Timing synchronization in string quartet performance: a preliminary study

Marco Marchini[1], Panos Papiotis[1], and Esteban Maestre[2,1] *

[1] Music Technology Group, Universitat Pompeu Fabra, Barcelona
[2] Center for Computer Research in Music and Acoustics, Stanford University, CA
{marco.marchini, panos.papiotis}@upf.edu
esteban@ccrma.stanford.edu

Abstract. This work presents a preliminary study of timing synchronization phenomena in string quartet performance. Accurate timing information extracted from real recordings is used to compare timing deviations in solo and ensemble performance when executing a simple musical passage. Multi-modal data is acquired from real performance and processed towards obtaining note-level segmentation of recorded performances. From such segmentation, a series of timing deviation analyses are carried out at two different temporal levels, focusing on the exploration of significant differences between solo and ensemble performances. This paper briefly introduces, via an initial exploratory study, the experimental framework on which further, more complete analyses are to be carried out with the aim of observing and describing certain synchronization phenomena taking place in ensemble music making.

Keywords: music performance, ensemble performance, synchronization, timing, tempo, string quartet

1 Introduction

Music performance as the act of interpreting, structuring and physically realizing a composition is a highly complex human activity with many facets: physical, acoustic, physiological, psychological, social, artistic, etc. [4]. Trained musicians are able to read and interpret a composition in the form of a music score, which may end up conveying very different emotions depending on how it is performed, i.e., how the content is transformed into musical sound. In fact, it is commonly acknowledged that there is an important part of expression or meaning already borne by the actual piece to be performed, and another part introduced by the performer when freely navigating the space of performance resources (e.g., timing deviations, dynamics modulations, etc.) resulting from a combination of praxis habits and certain constraints imposed by the structure and content of the score [2]. In the search for exploring and understanding the process of music

* The authors would like to thank collaborators from CIRMMT, McGill University for their support in hosting the recordings: Carolina Brum, Erika Donald, Vincent Freour, Marcello Giordano, and Marcelo M. Wanderley.

9th International Symposium on Computer Music Modelling and Retrieval (CMMR 2012)
19-22 June 2012, Queen Mary University of London

performance as an accessible instance of human cognition, some researchers of a variety of disciplines have tried to approach the challenge by looking at music performance as a goal-directed task, considering such task as driven by the sequence of symbolic events appearing in a music score [11].

Ensemble music performance can be regarded as one of the most closely synchronized activities that human beings engage in (actions coordinated to within small fractions of a second are considered routine even in amateur performance). Unlike speech, musical performance is one of the few expressive activities allowing simultaneous participation. As such, the potential of music as a basis for studying basic principles of non-verbal communication and entrainment of emotion is unparalleled [7]. Studies in computational modeling of music performance have confirmed the widespread consideration of tempo and dynamics as the two most prominent resources available for musicians to convey emotion or expression during performance [14] (e.g., by acting with creative freedom to carve personalized and aesthetically pleasing executions), thus represeting two major dimensions over which to extract relevant information from a performance recording of a certain piece. In agreement with the importance of explicitly considering metric, melodic and harmonic structures of scores when approaching the study of music performance from a computational perspective [1], previous researchers have based their work on pairing musicological charactersitics of musical scores with performance aspects, especially timing and/or dynamics [3, 13, 15, 12]. From these two dimensions, available for ensemble musicians to coordinate and successfully achieve their shared goal, a first clear choice for extracting synchronization-related information from joint performance is to analyze how timing modulations get synchronized in different situations (e.g., solo versus ensemble) and different musical contexts (e.g., by accounting for score structure).

The computational study of timing synchronization among ensemble performers has been approached in the past. A vast literature has been inspired by the concept of "participatory discrepancy" introduced in [6] by Keil. Following the Keil's directions, an objective measure of performer's time discrepancies for several bass players was out carried in [10]. However only a one-way synchronization could be observed since the musicians where recorded playing solo on top of a recorded tape. More interaction paradigms were considered in the work by Goebl and Palmer [5], where the focus was put onto exploring the influence of auditory feedback and musical role (e.g., leadership) on timing (note onsets) and motion (finger and head) synchronization phenomena among duets of pianists. A second relevant example of two-ways auditory/visual feedback is the work by Moore and Chen [9], which pursued micro-timing analyses from arm motion data acquired from two members of a string quartet while performing a relatively difficult, yet thoroughly rehearsed task. Findings of both works showed timing and/or motion synchronization as an essential cue for the exploration of basic social behaviors in coordinated action.

