# Evaluation of subjective impression of instrument blending in a string ensemble

Jithin Thilakan, Malte Kob
*Erich Thienhaus Institute, Detmold University of Music, 32756 Detmold, Germany, Email:*
*jithin.thilakan@hfm-detmold.de, malte.kob@hfm-detmold.de*

## Introduction

Blending of sound sources is an important aspect in situations such as music composition, orchestration, music recording techniques, room acoustic adaptation, etc [1, 2, 3]. The term 'Blending' refers to the perceptual fusion of two or more concurrent sounds by losing their individual distinctiveness [4, 5]. The impression of blending is observed to be a multi-dimensional phenomenon since it is influenced by different level factors such as the music composition and performance related characteristics, acoustic features of the performance space, the location and orientation of the musical instruments, hearing skill and experience of the listener, etc. In a joint musical performance, different levels of blending impressions between instruments are sought after by composers, musicians, recording engineers, and listeners according to their requirements. At the same time, due to the acoustic and musical factors involved, each musician/listener experiences a different blending impression as well.

The impression of blending can be analyzed in macroscopic (as a collective impact on listeners), and microscopic (as a detailed analysis of minute level variations) levels in both spectral and temporal domains. As a joint work between ACTOR [6] and VRACE [7] projects, the third ODESSA project recording and a macroscopic analysis of blending using a live listening test were conducted at Detmold Concert House with a string ensemble consisting of 9 violins and 18 listeners. Previous studies show that the acoustic features of concert halls, and the distribution and orientation of individual sources on stage significantly influence the resultant sound field of an ensemble and thereby influence the impression of blending [8, 9]. In addition to it, the coordinated action and joint strategies between musicians and the room acoustic feedback received by them are also observed to have an impact on their performance and therefore influence the resultant sound field of orchestra performance [10, 11, 12]. Hence, variations in room acoustic conditions, listeners' locations, and source positions have been diversified in this test to understand the influence of these factors on the perception of blending.

Furthermore, an initial step to evaluate blending at the source level (excluding the influence of acoustic environment) using signal analysis is performed on the recordings from the spot microphones attached to the violins. Potential signal and musical features that can predict the blending impressions are estimated from the signal analysis and validated against a pilot listening test conducted on selected sound samples.

## Methodology

### Listening test

The live listening test was performed at the Detmold Concert House with a string ensemble consisting of 9 violinists, and 18 test participants. The musicians who participated in the string ensemble were the students of the Detmold University of Music. Among the 18 test participants (7 female, 11 male), 14 of them had musical background with experience in playing musical instruments. To make the test participants aware of the objective of the test, the overall goal of the experiment, definition of blending and its perception aspects were explained to them at the beginning of the test. After that, the musicians were asked to play the musical pieces as a group of violins starting from 1 to 9 with an increment of 1 violin at each time. As a result, musicians got familiar with the piece and adapted with the joint performance with others whereas the test participants were able to develop an idea of the overall sounding impression with the increase in the number of violins.

