Proceedings of the 3rd International Conference on Music & Emotion (ICME3), Jyväskylä, Finland, 11th - 15th June 2013. Geoff Luck & Olivier Brabant (Eds.)

AURAL-BASED DETECTION AND ASSESSMENT OF REAL VERSUS ARTIFICIALLY SYNCHRONIZED STRING QUARTET PERFORMANCE

Panos Papiotis, Perfecto Herrera Marco Marchini, Esteban Maestre

Music Technology Group, Universitat Pompeu Fabra, Spain
panos.papiotis@upf.edu

Abstract

In a musical ensemble musicians can influence each other's performance in terms not only of timing but also in other aspects of the performance such as dynamics, intonation, and timbre. The goal of this work is to test whether this influence can be perceived by a listener from an audio recording solely. We utilize a set of string quartet recordings where every piece is recorded in two experimental conditions: the solo condition, where each musician performs alone; and the ensemble condition, where the musicians perform together after a brief rehearsal. Using state-of-the-art audio analysis/synthesis methods, we artificially synchronize the recordings in the solo condition note-by-note, thus generating a set of pseudo-ensemble performances where there is no interaction between the musicians. We then carry out a series of listening tests: first, the subjects are tasked with comparing the quality of the performance and the degree of coordination for the two recordings, without knowing that one of them is artificially synchronized. Then, we reveal to the listeners that one of the two versions is artificially synchronized and ask them to point out which recording is which. The results suggest that listeners cannot easily discriminate between the real and artificially synchronized recordings; furthermore, the accuracy of their judgements appears to be affected by the listeners' level of musical training as well as the piece that is performed.

Keywords: ensemble performance, listening experiment, interdependence

1. Introduction

Ensemble music performance is a special case of goal-oriented social collaboration where communication is carried out nonverbally, with the performers' intentions being mediated through expressive gestures and the produced sound (Keller, 2008). One can safely assume that the first step in achieving coordinated action is rhythmic synchronization; however, depending on the instrument, the musicians can also coordinate their actions in other aspects of performance (such as dynamics, intonation, and timbre).

Using computational means to detect evidence of this interdependence between the performers has proven to be a difficult yet achievable task (Papiotis et al, 2012). Our motivation behind this article is to assess whether the same can be achieved by human listeners; such an investigation can help us understand which aspects of music collaboration are most salient from a listener's point of view, as well as identify the skills that affect the listener's perception.

Previous work on this subject is limited. Glowinski et al carried out an experiment where subjects were asked to decide whether a recorded segment was performed solo or in an ensemble by observing only the first violinist of a string quartet ensemble (Glowinski et al, 2012). Besides the perceived performance condition, the subjects also rated the musi-

cian's expressivity and the expressed emotions of the performance, while also describing which of the musician's body features (head motion, arm motion, etc.) they focused on in order to make their assessment. Results did not show significantly different assessments for the solo and ensemble conditions, although the expressivity and expressed emotion ratings showed some significant interaction with the two conditions (solo, ensemble).

Listening experiments have been employed in similar tasks. Examples include judging whether a recorded performance was composed or improvised (Lehmann and Kopiez, 2010), and whether different excerpts had been played by the same performer (Gingras et al, 2011). Finally, listening experiments have also been used to evaluate the simulation of an orchestral violin section from a single recording (Pätynen, 2011).

Our aim in this work is to assess how reliably can human listeners detect evidence of musical interdependence when listening to recorded performances of an ensemble. Our methodology is to carry out a listening experiment where listeners compare real string quartet recordings to artificially synchronized solo recordings of the same piece. We utilize short piece excerpts of varying characteristics and investigate how the listeners' judgements are affected by them as well as the listeners' own background.

The rest of this article is organized as follows: in Section 2, we describe the music material that was recorded for the listening experiment, the processing that is applied to the recordings, and the experimental process. In Section 3, we present the acquired results; finally, in Section 4, we discuss the implications of our findings and offer some concluding remarks.

