PROCEEDINGS of ISMA 2019
International Symposium on Music Acoustics

13 - 17 September 2019 in Detmold, Germany

Investigation of the blending of sound in a string ensemble

Stefanos IOANNOU[1], Malte KOB

(1) HfM Detmold, Germany, stefanos@primehome.com
(2) HfM Detmold, Germany, kob@hfm-detmold.de

Abstract
Ensemble sound is a topic where, to date, only little research has been conducted. A fundamental question that still needs to be answered is how many violins are required for a number of violins to blend together in an ensemble.

As part oft the ACTOR project, A violin phrase from Tchaikowsky's 6th Symphony has been recorded with one, two, three, four, six and ten violins over piezo microphones, spot microphones, a decca tree, room microphones and a dummy head. These recordings were used to create a listening test, in which participants had to guess the number of violins playing and state if the recording sounded like an ensemble. Results of this test will be presented.

1 INTRODUCTION

When orchestras were first formed, they were far smaller than the typical modern orchestra. In 1730, J. S. Bach submitted his request for his church orchestra: 3 first violins, 3 second violins, 4 violas, 2 violoncellos, 1 double bass, 6 woodwinds, 3 trumpets and timpani. The first bigger orchestra was the Mannheim court orchestra, which in 1756 could boast with a 10.10.4.4.2. string section. This, however, was far from the norm. Most orchestras of the classical period had around 10 to 20 string players. In the course of the 19th century, orchestras with at least 10 first violins became common. Up to now this is the standard orchestra size for symphonic works (1). Reasons for this increase may vary, but it might be attributed to more wind instruments, such as clarinets and trombones, being used which caused the need for more strings or the tendency to build bigger concert halls, which required more instruments to achieve a full sound.

A string ensemble is extremely complex. It can be considered a number of incoherent sound sources, meaning that no matter how good the musicians are, no static phase relationship will be achieved between two instruments. Moreover, the distribution of the instruments on stage makes it difficult to predict the ensembles' timbre and radiation. Even the individual sound of the instruments is important for the sound of the ensemble as a whole. In a 2012 demonstration in Stuttgart, a violinist recorded a phrase of music multiple times using overdubbing. The resulting sound failed to create the illusion of a real ensemble. When the experiment was repeated using different violins each time, the result was much more natural sounding (1).

Due to the incoherency of the sound sources and the different distances to the microphones, the listener will localize the ensemble around the instrument nearest to him. The normal critical distance (defined as the distance from a sound source where the sound pressure of the direct sound and the reverberant sound are equal) is of little importance concerning the perceived relationship between direct and reverberant sound. A better measure would be the ratio of the direct sound to the closest instrument and the diffuse sound of the whole ensemble.

It is clear, that the transition from single sound sources to ensembles is acoustically very complex, with only little research having been published. This study aims to provide a first analysis of ensemble sound and instrumental blending. A fundamental question that needs to be answered is how many instruments are required for a number of string instruments to be perceived as an ensemble, rather than as individual instruments, as well as the influence the room has on the blending of the sound. For this, a listening test was designed with recordings made at the Salle Claude-Champagnie in Montral with the orchestra of the Université de Montréal. Furthermore, the "imperfections" of the players are to be examined, such as the phase relationship between the instruments and the intonation.

2 SETUP

The data used for the following investigation was collected on the 25th and 26th September 2019 during measurements at the Salle Claude-Champagnie in Montral as part of the ACTOR project. The measurements were made in collaboration with the orchestra of the Université de Montréal as well as researchers from the McGill University in Montreal and the Hochschule für Musik Detmold.

The measurements are based on recordings of P. I. Tchaikovsky's 6th Symphony in B Minor op. 74; specifically, the phrase of the first violins in the first movement from bar 89 to 97. This excerpt was first played by one violin, then again by 2, 3, 4, 6 and 9 violins.

