SYNCHRONIZATION OF INTONATION ADJUSTMENTS IN VIOLIN DUETS: TOWARDS AN OBJECTIVE EVALUATION OF MUSICAL INTERACTION

Panagiotis Papiotis, Esteban Maestre, Marco Marchini, Alfonso Perez

Music Technology Group,
Universitat Pompeu Fabra
Barcelona, Spain
{panos.papiotis, esteban.maestre, marco.marchini, alfonso.perez}@upf.edu

ABSTRACT

In ensemble music performance, such as a string quartet or duet, the musicians interact and influence each other's performance via a multitude of parameters – including tempo, dynamics, articulation of musical phrases and, depending on the type of instrument, intonation. This paper presents our ongoing research on the effect of interaction between violinists, in terms of intonation. We base our analysis on a series of experiments with professional as well as amateur musicians playing in duet and solo experimental set-ups, and then apply a series of interdependence measures on each violinist's pitch deviations from the score. Our results show that while it is possible to, solely based on intonation, distinguish between solo and duet performances for simple cases, there is a multitude of underlying factors that need to be analyzed before these techniques can be applied to more complex pieces and/or non-experimental situations.

1. INTRODUCTION

The study of music performance is a research field that is thoroughly multidisciplinary, combining elements from signal processing, computational musicology, pattern recognition, and artificial intelligence. The largest corpus of work within this field approaches the problem of performance analysis as follows: given the score for a musical piece P and a recorded performance of that piece E_p, the goal is to measure the deviations between P and E_p in terms of *timing* (onset times and note durations), *dynamics*, *articulation* and, based on the instrument type, *intonation*. Almost all existing approaches deal with the performance aspects of a single musician, and a thorough state of the art can be found in [1].

For the case of musical ensembles,such an analysis can be performed on two different levels: the *intrapersonal* level, where each musician is individually interpreting his/her own score, and the *interpersonal* level, where each musician provides and receives audiovisual feedback to/from other musicians, thus establishing a causal relationship between the performances.

This causal relationship and the influence mechanism that drives it can be perceived as an achieved *synchrony* between the musicians; taking the case of musical timing as an example, synchronized tempo curves and onset times would strongly indicate a joint performance, as opposed to each musician performing alone and subsequently joining the recordings.

In this work, we focus on the synchronization of intonation adjustments in violins: since the violin is a fretless instrument and therefore capable of producing continuous pitch, violinists performing in a duet or larger ensemble must work together to achieve harmonic consonance for the overall sound.

1.1. Objectives

In studying the interaction mechanism that results in the synchronization of intonation adjustments, we focus on three main objectives:

- *Detecting* evidence of synchronization, as well as measuring its strength
- *Analyzing* the cause and behavior of the synchronization mechanism
- *Simulating* the mechanism by means of a computational model.

Through the achievement of these objectives, we hope to contribute in two different ways: first, by detecting and quantifying the interaction between musicians as they perform – which has potential applications both in analyzing collaborative performances as well as aiding in their realization. Second, by measuring and simulating the interaction – which can be used to synthesize audio from separate, intelligent musical agents which are aware of, and capable of adapting to each other's expressive choices in order to better control the quality of the joint acoustic result.

In this article, we investigate towards the achievement of the first objective; to detect and quantify the synchronization mechanism, specifically for the case of violin intonation. To this end, our approach was to record violinists performing their part in an interactive set-up (i.e. together) and then separately; then, using each musician's score as a reference for the expected pitch, we calculate the deviation of the performed pitch contour from the score. Finally, we attempt to measure the coupling between the pitch deviation of violinist 1 and 2, using various interdependence measures.

1.2. Related work

Regarding synchronization in musical ensembles, an important amount of the existing work has been done from a cognitive point of view, such as Keller's theory of joint action in music performance[2].Keller focuses on three processes: *auditory imagery*, where the musician has his/her own anticipation of their own sound as well as the overall sound of the ensemble, *prioritized integrative attention*, where the musician divides his/her attention between their own actions and the actions of others, and *adaptive timing*, where the musician adjusts the performance to maintain temporal synchrony. The final process, essentially an error correction model where discrepancies between timing representations are detected, has its mathematical foundation in phase and period synchronizations; however, the main focus of this work is mainly theoretical and not immediately usable as groundwork for computational approaches.

