of the 9 GEMS items on a visual analogic continuous scale (from 1 to 100). English translation of the 9 GEMS items are: Wonder, Tenderness, Tension, Sadness, Transcendence, Joyful Activation, Nostalgia, Power, and Peacefulness.

4. RESULTS AND DISCUSSION

Do participants successfully distinguish between *Solo Vs Ensemble* performance? A Fisher's exact test showed no significant association of Condition (*Solo* vs *Ensemble*) with the Perceived Condition (*Perceived_Solo* vs *Perceived_Ensemble*) (p=.387, Figure 4). The same test has been performed using the participants' ratings weighted out by the level of confidence. In this latter case also, no significant association has been found. (p = .575).

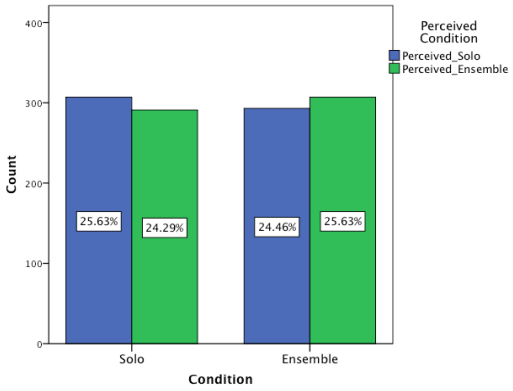

Figure 4. Bar Chart of the Experimental Condition (*Solo* and *Ensemble*) Vs Perceived Condition (*Perceived_Solo* and *Perceived_Ensemble*). The *Perceived_Solo* and *Perceived_Ensemble* responses are nearly equally distributed over the two experiment conditions

Received Operator Characteristics (ROC) curves were further employed to assess each participant's "diagnostic" accuracy. Though no effect for the whole group has been found (Area Under the Curve = .513, p = .453), individuals showed a wide range of performance accuracy, from .333 to .783

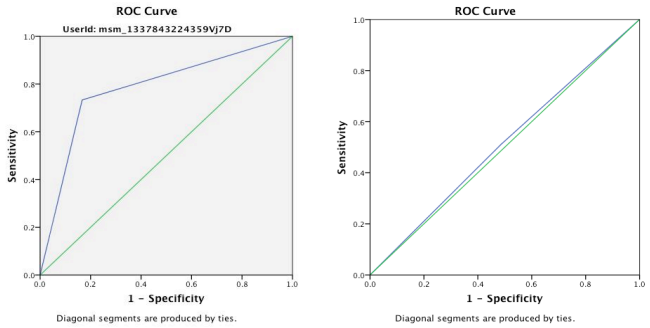

Figure 5. On left side, the ROC curve of the most successful participant (AUC=.783); on right side, the ROC curve considering ratings all 20 participants of the experiment (AUC=.513). AUC values closer to 1 indicate reliable distinction between the *Solo* vs *Ensemble* condition, whereas values near .50 indicate the predictor is no better than chance

TABLE I. FIXED EFFECTS OF GENERALIZED LINEAR MIXED MODEL. NON SIGNIFICANT EFFECTS ARE OMITTED TO SAVE SPACE

Fixed Effect *Target: Answer (***Perceived_Solo *vs* Perceived_Ensemble*)*[a]				
Source	***F***	***df1***	***df2***	***Sig.***
Level of Confidence	4.959	1	976	.026
Music Segment	7.968	4	976	.000[b]
Condition x Sadness	9.941	1	976	.002
Condition x Nostalgia	9.128	1	976	.003
Level of Conf. x Tenderness	4.033	1	976	.045
Segment x Joy	4.409	4	976	.002
Segment x Serenity	2.493	4	976	.042

a. Reference category: *Ensemble* Condition

b. Bonferroni-corrected statistical significance was .0011

As shown in Figure 6, participants reported a higher confidence level when they judged that the first violinist was playing within an ensemble. It seems that a common strategy was adopted to distinguish social cues from the analysis of the musician's behavior. The current results however do not provide sufficient details to understand which behavioral characteristics may have been used, as no effect of perceived body characteristics was found