In this paper, we present a preliminary study of timing synchronization phenomena among the four members of string quartet during performance. Timing information is extracted by processing multi-modal data acquired from real recording (providing a note-level score-performance alignment) of the execution of a simple musical passage that was unknown to the musicians. From annotated note onsets and offsets, a number of timing deviation analyses are carried out at two different temporal levels, focusing on the exploration of significant differences between solo and ensemble performances. Rather than with the aim of presenting a thorough study on the topic, this paper discusses an initial exploration experiment while introducing the framework and methods through which more complete and extended analyses are to be systematically carried out on a large corpus of quartet performance recordings being constructed at the moment.

The rest of the paper is structured as follows. Sec. 2 briefly introduces the general approach we envisioned and employed and explains what type of data we acquire from each experiment and summarizes the techniques used for preprocessing it. We then present some preliminary results on tempo in Sec. 3 and, finally, discuss them in Sec. 4.

2 Experimental framework

As verified by the above literature, the subject of collaborative musical performance is a very complex one. In order to obtain reliable results using computational means, the existence of valid hypotheses is of very high value; for that reason, we are working on an experimental framework which will provide a set of recordings where the studied relationships among the musicians are well defined and unambiguous. The final corpus of music pieces used, which will be detailed next, has been selected and modified using the help of a professional string quartet performer, and will in time be recorded by a number of different quartets.

The corpus is based on an exercise handbook for string quartets[3], intended for improving the "ensemble skills" of the quartet members. The material is divided into six categories, with each category containing a number of short exercises dealing with a different aspect of ensemble performance: *Intonation, Dynamics, Unity of Execution, Rhythm, Phrasing, and Tone production/Timbre.* An exercise consists of a simple, low difficulty score, together with annotations on what is the specific goal that must be achieved by the quartet.

We record the musicians' performance in three experimental conditions; solo (first sight), rehearsal, and ensemble. In the first condition (solo), each musician must perform their part alone without having access to the full ensemble score nor the instructions that accompany the exercise. In this way we wish to eliminate any type of external influence on the performance, be it restrictions imposed by other voices of the ensemble or instructions by the composer that are not in relation to the individual score of the performer. In the second condition (rehearsal) , following the solo recordings of each quartet member, the group of musicians is provided with the full ensemble score plus the composer instructions; they are then left to rehearse the exercise alone until they are able to fulfill

[3] Mogens Heimann - Exercises for the String Quartet.

the requirements of the exercise. In the third condition (ensemble), following the rehearsal, the quartet is finally recorded performing the exercise as a group.

In terms of data acquisition, both audio and motion capture data are recorded for each member of the ensemble; these data streams are synchronized in real time. Audio-wise, the individual signal from each musician is captured through the use of piezoelectric pickups attached to the bridge of the instrument. Instrumental - i.e. sound-producing - gestures such as bow velocity and force are also acquired through the use of a motion capture system, as detailed in [8].

For every recording a semi-automatic, note-level alignment between the performance and the music score is performed using a dynamic programming approach, a variation of the well-known Viterbi algorithm. This approach focuses into three main regions of each note: the note body and two transition segments (onset and offset). Different costs are computed for each segment, using features extracted by the audio (RMS audio energy, Fundamental frequency) as well as the bowing features described above. Finally, the optimal note segmentation is obtained so that a total cost (computed as the sum of the costs corresponding to the complete sequence of note segments) is minimized. This method, which can be seen in more detail in [8], has so far provided robust results that only in few occasions require manual correction. Through this alignment, it is then easy and accurate to extract detailed timing information for each performer. More complicated information such as the dynamics, timbre or articulation of the performance is extracted by combining the audio signal with the instrumental gesture features.

3 Preliminary study and initial results

The objective of the study presented here is to exploit content of some preliminary experiments and formulate new hypothesis to be tested in the next set of experiments. In this article we deal with some results arising from the experiment conducted with the exercise shown in Fig. 1. The exercise consists of an ascending and descending D major scale in thirds. The quartet is divided in two sections (violins in the first, viola and cello in the second) one alternating with the other. Musicians were instructed to play the score as if it was played by one instrument. We did not impose on them further constraints such as to follow a metronome.

