INFLUENCE OF RECORDING TECHNIQUE AND ENSEMBLE SIZE ON APPARENT SOURCE WIDTH

Renzhi Guo

School of Architecture
Rensselaer Polytechnic Institute
Troy, NY, USA
kerou273.15@gmail.com

Jonas Braasch

School of Architecture
Rensselaer Polytechnic Institute
Troy, NY, USA
braasj@rpi.edu

ABSTRACT

Previous studies have looked at how different concert halls with different lateral reflections affect apparent source width. Yet, the perceptual effects of different source distributions with different recording techniques on apparent source width are not well understood. This study explores how listeners perceive the width of an orchestra by using four stereo and one binaural recording techniques and three wave field synthesis ensemble settings. Subjective experiments were conducted using stereo loudspeakers and headphone to play back the recording clips asking the listeners to rate the perceived wideness of the sound source. Results show that recording techniques greatly influence how wide an orchestra is perceived. The primary mechanism used to judge auditory spatial impression differs between stereo loudspeaker and headphone listening. When western classical symphony is recorded and reproduced by two-channel stereophony, the changes in instrument positions in terms of increasing or reducing the physical source width do not lead to an obvious increase or reduction on the spatial impression of the performing entity.

1. INTRODUCTION

The early formation of the concept of apparent source width (ASW) began in the study of concert hall acoustics. Asking for what quantifies the spatial impression and how the auditory system perceives the quality of a performance rendered by the concert hall have been the driving questions in the past decades. Composer, mixing engineer, and architectural acoustician share a common goal which is to produce an effect of enlargement that might be counter-intuitive to what appears in the look. Specifically, using an octet to sound like an eighty-member symphony orchestra, widening the stereo image by panning the left and right channels, or making a medium-sized concert hall sound like twice its size. Based on universally accepted theory, the sense of space for the auditory system involves a correlation between two ears. This sensation is represented by psychoacoustic parameters—apparent source width (ASW) and listener envelopment (LEV) [1, 2]. While acousticians have been working on reaching a sound field by designing an enclosed space, audio engineers have been dealing with creating a sound field in an enclosed space by extending the artistic dimensions. Although acousticians would not name their topics as "3D audio" and sound engineers would not use the term "auralization" in both of their studies of sound field reproduction around 30 years ago, they are merging into one body of research community on spatial hearing. Important figures who attested to this fact are Tim Ziemer [3, 4, 5], Floyd Toole [6], Sascha Spors [7, 8], Jürgen Meyer [9], and Stefan Weinzierl [10, 11, 12]. The impression of listeners to aurally "see" the size of a performing entity is crucial to the success of both a concert hall and a reproduced sound field. Despite acousticians and audio engineers are merging, there are still gaps of consistent knowledge or consensus between what characterizes the perception of performing entities with different sizes and what makes it possible for an effective reproduction on the basis of human hearing.

Among acousticians, abundant studies have gone into spatial impression with the goal to investigate the effect of the room on the perception of the sound source, but seldomly the effect of the recording and reproduction technique combined with ensemble width on the perception of the sound source. Similarly, there is a lack of contribution to the perceptual studies for sound field reproduction led by audio engineers, let alone study on apparent source width. Early works have examined the testable predictors for quantifying apparent source width, such as IACC and LF [2, 13, 14, 15]. However, few research investigated apparent source width by directly recording the sound field of a symphonic performing entity with different recording techniques along with different ensemble arrangements. Similarly, research regarding perception of sound field reproduction have investigated directivity [16], distance [17, 18], loudness [19], audible artifact of focused sound sources [20, 21], off-center positions [22], localization [23], but few has been done on apparent source width. Even if there are studies on apparent source width, the signal or stimuli used were not a direct recording performed by a virtual orchestra in a concert hall, but synthesized signal, sine tone, noise sequence, convolution, or applying delays.

This paper uses the technology combination consisting of four stereo and one binaural recording techniques, one concert hall, three wave field synthesis ensemble settings, and two stereo reproductions, to investigate how recording techniques and physical source width (ensemble size) affect apparent source width. The authors aim to provide insights into how the auditory system perceive the width of a musical ensemble within the current sound reproduction technologies and help extend the knowledge on characterizing apparent source width.

2. METHOD

Beethoven's Symphony No. 8, performed by wave field synthesis in three ensemble settings, was recorded using four stereo recording techniques and one binaural technique at two distances in EMPAC concert hall. Perceptual experiments were conducted using stereo loudspeakers and headphone to play back the recording clips while asking the listeners to rate the perceived wideness of the sound source. The general method chain is shown in Figure 3.