A more practical approach on musical synchronisation can be found in [3], where the bowing gestures of two members of a skilled string quartet are studied, revealing evidence of interactive coupling in the synchronization of their movement along with a high precision in execution. Similar to Keller's approach on timing, measures of synchronization are based on approaching the musician's temporal behavior as an oscillating system, where the two musicians are coupled with each other. Another approach, similar in context but dealing more with the concept of dominance in social interaction can be found in [4].

Regarding intonation and melodic features in general, Kahlin's research[5] focused on formant frequencies in singing voice, and specifically on the effect of singing solo versus singing in a barbershop quartet, reporting that the singers strive to separate their own formants from the others; given the importance of accurate intonation in such as scenario, it is considered likely that the singers spread their formants in order to hear themselves better, which facilitates intonation.

On the specific subject of synchrony in intonation adjustments, we could not find any literature that has investigated the matter from a computational point of view. For the specific case of the violin however, there exist musicological approaches [6] and handbooks directed to violinists which discuss the hypothesis that such adjustments exist.

1.3. Outline

The remainder of this article is organized as follows: In section 2, we describe our signal acquisition process, as well as the pre-processing steps that we perform prior to our analysis; section 3 provides details about our choice for the interdependence measures; section 4 contains some experimental results of our analysis. Finally, in section 5, we discuss the results and provide some conclusions and our future directions.

2. SIGNAL ACQUISITION & PRE-PROCESSING

We conducted two series of experiments; the first round of experiments featured two professional violinists who have previous experience in performing as a duet and were familiar with the scores they were performing, while the second round of experiments featured two amateur violinists who had no experience of performing together, and without previous knowledge of the scores.

2.1. Recordings

Each piece was recorded in two discrete set-ups:

- a *solo* set-up, where each musician performed their part alone, and
- a *normal* set-up, where the musicians performed their respective part together as in a normal duet situation.

In order to reduce the complexity of the required task as well as motivate the musicians to focus on the performance unrestricted, the recordings were carried out without the use of a metronome. The pieces performed by the professional musicians were select excerpts from J.S. Bach's *Concerto for two violins* (BWV 1043) and L. Berio's *Duetti per due violini.* For the case of the amateur musicians, we opted for scores of much simpler difficulty, and thus the pieces used were the traditional piece *Greensleeves* played in unison by the two violinists, and a simplified excerpt from L. Berio's *Duetti per due violini*.

Each violinist was captured using piezoelectric pickups fitted on the bridge of the violin, while a large diaphragm condenser microphone captured the overall sound of the duet; all audio signals were captured with a sampling rate of 44100 hertz. Besides audio signals, we also recorded bowing gesture parameters using an EMF motion tracking device, using the method detailed in [7]; these signals were used to perform the audio-to-score alignment, as it is shown in the next section.

2.2. Score-performance alignment

In order to have a reference to which the intonation adjustments can be compared, it was necessary to align each performance to its respective score; this way, the score can be used as the 'expected' pitch, and the difference between this and the recorded pitch can be viewed as a score-free representation of the intonation. However, it is known that score-performance alignment is a difficult task, especially for the case of a continuous-excitation instrument such as the violin where the note onsets are varied and smooth.

Utilizing the captured bowing gestures as well as the audio information as input, we used an implementation based on the method described in [9],[10]. In this method, bow direction changes as well as more subtle measurements such as an estimation of the applied bow force provide the most probable candidates for note changes, combined with information extracted from the audio (such as the fundamental frequency curve and the root mean square energy of the recorded sound). These features are given as input to an implementation of the Viterbi algorithm, which calculates the temporal alignment between the score and the recorded performance.

2.3. Temporal alignment of experimental recordings

Since the recordings in the experimental set-ups were performed without a metronome, it was necessary to time-warp the performances in order to compare pitch deviations between violinists 1 and 2 without regarding timing information; in the solo recordings, for example, this comparison is impossible since the two recordings of violinists 1 and 2 were not temporally synchronized.