For each case we recorded 4 consecutive repetition of the score (Fig. 1). A score alignment has been executed on each of those 8 performances. The analyzed set of data thus consists of 512 aligned onsets (64 per performance). We also derived a *joint-performance alignment* consisting of 128 onsets where for each third chord we compute an onset given by the mean of the individual attacks of the two notes that form that chord. The goal of this exercise is self-evident in the score. The notes within each group have to blend together while allowing the blocks of semiquaver notes formed by each group to slot together in a temporal order. In addition to that, the requirement of achieving a good “unity of execution” means that the parts played by each group have to be connected

Fig. 1. Score of the exercise employed for the experiment.

to the part of the subsequent section without disruptions in terms of tempo and dynamics. It is also worth to notice that the slurs contained in the scores, by requiring the musicians perform with a certain bow direction might also pose some constraints to the synchronization process.
We divided results of the analysis into *Micro-* and *Macro-* tempo results. We consider *Macro-tempo* as the tempo experienced by a listener in a relatively long region of time, it can also change in time but rather slowly. *Micro-Tempo* comprises of slight anticipations of note events followed by a deferral of the subsequent events in a way that the result does not contribute to macro-tempo changes.

3.1 Macro-tempo

By assuming, for a moment, the tempo to be constant when no onset occurs we derive a bpm step function defined to be the corresponding beat per minute value of each duration. This curve is noisy due to the differences in duration of the notes. In order to remove high frequency content and derive an overall tempo behavior we convolve it with a gaussian curve of variance $\sigma > 0$. From

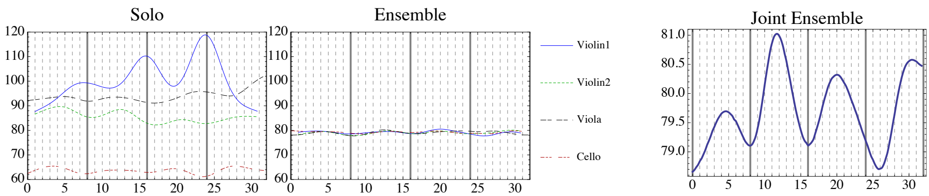
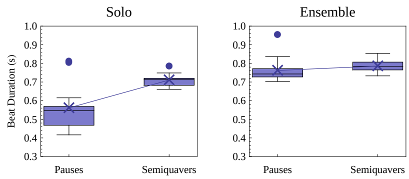
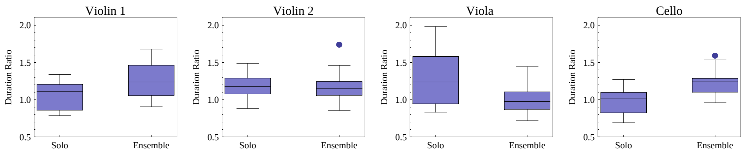


Fig. 2. Individual tempo curve of the four instruments for the solo (left) and the ensemble case (center). Vertical grid lines mark the boundaries of the repetitions (full line) and the beat start time (dashed line).The tempo curve of the joint ensemble performance is shown on the last plot (right).

the tempo curves derived for the solo/ensemble case we find that not only the mean tempo of the four musicians becomes the same, but also the variance of the tempo curve gets significantly smaller in the ensemble case. We can interpret this result both as an indicator that the freedom of the musicians gets restricted and as a result of the collaborative way in which the tempo is jointly shaped. Fig. 2 shows tempo derived with $\sigma = 1.67$ for the solo/ensemble case. As it is clear from the plots, the individual tempo curves contract to the same tempo when the musicians play together.



(a) Beat duration for the first violinist solo vs ensemble.



(b) SQDR of solo vs ensemble

Fig. 3. Box-and-whisker diagrams showing some results in micro-tempo.

In the solo excerpts we found a relationship between the alternation of pauses and semiquavers of each single voice and the corresponding duration of the beat. The tempo was kept differently by the musicians in the case of a pause then in the case of semiquavers. Also in this case we found differences from the ensemble recordings case where the discrepancy between pause and semiquaver duration gets smaller because of the interdependence among musicians. The results of this analysis are shown in Fig. 3(a) by box-and-whisker diagrams. A t-test shows a significant difference between pauses and semiquavers duration in the cases of solo violin 1 and solo viola. In the ensemble case only the viola is found to play pauses consistently shorter than semiquavers although the difference was small[4]. Regarding the variances, a χ^2-test at a significance level of 5% could find disjoint confidence intervals only for the first violin in the solo and the cello in the ensemble. In the remaining cases the amount of variation in duration across the pauses did not differ from the one of the notes.