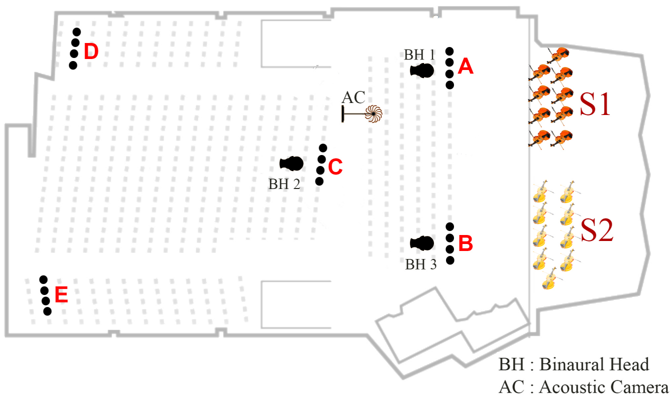


**Figure 1:** The sketch of the seating plan in Detmold Concert House.

The listening test was performed in two parts; in the first part, the listeners were advised to sit in the prescribed seats in the predefined locations of the Concert House (denoted as A, B, C, D, and E in Figure 1), and in the second part, the listeners were free to choose seats according to their individual preferences. In each part of the test, the acoustic conditions of the concert hall have been varied using the Wave Field Synthesis (WFS) system installed at Detmold Concert House. It includes natural acoustic conditions (no WFS) of 1.6 seconds reverberation and longer reverberation times (with WFS) of 2.3 seconds and 5 seconds in the first part and 2.3 seconds and 3.2 seconds in the second part. In addition to the artificial reverberation, the test conditions are varied by changing the seating arrangement between German (Violins are mainly radiated towards the listeners; denoted as S1 in Figure 1) and American (Violins are mainly radiated towards the rear wall of stage; denoted as S2 in Figure 1) seating. For each acoustic condition, an ensemble of violins with a variable number of players (1, 2, 3, 4, 6, and 9) played at random order. The test participants were requested to listen with their eyes closed and rate the sound after the music is stopped by giving their responses on the form. A conductor on the stage arbitrarily decided the number of violins needed to play in each take. For one particular take, two musical pieces were played in a sequence which are Symphony No. 6 in B minor, I. Adagio – Allegro non troppo by Tchaikovsky, and Sonata No. 12 in A flat major, II. Scherzo by Beethoven that altogether took almost 60 seconds to complete.

For each take, the listeners were asked to predict the number of violins played, and also to mention whether it sounded like an ensemble or not (yes or no). As per definition, the increase in the confusion in identifying the correct number of players in the ensemble corresponds to a higher degree of blending. After the performances in the 4 acoustic conditions, the listeners were able to choose their favorite seats in the Concert House and then the same process is repeated in this second part. The test altogether took 2 hours to complete with a small break in between parts 1 and 2.

### Instrument recordings

As mentioned earlier, considering the influence of factors such as the coordinated action and joint strategies between musicians, room acoustic feedback on musicians, etc., on the resultant sound field of an ensemble, the finest way to obtain the source signals of instruments to evaluate the blending between them in realistic cases would be to record the sources individually in a joint performance. Hence, in this study, the individual violins in the string ensemble are recorded using individual 'DPA 4099 Core Violin' clip-on microphones attached to the body of the instrument. A disadvantage of this particular approach would be the presence of room reflections (although it was found to be minimal in these microphone recordings), cross-talk between other instruments, and noise caused by the musician (breathing noise, scratching of the bow, etc.). In addition to the individual recordings, the performance of the ensemble was captured using stereo pair microphones, an acoustic camera, three binaural heads, and a portable binaural headphone recorder which were kept at different locations of the Concert House as given in Figure 1.

## Result and discussion

### Listening test

The overall ability of listeners in the prediction of the number of violins in all acoustic environments is shown in Figure 2. In general, the listeners tend to lose the ability to predict the correct number of violins with the increase in the number of violins played. This is in agreement with earlier findings on a similar problem in which the listener's ability to predict the number of sources is analyzed for different microphone recording techniques [2]. Whereas the performances of 1 and 2 violins were predicted very well with some outlier points, from 4 violins onwards a high variation in the predictions among listeners is observed. It should be noted that since the listeners were already biased as they knew about the maximum number of violins in this live listening test, the prediction of values above 9 would not be expected.