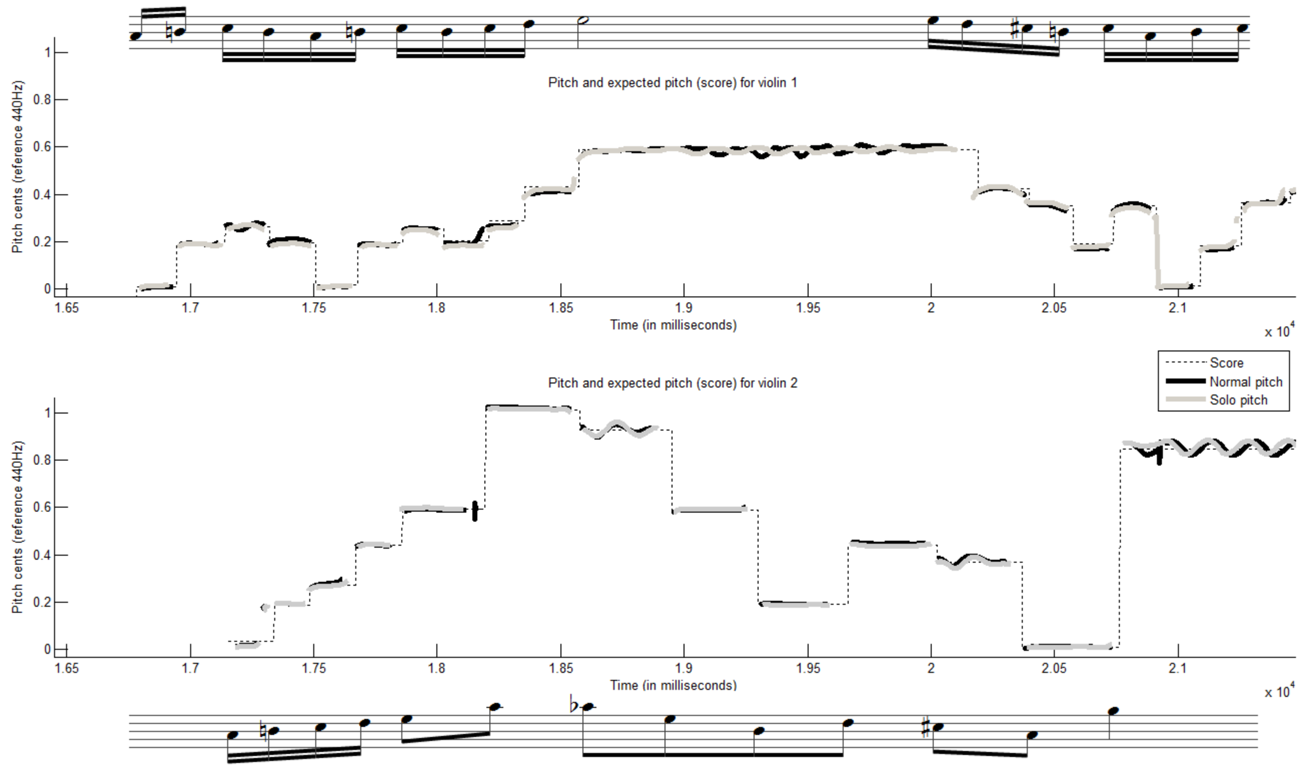

Figure 1: Expected pitch (score), recorded pitch in the normal set-up, and recorded pitch in the solo set-up, for an excerpt of J.S. Bach's 'Concerto for two violins'

This was achieved by applying a note-by-note temporal warping algorithm based on resampling the signal between note onsets and restoring its original pitch using an implementation of the phase vocoder algorithm, as described in [8]. We initially considered warping only the pitch contours instead of the recorded audio; however, besides producing an accurate and non-destructive temporal alignment (as seen in Figure 2) , this approach can be also very useful in performing user evaluation tests, where subjects can hear the normal duet recordings and the solo aligned-duet recordings and rate the quality of the duet's intonation - thus investigating whether intonation adjustments alone (i.e. with no temporal mismatch) can provide enough information to discriminate between solo and duet recordings. For this reason, all recordings were warped to match the onset timings of the normal recordings, in order to preserve the natural timing of a joint performance.