Fig. 3.1 shows the same tempo analysis for the joint ensemble performance. The most evident feature of this tempo curve is its relationship with the repeti-

[4] The difference between the mean duration of the four notes groups and that of the pauses was just 47 ms. This does not necessarily mean that the viola was not synchronized with the rest but it might mean that he was slightly anticipating the the others' first note onset and/or deferring the last note offset in the group of notes.

tion structure of the exercise. In fact, while remaining relatively constant (just slightly increasing through the performance) the performance tempo was indeed oscillating by speeding up in the center of the repetition and slowing down towards its boundaries. In table 3.1 the correlation with the pitch curve[5] is shown for all the cases. In the joint ensemble performance the correlation is 0.56. This confirms the overall tendency of the performance to speed up at higher pitches[6]. Despite the fact that we only have recoded few repetitions, this value of correlation is highly improbable to arise by chance. To quantify the significance we have used an empirical (Monte-Carlo) method. We generated sample random performances by perturbing in different ways the score with a gaussian noise to each note onset time. For each random performance we have performed the same macro-tempo analysis as the one performed on the real performance. Five groups of 2000 random performances were generated of respectively standard deviations 0.1, 2.5, 5, 10 and 25 ms. We then yield a value of pitch correlation for the tempo curve produced by each perturbed score and estimate variance and mean of the correlation. Assuming it to be a gaussian distribution we obtain an empirical p-value[7] for each noise amplitude. The resulting p-values are shown in Fig. 5 and, as you can see, the p-value is bounded by 2.5%. Remarkably, an increase of error variance σ^2 yields a decrease of p-values and not the other way around. Thus, it is even less likely to get a big correlation by adding a bigger noise than by adding a small one. In conclusion a confidence level of 98% can be considered to hold in all cases.

We have to notice that the excursion of the tempo curve is smaller than the just-noticeable-difference (JND). This means that the musicians are not aware of this fluctuations of tempo. Moreover, we can not still distinguish if this mechanism is directly related to repetition structure, pitch or to some more complex unconscious mechanism governing the performance.

	Solo		Ensemble		Joint Ensemble	
	Corr	Cov	Corr	Cov	Corr	Cov
Violin 1	-0.54	-33.65	0.59	2.87	0.56	2.78
Violin 2	0.5	8.5	0.75	3.34		
Viola	0.05	0.92	0.59	3.01		
Cello	0.72	6.1	0.35	1.01		

Fig. 4. Correlation of pitch with joint ensemble tempo curve.

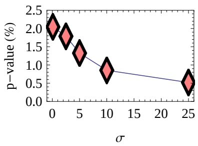


Fig. 5. p-value for different variances of gaussian noise for the empirical significance test.

[5] The pitch curve has values in number of semitones and has been constructed by taking the higher pitched note of each chord

[6] This is predicted by the well-known phrase-arch rule of Friberg et al. It is thus probably unrelated to pitch, and occurs only because the high pitches are in the middle of the phrase. However the performance we are analyzing here, far from being expressive, is just an exercise scale.

[7] The empirical probability of having a correlation as high as the one measured for the real performance (0.56).

3.2 Micro-tempo

Whereas macro-tempo can be related to global properties of the performance such as phrases or repetition patterns, micro-tempo is usually related to incidental local characteristics of the score.

At a shorter micro tempo scale, we found a consistent relation between the duration of each semiquaver and the position it occupies within groups of 4 semiquavers. An ANOVA test could confirm at a significance level lower than 1% the effect of metrical position on the joint-performance.