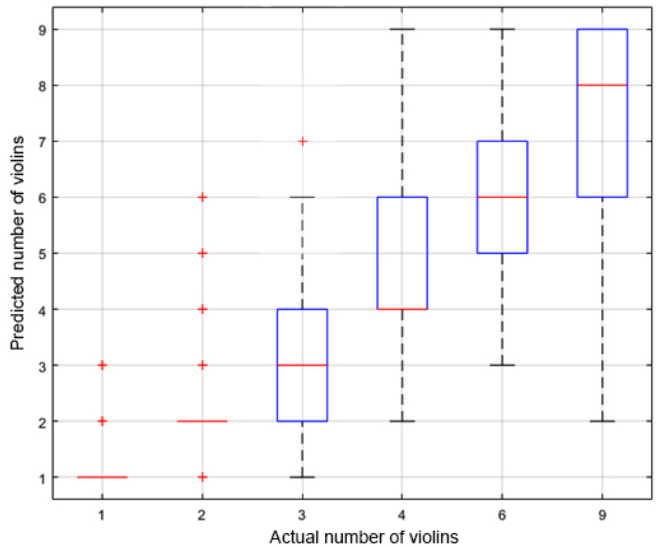


**Figure 2:** Listeners' prediction of the number of violins with the actual number of violins played

The variation of the prediction of the number of violins with the different acoustic environments is shown in Figure 3. When the artificial reverberation is introduced, the listeners tend to lose the ability to predict the correct number of violins. Interestingly, it is observed that the listeners were not able to perceive the difference between 6 and 9 violins in a reverberant acoustic environment. This shows that the blending is positively correlated with the reverberation of the acoustic environment, and supports the results from previous studies [9].

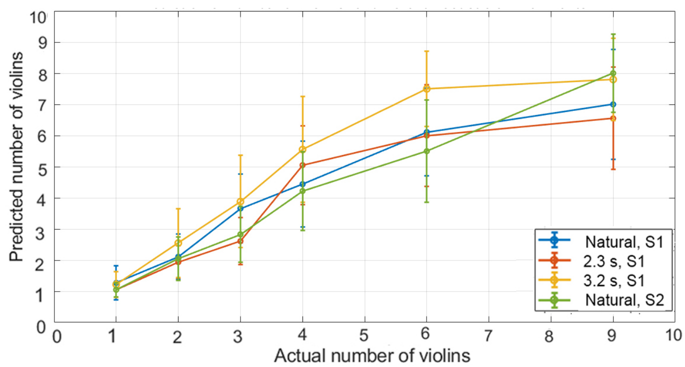


**Figure 3:** Variation in the prediction of number of violins with different acoustic environments

Figure 4 shows the overall variation in the prediction of the number of violins with different seating locations in the Concert House for different acoustic environments.

In this case, the listeners in the far locations (location D and E) tend to lose the ability to predict the number of violins played in comparison to the listeners in the near field where the direct sound from instruments have a strong influence. For low numbers of violins, the listeners at far locations are observed to overestimate, and for a high number of violins, they underestimate. Although the effect is minimal, this behavior positively correlates with the impression of blending.

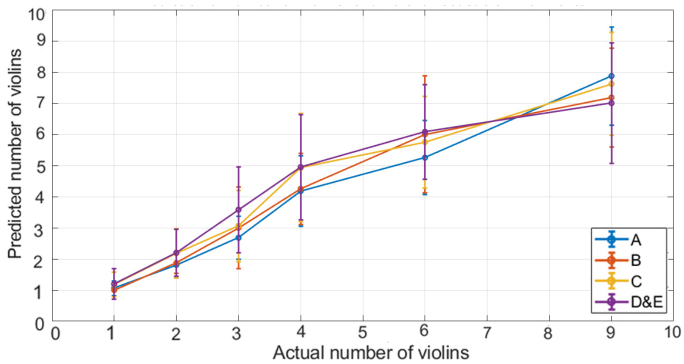


**Figure 4:** Variation in the prediction of number of violins with seating locations

As a more generalized approach, the responses of the same 12 participants who listened to the performance in the near-field (seats in front rows) and far-field (seats in back rows) locations in the two parts of test are analyzed. Figure 5 shows the percentage of correctness in the prediction of violins for the above-mentioned 12 participants who listened to the ensemble in the near-field with natural acoustics (denoted as 'direct sound field' with more influence of the direct sound from the instruments), and in the far-field with artificial reverberation (denoted as 'diffused sound field'). A significant difference in the percentage of correct predictions is observed between the direct and diffuse field, especially for 1 violin where it drops from 100 % correctness in the near-field to around 60 % in the diffuse field. In both cases, the percentage of correctness is observed to drop with an increase in the number of violins (as mentioned earlier, listeners in this test were already biased in the case of 9 violins). This clearly indicates that the acoustic environment plays an important role in the listener's ability to predict the number of violins and thereby influences the blending.