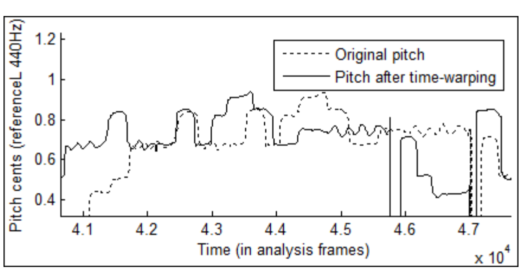

Figure 2: Extracted pitch before time-warping and after time-warping the source sound.

2.4. Audio feature extraction

The audio features that were used are the fundamental frequency contour, root mean square (RMS) energy and aperiodicity, all of which were extracted using the YIN algorithm[11]. The window size was set to 4096 samples, using a hop size of 32 samples. Both the F0 contour and the score were converted to pitch cents using 440 hertz as reference. The RMS energy and aperiodicity were used to filter out data points without a clear pitch content, while the octave errors produced from the pitch estimation were corrected using pitch guides; essentially an upper and lower bound for the F0 contour that is obtained by shifting the score-defined notes by a given threshold (±30 pitch cents). Finally, all time series were downsampled to a sampling rate of 1KHz, in order to facilitate the comparison of time series as well as reduce the computational load for the interdependence measures used afterwards. Figure 1 presents an example of the extracted, post-processed data.

2.5. Intonation adjustments extraction

In order to extract the intonation adjustments from the F0 contours, we consider the score as a reference or 'expected pitch', i.e. perfect, non-adjusted intonation. The deviation between the recorded pitch and this reference was initially used as the intonation adjustment. However, the score is translated to fundamental frequency using the equal temperament scale, as in keyboard or fretted instruments; this does not necessarily hold for the case of violin performance. Indeed, violinists make their choice of temperament based on a number of factors, such as the interval with the previous and next note, the string the note is being played on, as well as the type of instrument they are performing with.

This phenomenon was observed in our recordings as well; there were numerous notes in the score whose expected pitch had a systematic distance from the mean pitch performed by all violinists. Given that all measured pitch deviations varied within a very small range (within the range of 10 pitch cents), this systematically biased representation of the expected pitch introduced a lot of noise in our time series. We proceeded to bypass this problem as follows: for notes with a systematic distance larger than 5 pitch cents, we substituted the expected pitch with the mean F0 value of all performances for that particular note, in that particular piece. This choice was later validated by our results as increasing the separation between the *solo* and *normal* experimental set-ups.

3. INTERDEPENDENCE MEASURES

After extracting the pitch deviations for each violinist and each recording, the next step was to search for interdependence among the pitch deviations of violinist 1 and violinist 2. Several interdependence measures were considered and tested before we decided on which method to use. First we briefly discuss methods which did not provide useful, before finally describing the method we chose.

3.1. Preliminary interdependence measures

As a first indication to the nature of the possible interdependence between the two pitch deviations, we initially turned our attention to scatter plots between the values of the two time series. Figure 3 presents such a scatter plot, for the Greensleeves recording with the two violinists playing in *unison*, which is we considered as the simplest and clearest case of interdependence.

Linear and rank correlation. Naturally, the first obvious choice as a measure of interdependence was linear correlation. It can be observed from the scatter plots, however, that there is no visible correlation between the pitch deviations of the two violinists, and there is little difference between the solo and normal set-ups. The implications of such an observation lead us to the conclusion that, whichever synchronization phenomenon occurs is not consistent throughout the piece, and cannot be viewed as a process which is either invariant to time or cyclostationary.

Nevertheless, we calculated three correlation measures: product-moment (Pearson), Kendall and Tau correlation among the two time series. Unsurpriningly, the correlation coefficients had very low values for all experiments (≤[0.015] average), and failed to show statistically significant separation between the *solo* and *normal* experimental set-ups.

Mutual information. Another interdependence measure that was considered was mutual information, adapted for non-bivariate time series [12]. Mutual information is a dimensionless measure first applied to Information Theory, and loosely put, measures the difference between two types of joint entropy; the joint entropy of the two variables as measured from the data, and the joint entropy of the two variables as if they were independent.