Differences have been found also when comparing the solo performance with the ensemble performance. Since the general tendency is to play the first note of the group longer than the second we have focused, for the sake of simplicity, on the ratio between the duration of the first semiquaver duration and the second (SQDR). This simplification also enables us to compare the solo case with the ensemble case since the ratio is not directly dependent on tempo. Remarkably, we could prove at a significance level lower than 2% the effect of the two scenarios to the SQDR for first Violin, Viola and Cello. We can thus report an overall tendency to exaggerate the agogic accent of consecutive strong-weak semiquaver couples in the ensemble case respect to the solo. Whereas the second violin keeps maintaining a positive SQDR of 1.19 in both the cases, the first violinist and the cello increase theirs from 1.07 to 1.27 and from 1.0 to 1.24 respectively. Despite this general tendency, a different behavior was measured for the viola which was decreasing its SQDR from 1.29 to 1.02.

A further analysis of the precedence of the onset times seems to explain the different micro-tempo results of the musicians in the ensemble case. Analyzing the attack time of the musicians having synchronized notes we found out that the attacks of the cello were preceding the ones of the viola by a mean of 8 ms, and the first violin was preceding the second by 13 ms. Musicians employing an higher SQDR are thus also anticipating their partner on the average. This suggest that the use of contrast in successive notes could be used as a mean of communication between the musicians to better control the synchronization.

4 Discussion and future work

We have presented an experimental framework through which we assign the musicians of a string quartet the task of playing specifically chosen exercises after a brief rehearsal period. In this context we have shown a set of preliminary results on timing synchronization phenomena observing the differences between musicians playing alone or in ensemble.

In the macro-tempo and at the beat level we have observed broad reduction of the mean bpm and its total variation in each single instrument. This confirms the hypothesis that the constraints that musicians are required to follow end in favoring a more controlled execution. In the joint ensemble performance we have then detected a consistent correlation of the bpm with the phrase structure of the repetition. Despite the fact that the excursion was here within the JND for tempo changes we have shown that this behavior is unlikely to happen by chance. However, more experiments should be carried out to check if this behavior arises because of the repetition structure, because of the pitch contour or for more complex reasons.

The analysis micro-tempo, on the other hand, was pointing out generally a bigger variance between short contiguous notes in the ensemble than in the solo. By also looking at the precedence of onset attack time between musicians we have formulated the hypothesis that a bigger contrast between contiguous short note duration might be used by leaders to maximize the communication with the other musicians or improve the synchronization. This hypothesis should be taken into account systematically to design further experiments.

References

1. J. Beran and G. Mazzola. Analyzing musical structure and performance-a statistical approach. *Statistical Science*, 14(1):pp. 47–79, 1999.
2. D. Deutsch, editor. *The psychology of music.* Academic Press, 2nd edition, 1998.
3. A. Friberg. Generative rules for music performance: A formal description of a rule system. *Computer Music Journal*, 15(2):56–71, 1991.
4. A. Gabrielsson. *The performance of Music.* Academic Press., 1999.
5. W. Goebl and C. Palmer. Synchronization of timing and motion among performing musicians. *Music Perception*, 26(5):427–438, 2009.
6. C. Keil. Participatory discrepancies and the power of music. *Cultural Anthropology*, 2(3):275–283, 1987.
7. P. Keller. Joint action in music performance. *Emerging Communication*, 10:205, 2008.
8. E. Maestre. *Modeling instrumental gestures: an analysis/synthesis framework for violin bowing.* PhD thesis, Universitat Pompeu Fabra, Barcelona, Spain, Novembre 2009.
9. G. P. Moore and J. Chen. Timings and interactions of skilled musicians. *Biol. Cybern.*, 103:401–414, November 2010.
10. J. A. Prögler. Searching for swing: Participatory discrepancies in the jazz rhythm section. *Ethnomusicology*, 39(1):pp. 21–54, 1995.
11. H. Purwins and D. R. Hardoon. Trends and perspectives in music cognition research and technology. *Connection Science*, 21:85–88, June 2009.
12. R. Ramirez, A. Hazan, E. Maestre, and X. Serra. A data mining approach to expressive music performance modeling. *Multimedia Data Mining and Knowledge Discovery*, pages 362—380, 2007.
13. N. Todd. The dynamics of dynamics: A model of musical expression. *The Journal of the Acoustical Society of America*, 91(6):3540–3550, 1992.
14. G. Widmer and W. Goebl. Computational models of expressive music performance: The state of the art. *Journal of New Music Research*, 33(3):203–216, 2004.
15. G. Widmer and A. Tobudic. Playing mozart by analogy: Learning multi-level timing and dynamics strategies. *Journal of New Music Research*, 32(3):259–268, 2003.