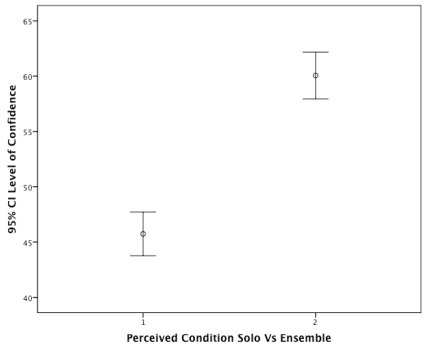

Figure 6. Level of Confidence (Error Bar plot) for the two perceived condition Solo Vs Ensemble

As shown in Figure 7, the Music Segment may have also affected the participants' answers ($X^2(4, n = 1198) = 78.394$, $p < .05$). The inspection of adjusted standardized residuals showed that Segment 4 was more likely to be considered as a *Solo* performance and Segment 5 as an *Ensemble* performance. The different writing style may explain this difference: in Segment 4, a fugato writing style sets all musicians at the same level by replicating the musical subject over the different instruments; all parts being equal with no leading part. In Segment 5, first violinist leads the ensemble, dialoguing mainly with the second violinist which ends the music phrase that first violin has started. These results suggest that when explicitly leading the ensemble, the first violinist exhibits specific attitudes and behaviors that may have been implicitly noticed by the participants, driving them to evaluate this segment more frequently as an Ensemble Condition. However, this did not lead to an increase in accuracy, since log-linear modeling revealed that accuracy did not differ across segments.

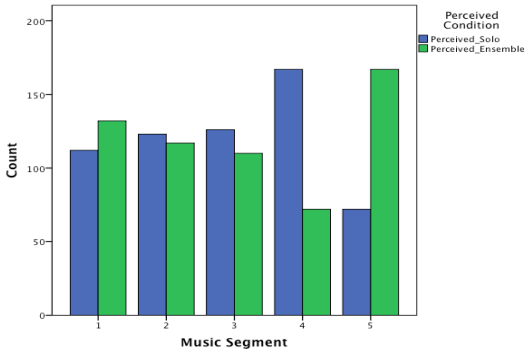

Figure 7. Bar Chart of the Music Segment (*Solo* and *Ensemble*) Vs Perceived Condition (*Perceived_Solo* and *Perceived_Ensemble*).

A third result of interest to understand participants' perceptions relates to the marginal interaction effect of the two emotions *Nostalgia* and *Sadness* with the Condition (*Solo* and *Ensemble*; see Table 1 and Figure 8). When the first violinist is correctly recognized as performing *Solo* or *Ensemble*, participants tend to ascribe him higher ratings of Nostalgia and Sadness.

The marginal interaction effect of Level of Confidence by Tenderness was due to a significant positive correlation between the two variables in the *Perceived_Ensemble* trials ($r = .23$), whereas the two variables did not show any linear association in the *Perceived_Solo* trials (see Figure 9).