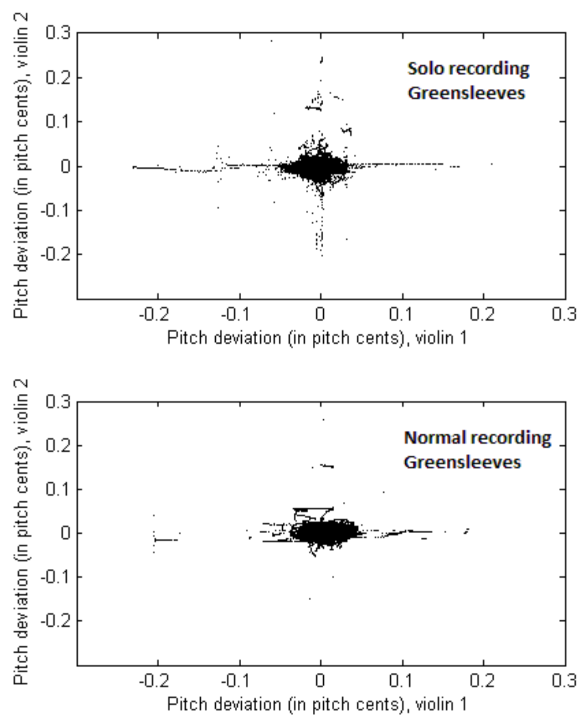

Figure 3: Scatter plots for Greensleeves between violinist 1 and 2, solo (top) and normal (bottom).

Thus, it quantifies the reduction in uncertainty about one random variable given knowledge of another. However, this measure too failed to provide separation between *normal* and *solo*, with the exception of the experiment shown in figure 3, where the violinists where playing the same melody.

Granger causality. One significant drawback of the previous measures is their lack of directionality; besides the overall degree of interdependence, it is also important to draw conclusions about the direction of influence, i.e. whether violinist 1 is influencing violinist 2 more than the opposite.

A measure that is capable of giving such an estimate is Granger causality [13], a statistical concept of causality that was first applied to Econometrics, and recently to Neuroscience. It poses the hypothesis that if variable X causes variable Y, then past values of X should significantly help in predicting future values of Y as opposed to simply using past values of Y to predict its own future, given that the data is normally distributed. The parameter that in most implementations has to be defined by the user is the maximum number of lags.

Although there are cases in auditory cognition where granger causality has been used as a measure of coupling[15], it soon became apparent that the nonlinearities of our time series were not the suitable input for this measure. The normalized causality value was very low (≤0.001) for all recordings and a large variety of lag values, while the separation was once again not consistent.

3.2. Non-linear coupling detection

For our final interdependence measure, we turned our attention towards methods which are widely used in computational neuroscience. There exists a variety of nonlinear interdependence measures that quantify the signature of directional couplings among two random processes, based on distances in reconstructed state spaces. Essentially, the dynamics of each time series are reconstructed using a given number of embedding dimension and a given time delay; then, a distance matrix is calculated by estimating the distance of each point from every other in the phase space. Finally, by evaluating distances of conditioned neighbors in the distance matrix, the directional coupling for the two variables is calculated.

Of these measures, we use the measure *L*, which was recently shown to be of higher sensitivity and specificity for directional couplings than previous approaches. For a more in depth explanation of the method as well as its mathematical formulation, we direct the reader to [14] where the method was originally introduced.

There are four main parameters that have to be given as an input:

- the *embedding dimension* (m), which is the number of past values to be included for the state space reconstruction
- the *time delay* parameter or *tau* (τ), which is the time delay in samples between each of the past values used, and
- the number of *nearest neighbors* (k), which is the number of closest points from the distance matrix to be used for the coupling calculation, and
- the *theiler window* (W), which is the range of points to be excluded from the distance matrix in order to discard too-close neighbors

Experimenting with the values of these parameters, it became evident that the most imporant ones where the embedding dimension (m) and the time delay (τ), since they were the ones who had the greatest impact on the outcome of the algorithm; the number of nearest neighbors was set to 3, and the theiler window to $2^*\tau$. From there on, we calculated the coupling strength between violinists 1 and 2 for each experimental recording for the following range of values for m and τ:

$$m = [2:20], \text{ and}$$
$$\tau = [10:360] \text{ milliseconds, with a step of 20 ms.}$$

The rationale behind the above ranges is fairly simple: 360 milliseconds were manually identified as the maximum vibrato period in our recordings, while 20 embedding dimensions is a commonly adopted upper limit for the computation of the nonlinear coupling.