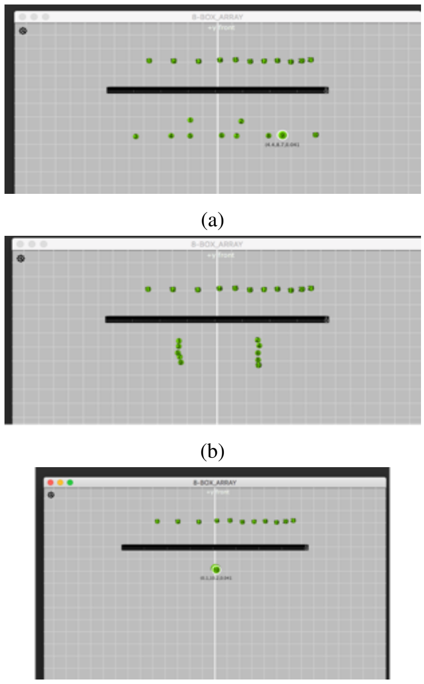
2.1. Music Material, WFS System and Source Distribution

For presenting a virtual orchestra through wave field synthesis, an anechoic music recording was used. The recording was performed by Orchester Wiener Akademie in performing the first, second, and fourth movements of Beethoven's Symphony No.8, conducted by Martin Haselbock and engineered by Christoph Bohm, David Ackermann, and Stefan Weinzierl in 2020 [10].

The current wave field synthesis loudspeaker array system in EMPAC concert hall can take up various configurations, such as linear, circular, or arrays with height difference. A linear configuration of WFS applied in this research. The array consists of 8 modules. Each module has 31 drivers (small loudspeakers) and 1 subwoofer. Both the subwoofers and the drivers are connected to the DANTE audio network. The small speaker drivers are Tang Band W2-2136S and the subwoofers are Sony SA-CS9. The length of each module is 1.87 meters, totaling a linear length of 15 meters. The signals run from 2 Apple Mac Pros, through a real-time spatial audio processor software Spat made by IRCAM, a French research suite focused on music acoustics and audio technology, combined with Max/MSP signal processing, to the synthesized wave field on the stage [24].

The instrumentations of Beethoven's 8th Symphony from the anechoic recording were sent to ten channels both in Reaper and Max/MSP for playback. The instrument channels were then arranged for three ensemble distributions using virtual panning spots in Max/MSP: orchestral, stereo, and mono. These three ensemble arrangements are exactly the same in instrument channels, but they display different physical source widths on the stage. The orchestral setting mimics the instrument positions of real orchestra. The stereo setting divides the orchestra into two parts on the stage, making the instruments closer to each other and shortening the physical expansion size on the stage. The mono setting gathers all the instruments to the center of the stage, making the physical ensemble width to the smallest.

Table 1 shows the ten instrument channels in Max/MSP. Figure 1 shows the three virtual source distributions. The sources are positioned in front of the loudspeaker array as marked from number 1 to number 10.



(a)

(b)

(c)

Figure 1: Three WFS virtual source distributions in Max/MSP: (a) Orchestral setting; (b) Stereo setting; (c) Mono setting. Note that the functioning virtual sources are only in front of the WFS array on the stage (below the black line in the picture), and the virtual spots numbered from 11 to 21 are not assigned in any settings.

2.2. Recording Techniques

2.2.1. Spaced Omni-Directional Pair

Two identical omni-directional microphones with a height of 2.21 meters were parallelly spaced 61 centimeters apart facing the sound source. The microphones used in this study were DPA 4006A.

2.2.2. ORTF Pair

Two identical cardioid microphones with a height of 1.91 meters were angled 110 degrees apart with a distance of 17 centimeters between the two capsules. The microphones used in this study were AKG CK1.

2.2.3. MS Pair

One cardioid microphone was placed towards the sound source for the mid channel, and one figure-8 microphone was placed facing laterally for the side channel. The height for the MS pair was 1.88 meters. The processing for MS was done using Reaper by 1) Splitting one stereo track that contains both the mid and the side signal into two mono tracks by which each mono track represents the mid signal and the side signal respectively; 2) Duplicating the mono track of the side signal and making one of them phase-inverted; 3) Panning the first side signal track hard-left and the second side signal track hard-right; 4) Rendering this result to a stereo wave

Table 1: Ten instrument channels as virtual panning spots in Max/MSP.

Instrument	Woodwind, Timpani	Brass, Timpani	Violin 1	Violin 1	Violin 2	Violin 2	Viola	Viola	Cello	Double Bass
Virtual Panning Spot	1	2	3	4	5	6	7	8	9	10

file. The signal gains for both mid and side were kept untouched in the whole process. The microphones used in this study were AKG C214 and AKG C414 B-ULS.

2.2.4. *Blumlein Pair*

Two identical figure-8 microphones with a height of 1.68 meters were used with the capsules placed at 90 degrees from each other. The microphones used in this study were Cascade Fat Head.

2.2.5. *Binaural Recording*

The microphones used in this study was Neumann KU 100 with a height of 1.8 meters.