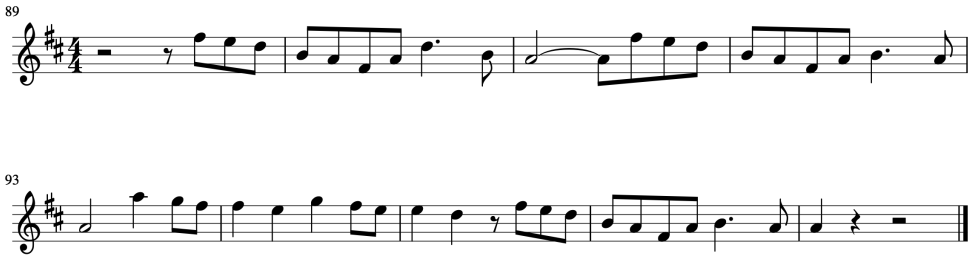

Figure 1 the violin theme

The first six violins were fitted with small piezo microphones, which were taped onto the tailpiece of the instruments. These were sent to an RME Octamic for amplification and then via Dante to the control room where they were recorded in the Pyramix Digital Audio Workstation.

Traditional recording methods were also implemented; 3 Schoeps MK4 microphones were used as spot microphones, 3 DPA4006 were used for a decca tree, a Neumann KU100 dummy head was placed in the second row and 2 Schoeps MK2S were placed in the balcony as room microphones. These went into RME Micstacy preamps and then via MADI into the control room.

3 Pitch analysis

The piezo microphone signals were used for a pitch analysis of each of the first six violins, to study if an ensemble sound is related to the pitch difference of the individual players. This analysis was done with the software Praat. The first five notes of the phrase where analyzed in each instrument and the mean and standard deviation in Herz and in cent were calculated. The orchestra tuning was 443Hz, so the reference pitches were based on this value. The Data is shown in Tables 1 to 5, with sharp notes (too high) marked in red, correct notes (correct to the nearest Hz) marked in green and flat notes (too low) marked in orange.

The data presented here has to be treated with care, due to the high amount of uncertainty in the measurements. Violinists play with a vibrato, which is done by moving the finger used to determine the length of the string, leading to a fluctuation in pitch. This makes it difficult to determine an exact pitch, as the pitch varies over the duration of the note; hence an average value had to be taken.

Note	Pitch (Hz)	Violin 1	Violin 2	Mean (Hz)	σ (Hz)	σ (cent)
f#	745.0	750	750	750.0	0.0	0.0
e	663.8	664	671	667.5	4.9	12.8
d	591.3	591	596	593.5	3.5	10.3
b	497.3	501	503	502.0	1.4	4.9
a	443.0	443	443	443.0	0.0	0.0
f#	372.5	373	375	374.0	1.4	6.5

Table 1 Pitch 2 violins

When only two violins play (table 1), the first violin plays most of its notes correctly and some to sharp, whereas the second violin generally plays a bit sharp. The standard deviation ranges from 0 to 12.8 cents, depending on the note played.

Note	Pitch (Hz)	Violin 1	Violin 2	Violin 3	Mean (Hz)	σ (Hz)	σ (cent)
f#	745.0	754	749	748	750.3	3.2	7.4
e	663.8	668	666	668	667.3	1.2	3.0
d	591.3	592	592	593	592.3	0.6	1.7
b	497.3	501	505	502	502.7	2.1	7.2
a	443.0	444	446	444	444.7	1.2	4.5
f#	372.5	376	375	376	375.7	0.6	2.7

Table 2 Pitch 3 violins

When three violins played (table 2), they all played quite sharp. There is not a single note where all violins had the exact same pitch. The range of the standard deviation was from 0.6 to 7.4 cents.

Note	Pitch (Hz)	Violin 1	Violin 2	Violin 3	Violin 4	Mean (Hz)	σ (Hz)	σ (cent)
f#	745.0	750	752	747	748	749.3	2.2	5.1
e	663.8	670	670	670	666	669.0	2.0	5.2
d	591.3	591	595	591	595	593.0	2.3	6.7
b	497.3	500	501	504	498	500.8	2.5	8.6
a	443.0	444	445	446	446	445.3	1.0	3.7
f#	372.5	376	374	378	375	375.8	1.7	7.9

Table 3 Pitch 4 violins

When four violins played (table 3), the tendency again was to play quite sharp, with only a few notes spot on. The range of the standard deviation was again reduced, ranging from 3.7 to 8.6 cents.