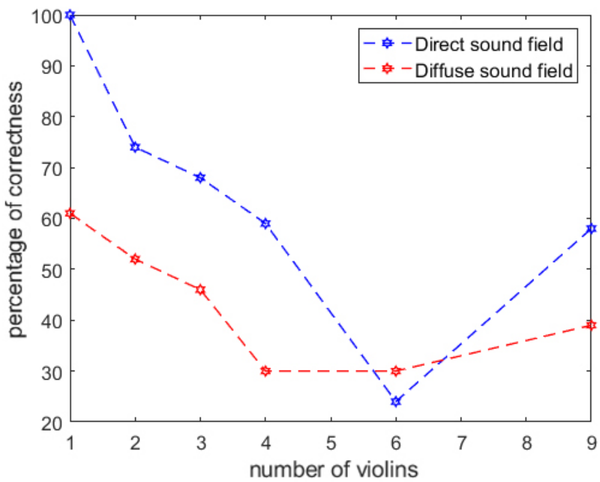


**Figure 5:** Variation in prediction in direct and diffuse sound fields

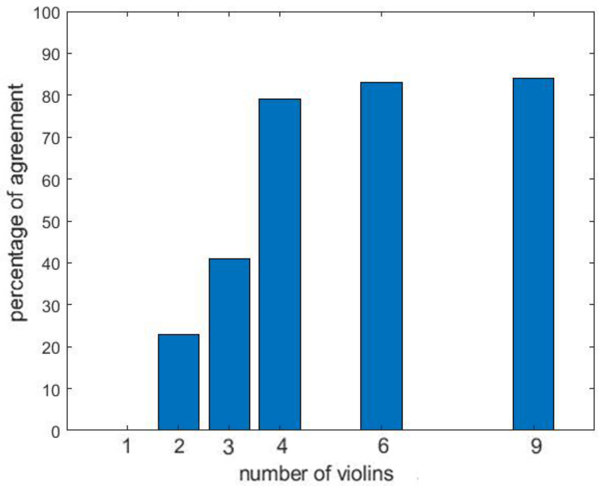


**Figure 6:** Agreement to the ensemble sound impression for different number of violins

Figure 6 describes the variation in the acceptance of the ensemble sounding impression for the different numbers of violins. For the question 'Does it sound like an ensemble or not?', participants gave a high agreement for ensemble sound impression from 4 violins onwards. Among different acoustic environments, this trend remained to be nearly the same.

## Signal analysis

Audio samples for blending evaluation are extracted from the spot-microphone recordings. A smooth high pass filter centered around 200 Hz is applied to reduce the breathing and bowing noise from the player. These samples had a minimal room acoustic contribution. Two pairs of violin signals, one with a high blending impression, and one with a poor blending impression playing the same musical piece are generated in two ways; (a) extracting 2 pairs from the recording, (b) altering one pair by introducing time delay, pitch and spectral variations, etc., in a controlled way to make the second pair. These violin samples are later evaluated using a pilot listening test.

As mentioned earlier, blending is a multidimensional phenomenon and hence the sound samples can differ in the blending impression due to many factors (eg. dissimilarity in timbre, asynchronous transients, difference in pitch values and individual loudness, etc.). From the evaluations on these samples, a set of potential audio and musical parameters (from spectral, temporal, and amplitude domains) are evaluated which can contribute to the prediction of blending between samples of different blending characteristics. Time windowed correlation of different parameters such as pitch values, the Mel spectrogram values, loudness envelope, spectral centroid, etc., is shown to have an influence on estimation of source-level blending among different kinds of samples.

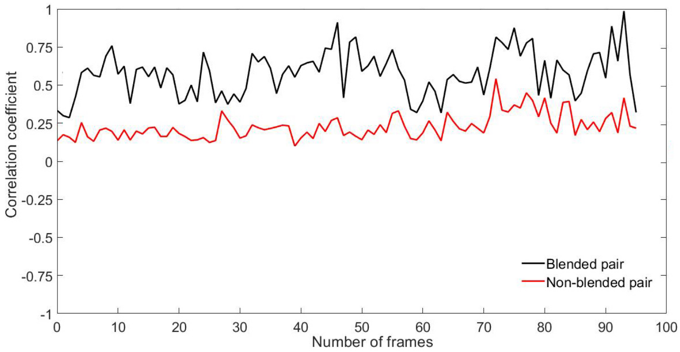


**Figure 7:** Moving window correlation of Mel scale spectrum values for blended and non-blended violin pairs

As an example, a moving window correlation of the Mel spectrogram values (that is more associated with human perception of sound spectrum) of individual channels for a time window of 100 ms, and overlap of 50 ms is used to characterize blending between instruments in the spectral dimension. The Mel spectrogram correlation for blended pair, and an altered non-blended pair of violin signals playing the same musical piece is given in Figure 7. A high correlation between source signals generally contributes to a better blending impression.