Given the number of tested values for the embedded dimension and the time delay described above, we performed 304 calculations of *L* for each recorded experiment; this was done to help us achieve greater statistical significance in our results.

The value of *L* increases with the amount of coupling strength, and is normalized between 0 and 1; higher coupling for violinist 1 than violinist 2 indicates that V1 casts a stronger influence on V2. Finally, the average coupling is calculated as the mean value of both coupling values, one for each violinist.

4. RESULTS

The results are divided in two categories – one for the recording of the amateur musicians, and one for the professional and more complex recordings.

4.1. Coupling results for amateur musicians

In all our recordings with the amateur musicians, three main empirical observations were made:

1. Violinist 2, being less adept at *prima vista*, was more focused on performing the piece correctly rather than adjusting to violinist 1. As a consequence, violinist 1 was mainly trying to adapt his intonation to violinist 2.

2. Performing a piece in unison (as in the case of *Greensleeves*) naturally exposes the mismatch in intonation the most, since the harmonic dissonance is much more apparent when the two violinists are performing the same melody, in the same tonal height; this was the case where the interaction between musicians was most evident.

3. Performing a piece where the melodic line of violin 1 is different from violin 2, made the detection of harmonic dissonance much more difficult. The same stands for the tempo of the performed piece, where slow tempos exposed bad intonation, and fast tempos also made it difficult for the musicians to keep their attention on their partner, presumably because of the cognitive load of the performance.

In Figure 4, the coupling strength for all values of *m* and *tau* is displayed as a grayscale mesh. It can be seen that, although *m* and *tau* seem to increase the overall coupling strength, the coupling value stays consistent throughout the repetitions of the *L* measure.

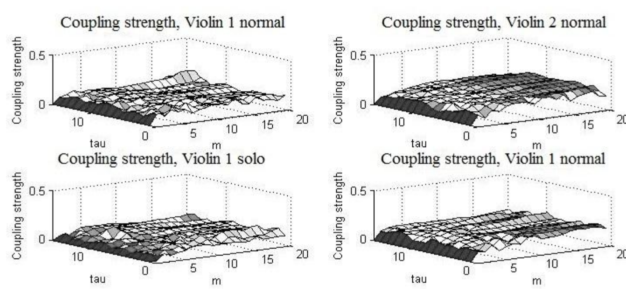

Figure 4: All calculated coupling strengths for the Berio duet recorded by the amateur violinists.

The average coupling strength for a given recording is given by taking the mean across all values of m and τ. Figure 5 shows the average coupling strength the Greensleeves recording:

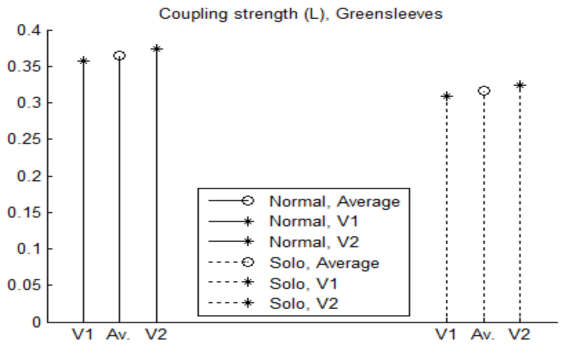

Figure 5: Overall coupling strength for *normal* and *solo* recordings of Greensleeves

Figure 6 shows the overall coupling strength for the Berio recording, featuring two repetitions of the *normal* set-up and two repetitions of the *solo* set-up:

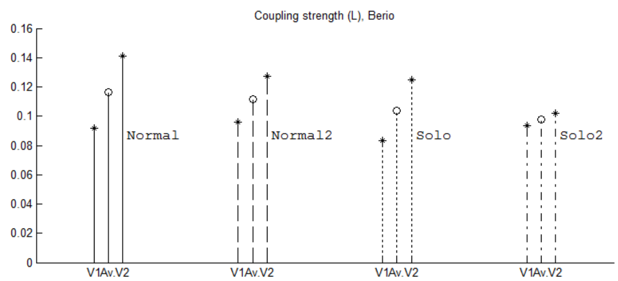

Figure 6: Overall coupling strength for two *normal* and two *solo* recordings of the L.Berio duet.