Note	Pitch (Hz)	Violin 1	Violin 2	Violin 3	Violin 4	Violin 5	Violin 6	Mean (Hz)	σ (Hz)	σ (cent)
f#	745.0	752	749	754	753	740	749	749.5	5.1	11.7
e	663.8	665	664	670	670	660	669	666.3	4.0	10.4
d	591.3	589	589	593	593	590	594	591.3	2.3	6.6
b	497.3	502	501	504	502	492	504	500.8	4.5	15.5
a	443.0	443	444	445	444	442	445	443.8	1.2	4.6
f#	372.5	375	376	375	377	368	372	373.8	3.3	15.3

Table 4 Pitch 6 violins

When six violins played (table 4), it can be seen that the 5th violin was consistently flat. However, again the general tendency was rather sharp, with a few correct notes. While during the increase from two to four violins, the range of the standard deviation was decreasing, in this measurement the range of the standard deviation increases dramatically, ranging from 4.6 to 15.5 cents. This is mainly due to the fifth violin playing flat.

Note	Pitch (Hz)	Violin 1	Violin 2	Violin 3	Violin 4	Violin 5	Violin 6	Mean (Hz)	σ (Hz)	σ (cent)
f#	745.0	743	753	750	750	741	745	747.0	4.7	10.8
e	663.8	662	663	668	672	662	675	667.0	5.6	14.4
d	591.3	590	590	591	593	590	594	591.3	1.8	5.1
b	497.3	503	496	502	500	501	498	500.0	2.6	9.0
a	443.0	443	442	444	447	443	446	444.2	1.9	7.5
f#	372.5	373	377	376	379	372	372	374.8	2.9	13.5

Table 5 Pitch all violins

The last recording was of all nine violins (table 5). Only the first six had piezo pickups, which could be used to evaluate the pitch. When all nine violins played, only violin 4 was always sharp. The standard deviation in pitch ranged from 5.1 to 13.5 cents, which is similar to the range when only six violins played.

Number of Violins	Minimum	Maximum	Range
2	0	12.8	12.8
3	0.6	7.4	6.8
4	3.7	8.6	4.9
6	4.6	15.5	10.9
9 (only 6 in measurement)	5.1	13.5	8.4

Table 6 Pitch summary

An interesting observation is the fact that the minimum standard deviation in pitch of a single tone is constantly rising. This suggests that the more instruments there are, the greater the range of pitches that is played. Up to four violins, the maximum standard deviation increases, but then increases again as more violins play. Especially the increase of the minimum standard deviation may suggest that an ensemble sound can be defined by this spread of different pitches played by each instrumentalist. Especially when one considers that each player's vibrato is slightly different, it might be possible that the spread of pitches heard at the same time is even larger than these measurements revealed.

4 Listening Test

4.1 Design

In order to answer the question of how many violins are needed to create the feeling of an ensemble a listening test was performed. The test consisted of 24 different examples in random order of the recorded violin phrase with different microphones and different numbers of musicians. The piezo microphones were not used, as they did not offer a natural sound similar to what can be heard in a room. The 24 variations consisted of 1, 2, 3, 4, 6 or 9 musicians and the spot microphones, decca tree, dummy head and room microphones. All examples were in stereo, except the ones recorded with the spot microphones, which were in mono.

The examples consisted of almost no mixing. The spot microphones were all mixed together at the same level. The three microphones of the decca tree were also mixed together at the same level, with the left microphone only on the left channel, the right microphone only on the right channel and the center microphone on both channels at −3dB. The dummy head and room microphone examples consisted of the left and right microphones panned to the left and right channel respectively.

For each example, the test subjects had to estimate the number of violins and state whether or not it sounded like an ensemble. The first question was to determine the accuracy to which one can hear the number of instruments – the lower the accuracy, the less individual instruments can be identified, hence sounding more like an ensemble. The second question was to determine how many violins are needed for the listener to conceive an ensemble, making it a more subjective question. The term “ensemble” was purposefully not defined, so as to not influence the listener in terms of what parameters are criteria for an ensemble sound.

The listening test was hosted with an online provider, allowing participation from all over the world. The instructions stated that the participants should take the test with headphones, as the examples from the dummy head are binaural. Over a period of five weeks, 54 people participated in the listening test, 13 of who only partially completed the test.

Participants came from all over Europe and North America.

4.2 Results

4.2.1 Spot microphones

Violins	Mean	StdDev	Mode	%Correct	EnsembleYes%	EnsembleNo%	Min	Max
1	1.0	0.1	1	98.1	3.7	96.3	1	2
2	2.1	0.3	2	90.7	13.0	87.0	1	3
3	3.1	0.9	3	51.9	35.2	64.8	1	5
4	4.1	1.8	4	25.9	77.8	22.2	1	10
6	4.4	1.5	3 and 4	7.4	74.1	25.9	2	8
9	5.2	2.5	3	1.9	88.9	11.1	1	15

Table 7 Spot microphone results

As the number of violins increases, the percentage of test subjects who felt that the example sounded like an ensemble playing also increases. The biggest difference is between three and four violins – whereas only 35% thought three violins sound like an ensemble, 78% thought that four violins sound like an ensemble. Interestingly, fewer people thought that six violins sound like an ensemble than that four violins sound like an ensemble.