2.3. Recording Procedure and Test Stimuli

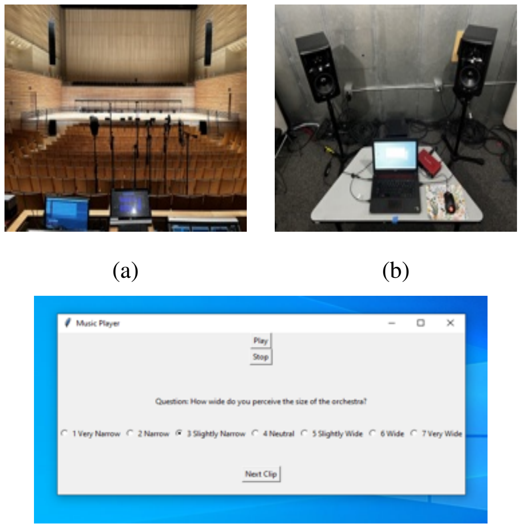
For the sending part, two Mac Pros work together doing separate tasks for playing back the orchestra in Reaper and controlling the virtual sound source in Max/MSP. The signals are then played back through wave field synthesis linear array on the stage. For the receiving part, four stereo microphone pairs and one binaural head, with ten channels totally, are plugged into SoundCraft Delta mixer, and then the mixer is connected to MOTU 1296 audio interface, and finally the audio interface is connected to a Windows computer.

The recording was made at two different distances away from the stage: in the front row and in the middle row. The front row was recorded first. In three source distribution settings, the orchestral setting was recorded first, then the stereo setting, and finally the mono setting. The three source distributions were recorded in both distances taking on the same order. The entire music material (the excerpts of the first movement, second movement, and fourth movement) was recorded in every source distribution setting at both distances. With two distances, three source distributions, and five recording techniques, there are thirty recording files totally.

The test stimuli are cut from the first movement of Beethoven's Symphony No. 8, bar 45 to bar 72 recorded in EMPAC concert hall. The orchestration of the first movement involves the following instruments: 2 flutes, 2 oboes, 2 B flat clarinets, 2 bassoons, 2 French horns, 2 F trumpets, Timpani, and string sections. There are in total of 30 stimuli from 3 ensemble settings by 5 recording techniques by 2 distances. Each stimulus is 25 seconds.

2.4. Experiment 1 (Stereo Reproduction)

Thirty stimuli introduced above were assessed through a stereo setup in a treated listening room. The room is small-sized, insulated inside a larger office, and installed with micro-perforated panels (MPP) on the four walls and the ceiling. The hardwood floor is covered by carpet, and four bass traps are sitting in the four corners of the room. The dimensions of the room are 3.76 m in length, 3 m in width, and 2.45 m in height. Seven male subjects with normal hearing participated in the first experiment. The subjects are coming from audio technology and recording class at Rensselaer Polytechnic Institute, aged from 19 to 21. Subjects are asked to rate the perceived width of the performing entity based on a 7-point Likert scale for every stimulus. The thirty stimuli were presented in a random order to the subjects. This listening test is run by a program written in Python, with a graphical user interface (GUI) for playing the stimuli and displaying a single question to



(a) (b)

(c)

Figure 2: Recording, reproduction, and listening experiment: (a) Microphone placement in the mid row; (b) Experiment lab top view; (c) Experiment GUI.

the subjects: "How wide do you perceive the size of the orchestra?" (See Figure 2 (c)) The loudspeakers used in this experiment are a pair of JBL 308P MKII, settled at ear height for a seated listener and are in the shape of an equilateral triangle with the listener, connected to the Windows laptop through a FocusRite 4i4 audio interface. (See Figure 2 (b))

2.5. Experiment 2 (Headphone Reproduction)

The second experiment is exactly the same in design and location as the first experiment, except for the reproduction method. The headphone used in this experiment is Beyerdynamic DT 770 Pro 250 ohm. Four male subjects from the same class as the subjects in the first experiment participated in the second experiment.

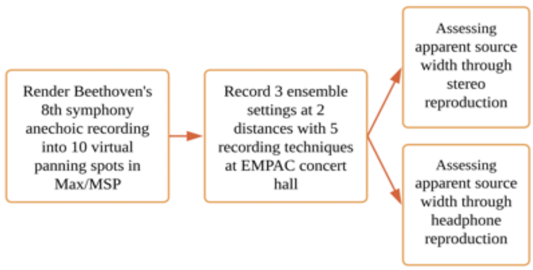
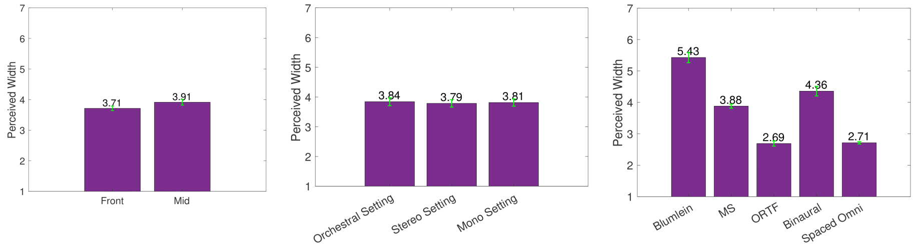


Figure 3: General method chain.