2. Method

2.1. Material

The recordings used for the listening experiment consist of five short excerpts of string quartet pieces, performed by a professional string quartet. All pieces were part of the quartet's current repertoire, and each excerpt was manually selected for its different qualitative characteristics, which we assessed with the help of a professional string performer. Table 1 shows a summary of each excerpt:

Table 1. Summary of the excerpts used for the experiment and their most salient characteristic.

ID	Piece	Dur.	Characteristic
P1	Borodin – String quartet nr.2 in D Major, 3rd Movt.	00:58	Phrasing
P2	Borodin – String quartet nr.2 in D Major, 1st Movt.	00:46	Dynamics, Intonation
P3	Beethoven – String Quartet nr. 4 (op. 18), 1st Movt.	00:36	Dynamics
P4	Beethoven – String Quartet nr. 4 (op. 18), 1st Movt.	00:42	Rhythm
P5	Beethoven – String Quartet nr. 4 (op. 18), 3rd Movt.	01:21	Rhythm, Phrasing

Each piece excerpt was recorded in two conditions: *solo*, where each musician performed alone without any previous rehearsal, and *ensemble*, where the musicians performed together after a brief rehearsal period. No metronome signal was provided in any of the recordings. The *solo* and *ensemble* recordings of each piece excerpt were carried out on separate days.

2.2. Data acquisition & processing

An individual audio signal from each performer was acquired through a piezoelectric pickup fitted on the bridge of each instrument. The use of pickup signals from each musician allows for efficient post-processing with minimal artifacts (due to the absence of room ambience).

All recordings were automatically score-aligned using a dynamic programming routine and manually corrected to ensure that the annotated note onset times are accurate.

2.3. Artificial synchronization

Given that the recordings were carried out without a metronome, it was necessary to artificially synchronize the *solo* recordings; moreover, since our goal was to assess whether listeners can detect musical coordination based on factors other than rhythmic synchronization, it was also necessary to ensure that the *solo* recordings had exactly the same note onset/offset times as the *ensemble* recordings. We applied state of the art time scaling techniques (Bonada, 2000) to apply a non-linear time stretch to the *solo* recordings using the *ensemble* recordings as reference: for each individual instrument, the audio signal is partitioned using the note onset times as anchor points; then, the duration of each *solo* note is altered to match the duration of the corresponding *ensemble* note in the score; finally, the *solo* waveform is shifted to coincide with the *ensemble* waveform.

We carried out a pilot test to assess whether any audio artifacts are introduced by this procedure using music technology researchers as subjects, without encountering any. Earlier variants of this time-scaling algorithm have been also used in listening experiments without introducing any significant bias (Honing, 2006).

2.4. Post-processing

Given that bridge pickup recordings have a certain 'nasal' quality, all four pickup (bridge vibration) signals were respectively convolved with body impulse responses (Maestre et al 2013).

In order to reconstruct the stereo image of a string quartet's sound, the four recordings in each excerpt were panned from left to right as follows: violin 1 (60% left), violin 2 (20% left), viola (20% right), cello (60% right). Finally, the gains applied to each instrument's audio signal were manually set using stereo recordings of each excerpt as reference; the same gain was universally applied to all recordings.

2.5. Experiment

The listening experiment was carried out through an online survey system. Each subject was asked to use headphones in order to ensure similar listening conditions. Before listening to any recordings, the following personal information was gathered:

1. Age
2. Gender
3. Amount of (formal or informal) musical training (None, Up to 2 years, between 3 and 5 years, more than 5 years)
4. (Conditional to Training) Experience with bowed string instruments

After this step came Phase 1 of the experiment: the subject listened to the five recording pairs (*solo* and *ensemble*) in random order within the experiment (*solo* first or *ensemble* first), but the same order across all subjects. It is important to note that, at this time in the experiment, the subject was not aware that only one of the recordings is from a 'real' ensemble. The subject was tasked with listening to each pair of recordings and comparing them in terms of *Quality of performance* and *Degree of coordination*; there was also the option of considering both recordings equal.

In Phase 2, the subject was then informed that one of the recordings is real while the other is artificially synchronized. Then, the subject listened to the same five recording pairs again, this time with the task of choosing the recording he/she believed to be the real ensemble recording. Similarly to Phase 1, the subject could answer '*I am unable to decide*'. Finally, a comments' form was provided for each excerpt where the subjects could specify what helped them make their decision.