## Conclusion

The results of the perceptual evaluation of blending show that the number of sound sources and the acoustic environment possesses high importance in the resultant impression of blending. The listeners ability to predict number of violins tend to lose with an increase in the sources, and this is positively correlated with the increase in the overall blending impression of an ensemble. In addition, the increased reverberation time and seating distance favor this effect. In this test, 4 violins seem to be a critical limit at which a high variance in the prediction of the number of violins played, and a high agreement to the ensemble sound impression are observed.

From analysing the dry recordings, a set of potential parameters to predict blending in different cases are estimated. By considering the multidimensional aspects of blending, a future goal would be to formulate relevant parameters and combine them with proper weighting to predict the blending at the source level.

## Acknowledgement

The authors thank the listening test participants for their kind cooperation. Thanks to Stefanos Ioannou, Boris Alex Bolles, and Dustin Eddy for support in the recording. The authors acknowledge the ODESSA project participants, a collaborative project in the context of the ACTOR project partnership. This work has been performed within the project VRACE (Virtual Reality Audio for Cyber Environments) which has received funding from the European Union's Horizon 2020 research and innovation program under the Marie Sklodowska Curie actions (grant agreement number 812719).

## References

[1] R. A. Kendall and E. C. Carterette, "Identification and blend of timbres as a basis for orchestration," *Contemporary Music Review*, vol. 9, no. 1-2, pp. 51–67, 1993.

[2] S. Ioannou and M. Kob, "Investigation of the blending of sound in a string ensemble," no. 1. Proceedings of ISMA 2019, 2019, pp. 42–49. [Online]. Available: http://pub.dega-akustik.de/ISMA2019/data/articles/000093.pdf

[3] M. D. Francisco, M. Kob, J.-F. Rivest, and C. Traube, "ODESSA - Orchestral Distribution Effects in Sound , Space and Acoustics : An interdisciplinary symphonic recording for the study of orchestral sound blending," in *Proceedings of ISMA 2019*, 2019, pp. 33–41. [Online]. Available: http://pub.dega-akustik.de/ISMA2019/data/articles/000090.pdf

[4] G. J. Sandell, "Roles for Spectral Centroid and Other Factors in Determining 'Blended' Instrument Pairings in Orchestration," *Music Perception*, vol. 13, no. 2, pp. 209–246, 1995.

[5] S.-A. Lembke, "When timbre blends musically: perception and acoustics underlying orchestration and performance," Ph.D. dissertation, McGill University Montreal, Canada, 2015.

[6] "Analysis, Creation, and Teaching of Orchestration (ACTOR) project." [Online]. Available: www.actorproject.org

[7] "VRACE - Virtual Reality Audio for Cyber Environments project." [Online]. Available: www.vrace-etn.eu

[8] J. Meyer, *Acoustics and the Performance of Music*. Springer, 2009.

[9] D. Lincke, "Instrumental Blending in Concert Halls," Technische UniversitÂ¨at Berlin Institut, Tech. Rep., 2020. [Online]. Available: www.ak.tu-berlin.de/menue/master_theses/

[10] S. A. Lembke, S. Levine, and S. McaDams, "Blending between bassoon and horn players: An analysis of timbral adjustments during musical performance," *Music Perception*, vol. 35, no. 2, pp. 144–164, 2017.

[11] S. V. Amengual, M. Kob, and T. Lokki, "Analysis of trumpet performance adjustments due to room acoustics Analysis of trumpet performance adjustments due to room acoustics," in *Proceedings of the International Symposium on Room Acoustics*, 2019.

[12] S. V. Amengual, W. Lachenmayr, and M. Kob, "Study on the influence of acoustics on organ playing using room enhancement," in *Proceedings of Third Vienna Talk on Music Acoustics*, 2015.