3. RESULT

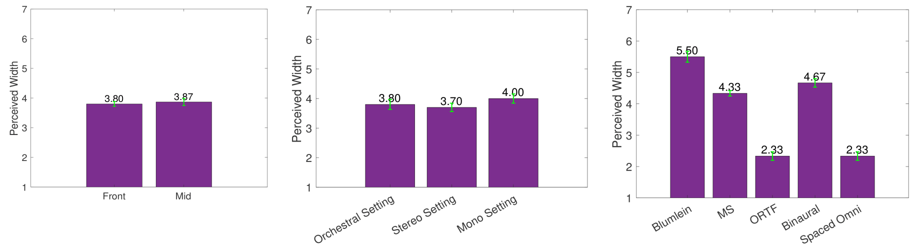
3.1. Experiment 1

Figure 4 and Figure 5 show the mean and the median values with error bars among the responses with regard to each variable: dis-



(a) (b) (c)

Figure 4: Mean response in stereo reproduction experiment: (a) Distances; (b) Ensemble settings; (c) Recording techniques.



(a) (b) (c)

Figure 5: Median response in stereo reproduction experiment: (a) Distances; (b) Ensemble settings; (c) Recording techniques.

tance, ensemble size, and recording technique in the first experiment. The data reflects that under stereo reproduction of classical orchestral music, virtually played by wave field synthesis, different source distributions play a minor role in the perception of auditory width. Mono setting may result in a wider perceived width over orchestral and stereo settings. Meanwhile, recording of the mid row which holds a longer distance to the stage is not obvious to discern the difference from what was heard in the front row through a two-channel stereo playback. And a further listening position in the concert hall results in a wider sound image in the reproduced sound field. The recording techniques directly impact how wide a musical source will be perceived. In reproducing the binaural recording through stereo loudspeakers, no crosstalk cancellation was performed. However, the responses predominantly recognize the binaural recording as spatially perceived wide. The Blumlein technique is considered as being able to produce a richer and all-inclusive spatial impression over the others through a stereo playback. Microphones with small diaphragms, such as the ones used for ORTF and spaced omni, produced a narrow source width and a thin spatial impression among the other recording techniques in stereo reproduction.

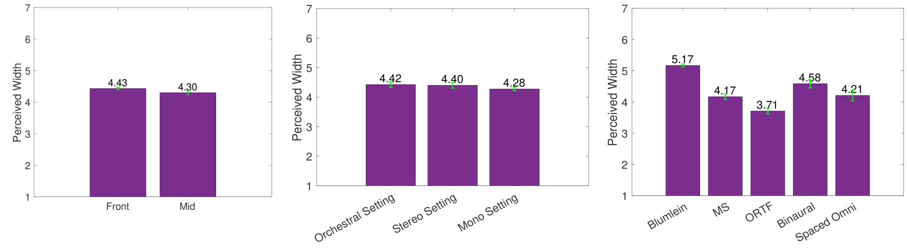
After visual inspection across different factors (distance (**DS**), ensemble size (**EN**), recording technique (**RT**)), which shows that the recording technique has the most significant differences between levels, a three-way ANOVA was performed to examine the interactions among factors (See Table 2). For effects of a single factor, the ANOVA result agrees with the mean and median calculations showed in Figure 4 and Figure 5. The main effect of distance on perceived source width was not statistically significant (F = 1.00, p = 0.32), indicating that changing capturing distance alone will not necessarily lead to a significant change in the wideness of an auditory image when played back in a stereo field. Similarly, the main effect of ensemble size on perceived source width was not statistically significant (F = 0.03, p = 0.97), implying that

Table 2: ANOVA result for stereo reproduction experiment.

Effect	*d.f.*	*F*	*p*	*SS*	*MS*
DS	1	1	0.32	2.1	2.1
EN	2	0.03	0.97	0.11	0.06
RT	4	26.97	0	225.88	56.47
DS × **EN**	2	0.85	0.43	3.54	1.77
DS × **RT**	4	2.03	0.09	17.02	4.25
EN × **RT**	8	1.33	0.23	22.27	2.78
DS × **EN** × **RT**	8	2.03	0.04	33.99	4.25

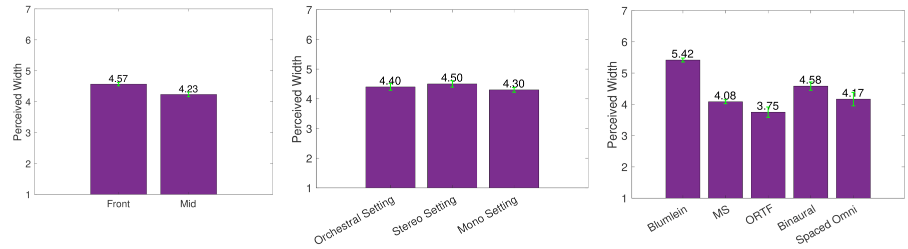
for a same music piece with the same number of instrumentalists, a wider ensemble arrangement on stage does not equal to a wider aural image of this ensemble. In contrast, the main effect of recording technique on perceived source width was highly significant (F = 26.97, p < 0.001), suggesting the importance of the capturing techniques for specific music pieces on influencing how wide an orchestra is perceived.