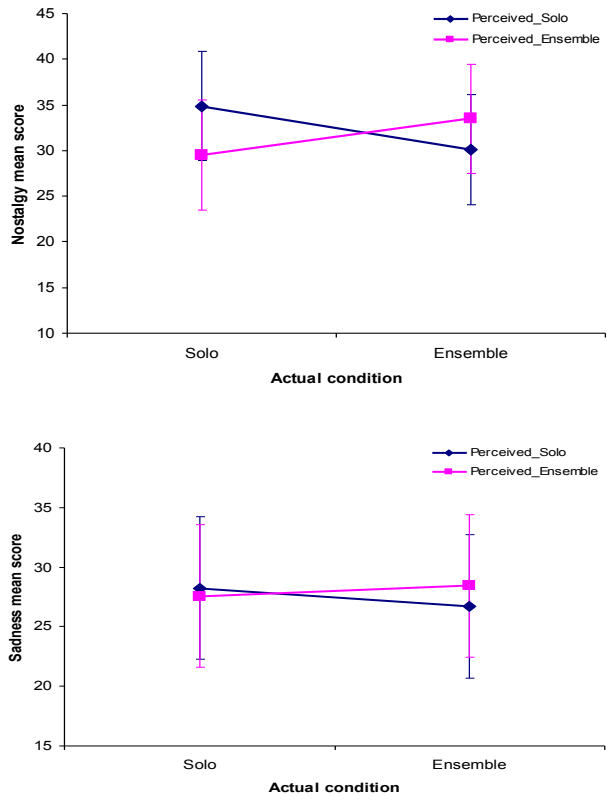

Figure 8. Plots of the interaction effect between Condition by Nostalgia and Condition by Sadness on perceived condition (solo vs ensemble).

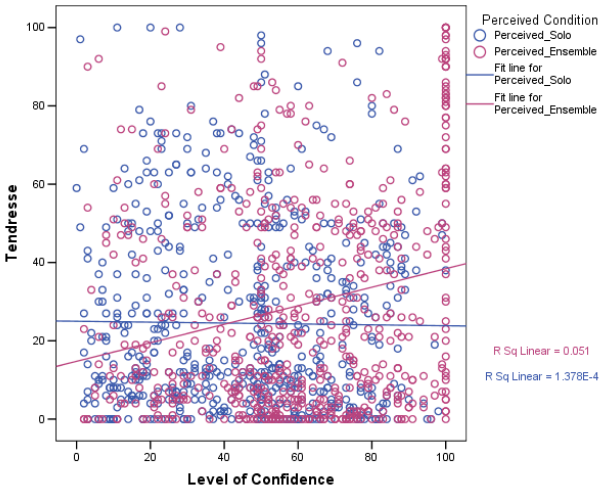

Figure 9. Plot of the interaction effect between Level of Confidence and Tenderness on perceived condition (solo vs ensemble)

This result suggests that in *Perceived_Ensemble* trials, a higher level of perceived Tenderness tend to be associated with higher levels of confidence of having provided a correct answer. Finally, the interaction of Segment with Joy was due to higher rating in Joy when a *Solo* condition was perceived in segments 1, 3 and 4, whereas the interaction of Segment with Serenity was due to higher rating in Serenity when a *Solo* condition was perceived in segments 1, 3 and 5 (Figure 10)

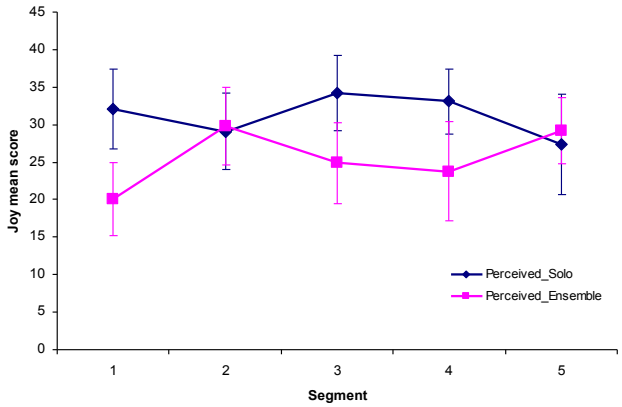

Figure 10. Plots of the interaction effect Segment by Joy and Segment by Serenity on perceived condition (solo vs ensemble)

All the marginal interaction effects observed so far may characterize participants' observations, which are specific to the first violinist performance. In this perspective, the idiosyncrasy of the musician's movement or his own playing style for example may have biased the experiment and may prevent us from generalizing the results. Yet, these interaction effects have detailed an intriguing aspect of social cognition, which is how the perceived emotions expressed by the musician can act as a moderator on the judgment of the observers.