As the number of violins increases from one to four, the mode (most common answer) from the test matches the number of violins. Also, the mean is extremely close to the correct number of violins, with the standard deviation increasing as the number of violins increases. This is also due to the fact that the number of participants whose estimate was correct decreases as the number of violins increases. While for one and two violins over 90% of participants were correct, only 52% and 26% were correct for three and four violins

respectively.

For six and nine violins, the number of violins was vastly underestimated as, on average, 4.4 and 5.2 violins respectively. The mode was also much lower, between 3 and 4 and the number of correct estimates was negligible.

4.2.2 Decca Tree

Violins	Mean	StdDev	Mode	%Correct	EnsembleYes%	EnsembleNo%	Min	Max
1	1.0	0.2	1	95.7	4.4	95.6	1	2
2	2.1	0.5	2	73.3	15.6	84.4	1	3
3	3.4	2.3	2	26.1	43.5	56.5	1	10
4	3.8	1.4	3	28.3	66.0	34.0	1	8
6	4.6	2.8	3	8.7	73.9	26.1	1	12
9	5.7	2.4	5	0.0	97.8	2.2	1	12

Table 8 Decca tree results

With the decca tree microphones, the number of subjects perceiving an ensemble also increased as the number of violins increases. The biggest increase was from two to three violins (from 16% to 44%) and from six to nine violins (74% to 98%). It is noteworthy, that with all nine violins playing, almost all subjects stated that they perceived an ensemble.

For one and two violins, both the mean and the mode match the number of violins playing. However, for two violins the accuracy of the test subjects was a lot lower than for one violin.

For three and four violins, the number of correct estimates is significantly lower (26% and 28% respectively). Also, the mode is too small in both cases. However the mean is still well within one standard deviation of the actual number of violins. This suggests that while subjects found it difficult to guess the correct number of violins, they were not far off.

For six and nine violins, the number of correct guesses is negligible. As with the spot microphones, the number of violins was underestimated in both cases. With six violins, the mean is still within one standard deviation of the correct number of violins, but with nine violins this is not the case. This may explain the large difference in ensemble perception between six and nine violins.

4.2.3 Dummy Head

Violins	Mean	StdDev	Mode	%Correct	EnsembleYes%	EnsembleNo%	Min	Max
1	1.2	0.5	1	87.5	10.0	90.0	1	4
2	2.5	1.9	2	64.1	15.4	84.6	1	12
3	3.3	2.5	2 and 3	25.0	50.0	50.0	1	12
4	4.1	1.9	4	37.5	85.0	15.0	1	10
6	4.2	2.2	3	10.0	75.0	25.0	1	10
9	5.3	2.8	5 and 6	0.0	87.5	12.5	1	12

Table 9 Dummy head results

The dummy head is a head model with microphones in the ears. This way head related transfer functions (HRTFs) are included in the recording (2). When listening over headphones, one can hear almost exactly what could have been heard in that place in the room. Due to this, the dummy head is the microphone that most exactly represents what a listener in the concert hall would have heard.

For one and two violins, the mean is further away from the correct number of violins than in with the previous microphones, however it is still quite accurate and within one standard deviation. However the standard deviation is also larger. The percentage of subjects correctly estimating the number of violins is also a lot less.

For three and four violins, the mean is also quite accurate; however, the number of correct estimates is a lot lower. At three violins, half the subjects perceived the sound as an ensemble, while at four violins 85% of the subjects stated that the examples sounds like an ensemble.

As in the previous section, for six and nine violins the number of violins was underestimated and very few estimates where correct. Surprisingly, more people thought that four violins sounded like an ensemble than six.