For combined effects, **DS** × **EN** (F = 0.85, p = 0.43) showed that the effect of distance on perceived source width did not depend on the ensemble size, and vice versa. In other words, for a performing entity who's playing the same piece, the listener's perceived source width affected by short or long distances against the stage did not differ no matter how this orchestra changes its instrument positions on stage. Conversely, **DS** × **RT** (F = 2.03, p = 0.09) was statistically significant, implying that the effect of distance on perceived source width varied depending on the recording technique used, and vice versa. This emphasizes the consideration for both the capturing techniques and the capturing position in pro-



(a) (b) (c)

Figure 6: Mean response in headphone reproduction experiment: (a) Distances; (b) Ensemble settings; (c) Recording techniques.



(a) (b) (c)

Figure 7: Median response in headphone reproduction experiment: (a) Distances; (b) Ensemble settings; (c) Recording techniques.

ducing preferred auditory results. Similarly, the **EN** × **RT** (F = 1.33, p = 0.23) interaction was marginally significant, suggesting potential dependency of the effect of ensemble size on perceived source width on the recording technique, and vice versa. This tells that the perceived source width caused by physically varying the ensemble width may not produce consistent results because of the capturing techniques used. Finally, the data for **DS** × **EN** × **RT** (F = 2.03, p = 0.04) combined effect was statistically significant. This three-way interaction suggests that in stereo reproduction, the effect of both distance and ensemble size on the perception of source width may largely depend on how a specific music passage was captured. And this inter-factor dependence from the three-way interaction highlighted how the human auditory system produces the impression of the wideness of a sound image.

3.2. Experiment 2

Figure 6 and Figure 7 show the mean and the median values with error bars among the responses with regard to each variable: distance, ensemble size, and recording technique in the second experiment. Based on the data of distance and ensemble setting, when reproducing a sound field, headphone may yield the opposite outcome to what loudspeakers produced. The listening position that has a longer distance to the stage can generate a sense of wider source width when played back through loudspeakers, but a narrower sound image when played back through headphone. Similarly, mono setting has less possibilities to create a wide perception in headphone, which is the opposite case for loudspeakers. And orchestral setting which has more spatially separated sound sources yielded wider width perception. However, what experiment 2 and experiment 1 have in common is that different ensemble sizes are less likely to influence the perceived width of a classical orchestra. In addition, Blumlein technique and binaural technique have a solid capability to reproduce a more bountiful sound field through both stereo loudspeakers and headphone. What the small-diaphragm microphones lack in a two-channel stereophonic field, they make up for in headphone. ORTF and spaced omni are able to have a decent performance when listened back through headphone. And spaced omni technique is able to generate a richer sound field over MS technique and ORTF technique through headphone.

Table 3: ANOVA result for headphone reproduction experiment.

Effect	*d.f.*	*F*	*p*	*SS*	*MS*
DS	1	0.36	0.55	0.53	0.53
EN	2	0.17	0.84	0.52	0.26
RT	4	4.76	0	28.45	7.11
DS × **EN**	2	1.75	0.18	5.22	2.61
DS × **RT**	4	0.43	0.79	2.55	0.64
EN × **RT**	8	0.66	0.72	7.90	0.99
DS × **EN** × **RT**	8	2.53	0.02	30.20	3.78

After the mean and median calculation, a three-way ANOVA followed (Showed in Table 3) to see the interactions among factors. Table 3 and Table 2 have many similarities. For example, as main effect, while distance (F = 0.36, p = 0.55) and ensemble size (F = 0.17, p = 0.84) along did not show significance on the dependent variable (perceived source width), recording technique (F = 4.76, $p < 0.05$) had a great influence on the dependent variable, which agrees with the ANOVA for stereo reproduction, highlighting the consistent role of recording technique across different playback methods. Similarly, the three-way combined effect (F = 2.53, p = 0.02) showed that in headphone reproduction, while the two-way interaction may not greatly affect how wide a music ensemble sounds in our aural architecture, our judgment relies strongly on how all three factors are simultaneously considered in the design of an auditory delivery, because the influence of any one of the factors depends on the levels of the other two factors. However, the **DS** $\times$ **RT** (F = 0.43, p = 0.79) interaction which showed significance, was not so at all in a headphone environment, which marks the obviously different, or even contrary listening outcomes between headphone and stereo loudspeaker.