3. Results

We analyzed the responses of 74 subjects (51 males). The mean age of the subjects was 32 years old (standard deviation = 11). 39 subjects had received more than 5 years of musical training, while 8 subjects had experience with bowed string instruments.

An overview of the subjects' responses can be seen in Figures 1 and 2; Figure 1 shows which recording was rated with a higher 'performance quality' per excerpt across all sub-

jects, while Figure 2 shows which recording was rated with a higher 'degree of coordination'.

One can already observe that each excerpt elicits a different response from the subjects. Especially the last two excerpts seem to be the most difficult to compare; given that we selected those two excerpts as examples of rhythmic coordination, it seems plausible that by making the *solo* and *ensemble* recordings identical in terms of note onsets and offsets, we are equalizing them in the aspect of the performance on which the musicians were most focusing on.

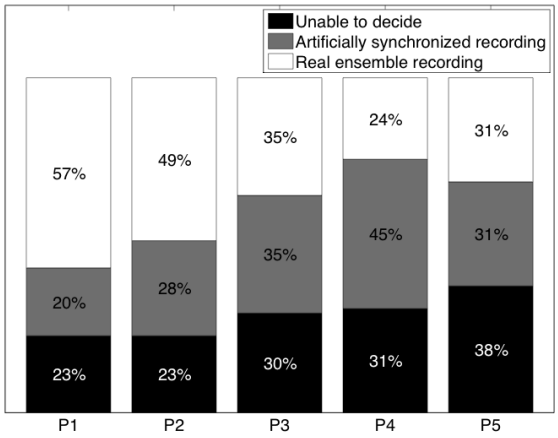

Figure 1. Collected responses for all subjects regarding Performance Quality (Phase 1). See Table 1 for the meaning of the different bars (Px).

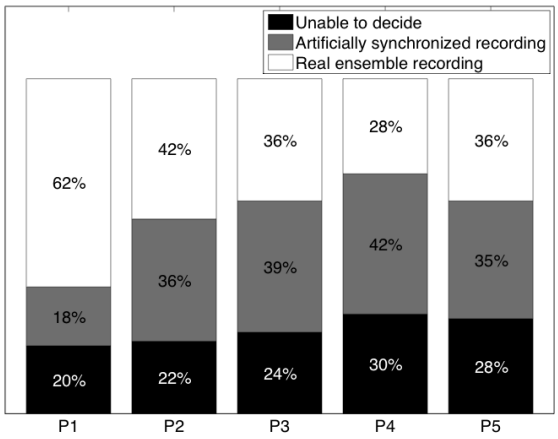

Figure 2. Collected responses for all subjects regarding Degree of Coordination (Phase 1).

Another observation that can be made from the above figures is that the subjects' ratings for 'performance quality' and 'degree of coordination' appear to be in relative agreement; this was confirmed by measuring the Spearman's rank correlation coefficient between these two factors per excerpt; the obtained *rho* values are as follows: P1: 0.84, P2: 0.73, P3: 0.72, P4: 0.66, P5: 0.71 (p-value<0.001 for all cases).

Regarding Phase 2 of the experiment, Figure 3 shows which recording was chosen as the real quartet recording across all subjects, per excerpt.

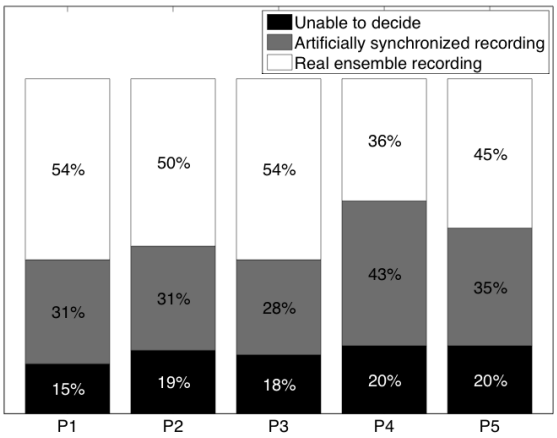

Figure 3. Recording finally chosen as "real" by all the subjects (Phase 2).