From the above calculated values, it can be seen that the coupling strength is capable of providing consistent separation between the *normal* and *solo* recordings, albeit with a small margin for more complex scores. Moreover, it is also seen that the coupling measure employed is capable of indicating the direction of the influence mechanism between the violinists; in both figures, violinist 2 has a stronger influence on violinist 1 than the opposite.

Moreover, it is observed that the coupling strength is significantly decreased when the musicians have more complex scores; in figure 5 the two violinists are performing the same score, while in figure 6 they are performing different melodic lines.

4.2. Coupling results for professional musicians

In the experiments with professional musicians, our main observation was that, since the musicians were already familiar with each other's playing, as well as the performed pieces, they could reproduce their intonation with remarkable accuracy throughout the recordings; thus shifting their attention more towards the timing and articulation aspects of the performance. This became particularly evident after listening to the time-warped recordings, where it was nearly impossible to distinguish between the *normal* recording and the *solo*. Evidence of this statement can be seen in one of the very first figures, Figure 2, where it is evident that the *solo* and *normal* recordings have remarkably similar F0 curves.

Figure 7 shows the overall coupling strength for the Bach recording, while figure 8 shows the overall coupling strength for the L.Berio recording.

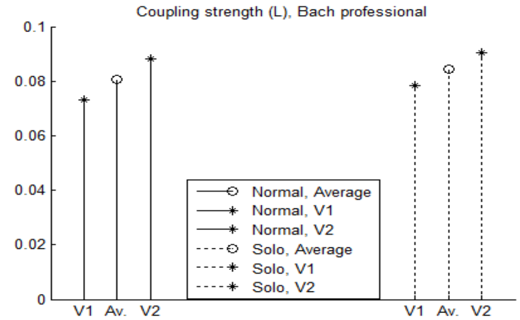

Figure 8: Overall coupling strength for *normal* and *solo* recordings of the J.S. Bach duet, with the professional musicians

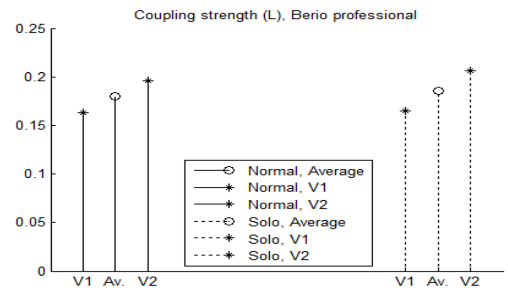

Figure 8: Overall coupling strength for *normal* and *solo* recordings of the L.Berio duet, with the professional musicians

It is evident that these cases are much harder to separate; this is attributed both to the skill of the violinists, as well as the complexity of the score. The direction of influence is consistent throughout all four recordings, although the tempo and complicated harmonic relationship between the two scores makes it difficult to validate as a result.

5. CONCLUSIONS & FUTURE WORK

In this paper we have presented a preliminary method for measuring the synchronization mechanism behind intonation adjustments in violin duets, based on the pitch deviation of each violinist from his/her respective score. An analysis procedure for the recorded material has been outlined, as well as some considerations on specific performance aspects of the violin. We tested a number of interdependence measures before concluding on the measure that provides the best results, and the final chosen measure appears to validate the hypothesis that violinists are influenced by each other's intonation when performing together, at least for the simple cases of non-professional musicians.

However, it has been seen that the coupling strength is dependent on a multitude of factors; namely the complexity of the piece, the skill of the violinists, and the harmonic relationship between the two performed melodic lines. In order to obtain clearer separation between solo and duet recordings and to study this synchronization phenomenon without coloration from the scores, it is necessary to push towards two main improvements.