4. DISCUSSION

4.1. Experiment 1

The data tells the fact that for EMPAC concert hall, the spatial impression is not deteriorated due to longer distance. Instead, the mid row is able to generate a wider, unique source width compared to the front row. Under a treated condition where the room size, the absorption, and the loudspeaker-listener angle are well contained, the spatial impressions on different locations, such as the front row and the mid row in the concert hall, are well preserved in a two-channel stereophonic field. Listening in a concert hall at a position further away from the stage does not necessarily lead to a narrow auditory image, and this experience can be reproduced by rigorous stereophonic means. Wide sound image is favored, but if the sources are distributed too wide, probably the auditory perception will not be satisfied. This is what human hearing mechanism does to the spatial sound in the context of music. The binaural hearing system functions to the highest degree to integrate harmonic signals rather than separate these signals. When listening to the orchestra in the mid row, higher order reflections as well as sufficient low frequencies are able to reach the listener, helped generate a greater spatial impression in the auditory system through the integration of those harmonic signals, resulted in an enlargement of the sound source in a harmoniously wide sound image. When played back through stereo loudspeakers, those harmonic signals are reproduced by bringing down the room reflections broadbandly and by calibrating the length and angle between loudspeakers and listener, and thus, the real aural image of the concert hall is conveyed by the phantom image from two loudspeakers. Comparatively, when listening to the orchestra at a far front position, the sound source expands to its largest distribution degree in front of the two ears, but the brain is confused because it is busy finding which individual source to focus on, since the sound source as a whole is too wide according to the scale between the listener and the sound source. In other words, the sound sources are too discrete and thus there is not enough space for the brain to shape an integrated auditory image of the sound source. Consequently, the sense of spaciousness will have a hard time applying to the perception of the sound source.

Based on verbal inquiry, participants cannot tell that the orchestra was virtual, but the data from ensemble settings show interesting results. The stereo setting is rated as the least wide in perception, followed by the mono setting and orchestral setting. When all the virtual sources are gathered in the center of the stage, the low frequencies are outstanding, which explains why mono setting could have a wide source perception. These low frequencies from cello, bass, and timpani interact with the concert hall from a point that it masked some of the violin's spectrum and boosted the whole orchestra to an extent that the perceived wideness of the sound source is increased. The result of the orchestral setting is expected, which illustrates the belief that wide physical source width produces wide apparent source width. However, the result shows that the circumstance where the physical source is distributed widely does not necessarily outperform the circumstance where the physical source is not wide. It is surely important that listeners need to hear individual signals from individual instruments expanded across the stage, or across the two loudspeakers, but what might be more important is that listeners need to hear an effect that the awareness of discerning the existence of individual instruments does not predominate over the sensory spontaneity of aurally "looking" at a bigger sound picture as a whole that the composer really tries to create. In other words, what listeners first care about in terms of auditory spatial impression led by sound source wideness is that if these individual instruments grouped in an ensemble can serve the need of creating a unique aural image by bringing distinctive spectra into harmony, rather than about telling the locations of each individual instrument. In reality, where players physically sit across the stage to form a whole body of orchestra or small chamber ensemble, the cello and the violin do not have the chance of being fused together. The sound image in a concert hall always set the violin and cello to be separate auditory events when they played together. The virtual orchestra played by wave field synthesis, however, has the ability to take everything to one point which is the mono setting, to enrich the spectrum of string pizzicato as one body, and to boost the tutti by fusing the timpani, cello, and bass in one place.

The Blumlein technique yielded the most outstanding result. The ribbon diaphragm and the radiation pattern helped generate a solid spatial impression that the width of the sound source is rich and wide. Interestingly, the binaural recording without crosstalk cancellation yielded a wide source perception. This may raise the question of the claim that binaural recording will be destroyed by the loudspeaker crosstalk. We need to further verify how much the recording had been destroyed by the crosstalk. As another coincident technique, MS received a fair score as well, but because of the difference in the polar pattern and the diaphragm, the MS did not reach the Blumlein result. The MS emphasized high strings, such as violins, but lacked the support from bass and timpani as well as half of the hall reverberation when compared to Blumlein. The binaural technique did not provide the amount of concentration on bass as the Blumlein did, but because of the omnidirectional microphone capsules, sufficient amount of early and late reflections in all frequency bands are captured, which amplified the spatial balance of the music passage, resulting in a balanced source perception. Compared to ORTF, spaced omni delivered more bass and less sharp high frequencies, resulting in a warmer and wider sounding. In the opposite, ORTF delivered excellent instrument separation in which the attack from distinctive instruments is stressed more than decay, sustain, and release, but this delivery did not help much in boosting the perception on the width of the orchestra. In other words, the ORTF provided more rhythmic sounding, but since the reverberation from mid and low frequencies are not enough for accompanying the direct sound, the auditory perception will not link the ORTF recording to a broad aural image.