Again, it can be seen that listeners encounter difficulties in detecting the real quartet recording, with some piece excerpts showing higher accuracy than others in the same way as in Phase 1.

So far, we have not investigated the effect of musical training on the subjects' responses; moreover, although we have seen that different excerpts provide varying results, the effect of each excerpt remains to be seen. Given that the variable on which the effect of training and excerpt we want to investigate is discrete we performed a logistic regression on the binary outcome of each comparison (=YES for the cases where the real quartet recording was chosen and =NO otherwise). The results can be seen in Tables 2, 3 and 4:

Table 2. Logistic regression results for Performance quality.

Coefficient	Estimate	Std. error	p
Excerpt P1	-0.404	0.320	0.207
Excerpt P2	-0.754	0.324	0.020
Excerpt P3	-1.361	0.341	<0.001
Excerpt P4	-1.932	0.370	<0.001
Excerpt P5	-1.562	0.350	<0.001
Training	0.277	0.108	0.010
String ex.	1.409	0.399	<0.001

Table 3. Logistic regression results for Degree of coordination.

Coefficient	Estimate	Std. error	p
Excerpt P1	-0.285	0.319	0.370
Excerpt P2	-1.161	0.330	<0.001
Excerpt P3	-1.402	0.337	<0.001
Excerpt P4	-1.794	0.354	<0.001
Excerpt P5	-1.402	0.337	<0.001
Training	0.363	0.107	<0.001
String ex.	0.667	0.370	0.071

Table 4. Logistic regression results for Final assessment.

Coefficient	Estimate	Std. error	p
Excerpt P1	-0.568	0.309	0.066
Excerpt P2	-0.738	0.311	0.017
Excerpt P3	-0.568	0.309	0.067
Excerpt P4	-1.314	0.325	<0.001
Excerpt P5	-0.964	0.315	0.002
Training	0.356	0.101	<0.001
String ex.	0.068	0.357	0.847

From the above results one can observe that the amount of musical training has a significant positive effect on the outcome; that is, subjects with higher amounts of musical training tend to be more accurate. Regarding experience with bowed string instruments, we could detect a significant positive effect only on the assessed performance quality; the small amount of subjects with string experience (8 out of 74) makes conclusive results difficult to achieve, and we believe that a more thorough investigation of the matter is called for.

Regarding the excerpt type, we can observe that excerpts P2, P4 and P5 seem to have the most significant effect on the subjects' ratings, at least for the final decision in Phase 2; for Phase 1 decisions, excerpts P2 to P5 all seem to significantly affect the subjects' ratings. We did not find any significant interaction between the coefficients, although the skewed distribution of some variables (such as experience with string instruments) makes interaction assessments difficult.

Finally, we wanted to compare the subjects' ratings with the estimated amount of interdependence in a music ensemble, as computed in terms of three aspects of the performance: Dynamics, Intonation, and Timbre (Papiotis et

al, 2012; Papiotis et al, 2013). Continuous audio and bowing gesture features are extracted as descriptors of the performance; then, computational measures of interdependence are applied between pairs of these features in order to assess the degree to which the musicians influence each other's performance. For the five piece excerpts used in this study, we computed the amount of interdependence on both the ensemble as well as the solo recordings; we then calculate the difference between ensemble and solo interdependence for each of the three aspects of the performance (Dynamics, Intonation, Timbre). In the excerpts where higher interdependence was measured for the solo condition, we simply assign a value of zero.

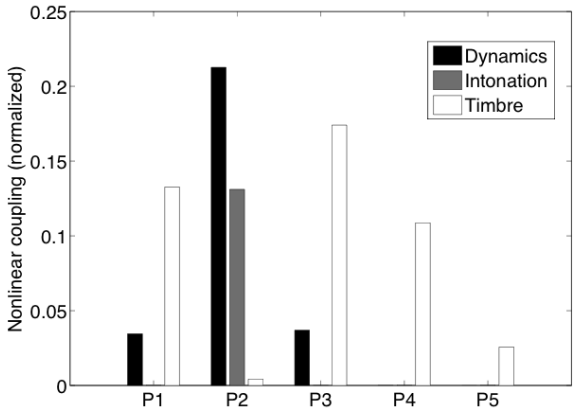

Figure 4. Estimated amount of interdependence for each excerpt, in terms of *Dynamics, Intonation* and *Timbre*.