4.2.4 Room Microphones

Violins	Mean	StdDev	Mode	%Correct	EnsembleYes%	EnsembleNo%	Min	Max
1	1.0	0.2	1	97.5	0.0	100.0	1	2
2	2.0	0.6	2	67.5	10.0	90.0	1	4
3	2.4	1.0	2	20.0	12.8	87.2	1	6
4	3.0	1.3	3	30.0	40.0	60.0	1	7
6	4.1	2.2	3	5.0	69.2	30.8	1	10
9	4.4	1.8	3	2.5	71.8	28.2	1	9

Table 10 Room microphones results

The room microphone examples follow the general tendency of the other microphones. However, for four or more violins, the number of violins was ever more underestimated than in the other examples. Also, even with all nine violins playing, only 72% of test subjects perceived the sound to be that of an ensemble. The maximum estimates for how many violins were playing were also lower. Even though test subjects could not accurately say how many violins were playing, they still did not perceive the sound as an ensemble. This suggests that just because one cannot distinguish individual instruments, this does not mean that an ensemble sound can be presumed.

4.2.5 Summary

The recordings made with the decca tree showed the best results for ensemble sound for when all violins played. The decca tree, due to its large distances between the microphones, may produce extremely uncorrelated signals, which are known to give a good feeling of envelopment (3). However, uncorrelated stereo signals alone do not seem to be enough, as the room microphones are also very uncorrelated, but do not prove useful for creating an ensemble sound. The decca tree is also the only microphone system, which due to its proximity to the orchestra has a relatively large percentage difference in distance from one desk to the next, causing a good depth perception (the last violin desk has a multiple of the distance to the decca tree that the first desk has). Perhaps it is this depth that may cause this ensemble sound.

Except from the room microphone examples, four violins seem to be enough to create a feeling of an ensemble in at least half the test subjects. However the more instruments play, the more subjects get this feeling. In the dummy head and spot microphones, four violins sounded to more test subjects like an ensemble than 6 violins. A pitch analysis showed that when six violins played, the range of pitches was larger than when nine violins played, suggesting slightly worse intonation. This might be the cause for this effect, as the deviation in intonation of the players could lead to the group sounding less homogenous, and hence less like an ensemble.

The reason for the worse ensemble sound in the room microphones might be because of the apparent width of the ensemble. The room microphones are so far away, that the angle between the left and right edge of the stage becomes small. This leads to a reduced width during reproduction, which opposes the idea of many instruments spread over the stage.

While one violin playing had a very high rate of correct prediction no matter the microphone setup, only the

spot microphones had a high rate of correct prediction for two violins. Also, for three and four violins the mean was closest to the correct number of violins with very small standard deviations. The spot microphones were the closest to the individual violins and as a result recorded less reverberant room sound and more playing sounds such as bowing. This might make individual violins more distinguishable.

Six and nine violins are often estimated as being a lot less. This suggests saturation at four violins, at which point one cannot distinguish individual violins. Perhaps, even when more violins play, one can distinguish a few violins out of the whole, and hence the responses yielded such low numbers.

5 CONCLUSION

The transition between being able or not to estimate the correct number of violins or at least come close to the correct number seems to be about at six violins, regardless of the microphone setup. However, for the listener to perceive an ensemble, more violins lead to better result. The microphone setup also influences this, with the decca tree offering the most ensemble-like sound, while the room microphones were the worst.

Intonation may influence the listeners' perception, as six violins have in some examples received a lower ensemble rating than four violins, which could be due to the larger deviation in intonation. If this is the reason for this effect, an ensemble sound can be achieved through good intonation, meaning that sounds will little deviation in frequency blend better together.

The terms blending and ensemble sound do not seem to be the interchangeable. Especially with the room microphones it can be seen that while the number of violins cannot be accurately estimated (high blending of individual instruments), the result still does not sound like an ensemble to all listeners. However, to achieve an ensemble sound, a high amount of blending seems to be required.

The results of this study apply to the recordings done by the orchestra of the Université de Montréal in the Salle Claude-Champagnie. For definitive results, this study would need to be replicated with different orchestras in different concert halls. Especially interesting would be a study with a professional orchestra which has been playing together for years, in order to analyze their intonation and phase relationships and the effect this has on the listener's perception. Also, while the mean estimate for four violins is about the correct value, six violins are largely underestimated. In future studies, examples with five violins playing the theme should also be included.

REFERENCES

[1] Meyer, J. (2015): Akustik und musikalische Aufführungspraxis. Bergkirchen: PVVMEDIEN

[2] Boré G., Peus S. (1999): Microphones. Berlin, Neumann GmbH

[3] Wittek, H. (2015): Stereo, 5.1 and 3D-Audio, Tonmeistertagung 2014