4.2. Experiment 2

Based on the opposite result from the first experiment and the second experiment regarding distance, it is reasonable to conclude that for headphone listening, the interaural decorrelation is more likely on determining the perceived width of the sound image. In an internalized sound field where the low frequencies are not physically surrounding the listener and are not directly encountering human body, separately distributed signals are considered first than the mid and low energies. Compared with perceiving the sound pressure of hard-panned signals from two loudspeakers, headphone helped more in integrating the sound pressure of those widely distributed signals into one aural image. Mid and low frequencies are equally important as binaural decorrelation in shaping the wide source perception in headphone, but because the brain treats these low frequencies differently between listening in a physical sound field and listening in an internalized sound field, the auditory spatial impression as well as the perceived width of the sound source will become different aural architectures between loudspeakers and headphone. In headphone listening, the brain tends to claim a bigger auditory image when the sound pressure from those signals is present and more active at the ear drum, which was the case for the recordings made in the front row. In other words, the brain uses more direct sound to judge the width of a performing entity when listening on headphone. In a stereo listening scenario where there is no visual reference on how the music is played by different performers, the brain is blind to the information of the sound sources. In headphone listening, however, the brain is even more blind and more dependent on the loudness of the signal to make a judgment on the width of the sound source. Because of the blindness, the aural architectures between listening through headphone and listening through stereo loudspeakers differ in a way that for headphone, further recording distance increases the depth of the sound source, but not necessarily the width of the sound source, whereas for stereo loudspeakers, further recording distance increases both the depth and the width of the sound source.

The mono setting does not seem to overtake the orchestral setting when played back through headphone. This is caused by the same reason why the recordings in the mid row yielded a decrease in wideness compared to the recordings in the front row when played back through headphone. Besides, as mentioned in the discussion for loudspeaker experiment, the mono setting brings a different spectrum that the mid and low frequencies of the entire orchestra are somehow boosted, affecting listener's emotion greatly while restraining the thinking on localization of individual sound sources. However, in headphone listening, there is no way to physically feel the interfered vibration on the desk and floor by the whole body. Listener's attention will be restrained from emotional feelings in this case, and be paid more on seeking the location of every instrument. Nevertheless, the different source distributions are subtle to display a big difference in the perception of the source width. What this indicates is that the music is written and captured in a way that a change in the position of individual instruments is less likely to replace the perception of the width of the ensemble. Even though the width of the orchestra has been increased or reduced, the number of instruments has been kept constant in all three ensemble sizes. The results from both experiments showed that for a same piece of music that has an unchanged instrumentation and number of performers, a physically narrow ensemble width on the stage will not necessarily deliver a narrow auditory image, and vice versa.

In both experiment 1 and experiment 2, Blumlein and Binaural technique have been recognized as being able to generate a wide spatial impression of the sound source. This supports that mid and low energies, along with diffusive reflections from multiple directions, are equally important for reproducing the performance of a classical orchestra through both headphone and stereo loudspeakers. Certainly, spaced omni provided more bass than ORTF and MS, but it has not been to a point that satisfies the auditory system when compared to Blumlein played back on headphone. The data show that the rankings of recording techniques in terms of how much spatial impression of the sound source they produce are consistent between loudspeaker reproduction and headphone reproduction, except for the position between MS and spaced omni. The reason why spaced omni is able to overtake MS on headphone is that the high density of reflections coming from all the directions are restored, which is lacking on MS. The reasons why spaced omni is way behind MS on loudspeakers are 1) These diffuse reflections don't sustain enough time to arrive at the eardrums of a listener to elicit an emotional response; 2) The bass is more or less attenuated by the rather flat frequency response of the spaced omni microphones so that the listeners neither encountered enough bass for enlarging the auditory spatial impression nor perceived enough crescendo in the music. The MS did not emphasize bass either, but its emphasis on high strings let the listeners hear and feel the dynamic changes in the music passage and a detailed play of crescendo from violins in both stereo and headphone reproductions.

5. CONCLUSION AND FUTURE WORKS

This paper investigated the effect of stereo recording and reproduction techniques and the effect of ensemble sizes on apparent source width. Four stereo recording techniques and one binaural recording technique were evaluated through stereo loudspeaker reproduction and headphone reproduction. Chief findings include 1) The Blumlein technique is the most suitable one among the others for delivering western classical symphonies by satisfying the auditory spatial impression on the width of the sound source and by a perfect capture of the musical dynamics which resulted in an obvious emotional resonance; 2) Interaural decorrelation is the major factor in determining the width of the sound source in headphone listening, whereas sufficient mid and low frequencies plus higher order reflections are considered more for a wide source perception when listening to stereo loudspeakers; 3) ORTF technique is more ideal for chamber ensembles and light music where the heavy and grand sounding made by much low registers are less likely to occur; 4) Regardless of the recording techniques used, for a same music composition that has an unchanged instrumentation and number of performers, changing the physical width of the ensemble do not lead to a significant change of perceived source width.