5. CONCLUSION

This study is a first, pilot attempt to investigate social behavior in music performance and to identify a set of non-verbal cues explaining the phenomenon. The experimental data collected so far using audio-video recordings have indicated that non-expert participants may have difficulties in distinguishing two modalities of interpretation of a first violinist: when playing alone (solo) and when playing with the other musicians of a string quartet (ensemble). However, the analysis of the participants' ratings, including their evaluation of musician's expressivity and emotions, seemed to suggest original strategies for decoding social behavior: when perceiving the Ensemble condition, participants tended to be sensitive to the music segment where the first violinist has clear leadership and they tended to assess correctly identified solo and ensemble performances with higher ratings of Nostalgia and Sadness.

Future work is needed and may include the use of point-light displays of the first violinist based on the collected motion capture data during the recordings. This new material, which captures in more detail the kinematic features of the performance, should enable to achieve a better understanding of the behavioral cues used by the participants. Other possible tracks for future research may include some changes in the procedure used to collect participants' data, such as: (i) addressing one modality at a time to have a better control on behavioral cues that have effect on participants' ratings; (ii) addressing experts (creation of focus group) and (iii) correlating the results of the perceptual experiments (participants' ratings) with the results from the automated behavioral analysis of musicians.

6. ACKNOWLEDGMENTS

The research presented in this paper has been partially funded by EU FP7 ICT FET SIEMPRE (Social Interaction and Entrainment using Music PeRformance Experimentation) Project No. 250026 (May 2010 - June 2013).

7. REFERENCES

[1] E. Brunswik. Perception and the Representative Design of Psychological Experiments. University of California Press, Berkeley, 1956.

[2] A. Camurri, P. Coletta, G. Varni, and S. Ghisio. Developing multimodal interactive systems with EyesWeb XMI. In Proceedings of the 7th international conference on New interfaces for musical expression, pages 305–308. ACM Press New York, NY, USA, 2007.

[3] G. Castellano, M. Mortillaro, A. Camurri, G. Volpe, and K. Scherer. Automated Analysis of Body Movement in Emotionally Expressive Piano Performances. Music Perception, 26(2):103–119, 2008.

[4] S. Dahl, F. Bevilacqua, R. Bresin, M. Clayton, L. Leante, I. Poggi, and N. Rasamimanana. Gestures in Performance. Musical Gestures: Sound, Movement, and Meaning, 2009.

[5] S. Dahl and A. Friberg. Visual Perception of Expressiveness in Musicians' Body Movements. Music Perception, 24(5):433–454, 2007.

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

[7] D. Glowinski, P. Coletta, G. Volpe, A. Camurri, C. Chiorri, and A. Schenone. Multi-scale entropy analysis of dominance in social creative activities. In Proc. of the Intl Conf on Multimedia, pages 1035–1038. ACM, 2010.

[8] P. Juslin and E. Lindstrom. Musical expression of emotions: Modelling listeners' judgements of composed and performed features. Music Analysis, 29(1-3):334–364, 2010.

[9] P. Juslin and J. Sloboda. Music and emotion: theory and research. Oxford University Press, Oxford, 2001.

[10] P. Juslin and J. Sloboda. Handbook of music and emotion: theory, research, applications. Affective Science. Oxford University Press, 2010.

[11] P. Keller and M. Appel. Individual differences, auditory imagery, and the coordination of body movements and sounds in musical ensembles. Music Perception, 28(1):27–46, 2010.

[12] V. Sevdalis and P. Keller. Cues for self-recognition in point-light displays of actions performed in synchrony with music. Consciousness and cognition.

[13] J. Sloboda. The acquisition of musical performance expertise: Deconstructing the" talent" account of individual differences in musical expressivity. 1996.

[14] W. Thompson, P. Graham, and F. Russo. Seeing music performance : Visual influences on perception and experience. Semiotica, 2005(156):203–227, 2005.

[15] B. Vines, C. Krumhansl, M. Wanderley, and D. Levitin. Cross-modal interactions in the perception of musical performance. Cognition, 101(1):80–113, 2006.

[16] M. Zentner, D. Grandjean, and K. Scherer. Emotions evoked by the sound of music : characterization, classification, and measurement. Emotion, 8(4):494–521, 2008