As it can be seen in Figure 4, the findings are in agreement with the experiment results; the lowest amounts of interdependence are encountered in the P4 and P5 excerpts, while the highest amount of interdependence is encountered in the P2 excerpt (which was found to significantly affect the subjects' response through the logistic regression). On the other hand, overall interdependence values that are averaged across the duration of a recording cannot perfectly relate to human perception, where a short passage or small detail might be enough to make a judgement. This can be reflected in the listeners' ratings of excerpts P1 and P3 which demonstrate higher accuracy than excerpt P2 (whereas interdependence for excerpt P2 is overall higher than for excerpts P1 and P3).

4. Discussion

In this paper, we investigated the listeners' capability (or lack thereof) in discerning between real and artificially synchronized recordings which have the same degree of rhythmic synchronization. In general, our findings suggest that this is a difficult task that is significantly affected by the piece that is being performed and by the aspects of the performance it draws most focus on; however it is seen that synchronization, while of high importance, is not the only aspect of ensemble performance that is reflected through the acoustic result.

It has also been shown that musical training can improve the listeners' capabilities for correct discrimination between real and artificially synchronized performances.

Finally, although methods recently applied to quantifying musical interdependence in string quartets seem to be in agreement with the listeners' judgements, differences were observed for specific excerpts.

This has been an exploratory work, and there are many areas in which investigation can be improved and expanded. A more diverse selection of musical pieces as well as more participants with string performance experience should be included in further refinements of the experiment, while a more thorough analysis on computational methods of interdependence, score-level features, and their relation to the listeners' judgement should be attempted.

6. Acknowledgments

The work presented on this document has been partially supported by the EU-FP7 FET SIEMPRE project and an AGAUR research grant from Generalitat de Catalunya. The authors would like to thank Jordi Bonada for providing the audio time-scaling algorithm.

References

Bonada, J. (2000) Automatic Technique in Frequency Domain for Near-Lossless Time-Scale Modification of Audio. *Proceedings of the International Computer Music Conference*, Berlin, Germany

Gingras, B., Lagrandeur-Ponce,T., Giordano,B.L., McAdams,S. (2011). Perceiving musical individuality: Performer identification is dependent on performer expertise and expressiveness, but not on listener expertise. *Perception*, 40:1206-1220.

Glowinski, D., Torres-Eliard, K., Chiorri, C., Camurri, A. and Grandjean, D. (2012) Can naïve observers distinguish a violinist's solo from an ensemble performance? A pilot study. *ACM-ICMI SBM workshop*, Santa Monica, California, USA.

Honing, H. (2006). Evidence for tempo-specific timing in music using a web-based experimental setup. *Journal of Experimental Psychology: Human Perception and Performance*, 32:780 –786

Keller, P. (2008). Joint action in music performance. *Emerging Communication*, 10:205.

Lehmann, A. and Kopiez, R. (2010). The dfficulty of discerning between composed and improvised music. *Musicae Scientiae* 14:113-129.

Maestre, E., Scavone, G., Smith, J.O. (2013) Digital modeling of bridge driving-point admittances from measurements on violin-family instruments. Stockholm Musical Acoustics Conference (SMAC) 2013 (submitted)

Papiotis, P., Marchini, M, and Maestre, E. (2012). Computational analysis of solo versus ensemble performance in string quartets: Dynamics and Intonation. *In Proceedings of the 12th International Conference of Music Perception and Cognition* (ICMPC12), Thessaloniki, Greece.

Papiotis, P., Marchini, M, and Maestre, E. (2013). Multidimensional analysis of interdependence in a string quartet. *In Proceedings of the International Symposium on Performance Science (ISPS2013)*, Vienna, Austria.

Pätynen , J. (2011). A virtual symphony orchestra for studies on concert hall acoustics. Doctoral Dissertation, Aalto University, Finland