First, we believe that the use of F0 contours as the only extracted feature does not convey the real phenomenon well enough, since the pitch perception of real instruments such as the violin is strongly related to psychoacoustics factors such as the loudness and timber of the produced sound. To this end, an objective measurement of harmonic consonance/dissonance has to be included as a feature, to approach more the human perception of pitch and intonation.

Second, in order to make the coupling detection independent from the performed scores, it is necessary to post-process the scores with an algorithm that analyzes the intervals between the two violins, and adjusts the expected pitch (or the lack thereof) according to the harmonicity of the interval; this way, very harmonic intervals between the two melodic lines will greatly penalize 'bad' intonation, while inharmonic intervals will contribute much less to the coupling strength.

6. ACKNOWLEDGEMENTS

We would like to thank Dr. Ralph Andrzejak for his valuable suggestions on interdependence measures, as well as his advice and help in adapting the *L*-measure algorithm specifically for our case. We would also like to thank the anonymous reviewers for their insightful comments.

Finally, the work presented on this document has been partially supported by the EU-FP7 ICT SIEMPRE project.

7. REFERENCES

[1] Widmer, G., & Goebl, W. : Computational Models of Expressive Music Performance: The State of the Art. Journal of New Music Research, 33(3), 203-216. (2004)

[2] Keller, P. E. : Joint action in music performance. Emerging Communication, 10. IOS press.(2005)

[3] Moore, G. P., & Chen, J. : Timings and interactions of skilled musicians. Biological cybernetics, 103(5), 401-14. (2007)

[4] Glowinski, D., Coletta, P., Volpe, G., Camurri, A., Chiorri, C. and Schenone, A. : Multi-scale entropy analysis of dominance in social creative activities. ACM International Conference on Multimedia, pp. 1035-1038. (2010)

[5] Kalin, G. : Formant Frequency Adjustment In Barbershop Quartet Singing. Doctoral dissertation, KTH, Department of Speech, Music and Hearing. (2005)

[6] Mason, J.A. : Comparison of Solo and Ensemble Performances with Reference to Pythagorean, Just, and Equi-Tempered Intonations, Journal of Research in Music Education Vol. 8, No. 1 pp. 31-38 (1960)

[7] Maestre, E., Bonada, J., Blaauw, M., Perez, A., & Guaus, E. : Acquisition of violin instrumental gestures using a commercial EMF device. International Computer Music Conference (p. 386–393). San Francisco (2007)

[8] Ellis, D.P.W. : A phase vocoder in MATLAB, http://www.ee.columbia.edu/~dpwe/resources/matlab/pvoc/, last retrieved on (March 2011)

[9] Maestre, E. : Modeling instrumental gestures: an analysis/synthesis framework for violin bowing. Doctoral dissertation, Univ. Pompeu Fabra, Barcelona, Spain. (2009)

[10] Perez, A : Enhancing Spectral Synthesis Techniques with Performance Gestures using the Violin as a Case Study. Doctoral dissertation, Univ. Pompeu Fabra, Barcelona, Spain. (2009)

[11] De Cheveigné, A., & Kawahara, H. : YIN, a fundamental frequency estimator for speech and music. Journal of the Acoustical Society of America, 111(4), 1917-1930. (2002)

[12] Dionisio, A., Menezes, R., & Mendes, D. A. : Mutual information: a measure of dependency for nonlinear time series. Physica A: Statistical Mechanics and its Applications, 344(1-2), 326-329. (2004)

[13] Granger, C.W.J. : Investigating causal relations by econometric models and cross-spectral methods". Econometrica 37 (3), 424–438. (1969)

[14] Chicharro D., Andrzejak R.G. : Reliable detection of directional couplings using rank statistics. Physical Review E, 80, 026217 (2009)S. Haykin, *Adaptive Filter Theory*, Prentice Hall, Englewood Cliffs, NJ, USA, second edition, (1991)

[15] Londei, A., D'Ausilio, A., Basso, D., Sestieri, C., Del Gratta, C., Romani, G. L., et al. Brain network for passive word listening as evaluated with ICA and Granger causality. *Brain Research Bulletin*, *72*(4-6), 284-292. (2007)