The technical combination not only galvanized the inventions of new fashions from academia to industry but also energized unqualified collaborations and communications among artists in producing new arts. In a futuristic era of immersive art, where both scientists and artists take a big leap together, several issues must be recognized for further investigations: More stress should be put on a rigorous categorization of the kinds of reproduction approaches in terms of the kinds of arts. The artistic intentions should be divided into two-channel stereophony, surround sound, and wave field synthesis. Similarly, the discussion about auditory spatial impression should be divided into when the performers are in front of the listener and when the performers are surrounding the listener. More documents are needed to clarify which recording technique is the optimal choice that aids a specific compositional intention.

More research is needed to investigate how music with fewer instruments (small in ensemble size, or physical source width) is able to generate a big sound image.

Taking an example of a recent endeavor to spatial audio and spatial hearing made by Stefan Weinzierl et al. [11], more subjective assessments are needed to understand what those recording techniques beyond two-channel stereo recording can bring to the table for artistic creations, and more efforts need to be put on documenting the success and drawbacks of specific recording techniques with specific projects so that the thinking and the sharing will not be only kept at the industries but also at the academia. In addition, in making these spatial recordings as well as auralizations, more high-quality anechoic materials with more music genres and signal types are needed to contribute to the subjective investigations.

6. ACKNOWLEDGMENT

The authors would like to thank Tim Ziemer and Prithvi Ravi Kantan for constructive feedback and helpful remarks on the initial manuscript.

7. REFERENCES

[1] D. Griesinger, "The psychoacoustics of apparent source width, spaciousness and envelopment in performance spaces," *Acta Acust. united Acust.*, vol. 83, no. 4, pp. 721–731, Jul. 1997.

[2] T. Okano, L. L. Beranek, and T. Hidaka, "Relations among interaural cross-correlation coefficient ($IACC_E$), lateral fraction (LF_E), and apparent source width (ASW) in concert halls," *J. Acoust. Soc. Am.*, vol. 104, no. 1, pp. 255–265, Jul. 1998.

[3] T. Ziemer, "Source width in music production: Methods in stereo, ambisonics, and wave field synthesis," in *Studies in Musical Acoustics and Psychoacoustics*. Cham, Switzerland: Springer Nature, 2017, pp. 299–340.

[4] T. Ziemer and R. Bader, "Psychoacoustic sound field synthesis for musical instrument radiation characteristics," *J. Audio Eng. Soc.*, vol. 65, no. 6, pp. 482–496, Jun. 2017.

[5] L. Böhlke and T. Ziemer, "Perceptual evaluation of violin radiation characteristics in a wave field synthesis system," in *Proc. Mtgs. Acoust.*, Boston, MA, USA, Jun. 2017, paper 4pMU11.

[6] F. Toole, *Sound Reproduction: The Acoustics and Psychoacoustics of Loudspeakers and Rooms*. New York, NY, USA: Routledge, 2017.

[7] J. Ahrens and S. Spors, "Wave field synthesis of a sound field described by spherical harmonics expansion coefficients," *J. Acoust. Soc. Am.*, vol. 131, no. 3, pp. 2190–2199, Mar. 2012.

[8] N. Hahn, F. Winter, and S. Spors, "2.5D local wave field synthesis of a virtual plane wave using a time domain representation of spherical harmonics expansion," in *Proc. 23th Int. Congr. Acoust.*, Aachen, Germany, Sep. 2019, pp. 1132–1139.

[9] J. Meyer, *Acoustics and the Performance of Music: Manual for Acousticians, Audio Engineers, Musicians, Architects and Musical Instrument Makers*. New York, NY, USA: Springer Science & Business Media, 2009.

[10] C. Böhm, D. Ackermann, and S. Weinzierl, "A multi-channel anechoic orchestra recording of Beethoven's Symphony No. 8 Op. 93," *J. Audio Eng. Soc.*, vol. 68, no. 12, pp. 977–984, Jan. 2021.

[11] D. Ackermann, J. Domann, F. Brinkmann, J. M. Arend, M. Schneider, C. Pörschmann, and S. Weinzierl, "Recordings of a loudspeaker orchestra with multichannel microphone arrays for the evaluation of spatial audio methods," *J. Audio Eng. Soc.*, vol. 71, no. 1/2, pp. 62–73, Jan. 2023.

[12] F. Brinkmann, L. Aspöck, D. Ackermann, S. Lepa, M. Vorländer, and S. Weinzierl, "A round robin on room acoustical simulation and auralization," *J. Acoust. Soc. Am.*, vol. 145, no. 4, pp. 2746–2760, Apr. 2019.

[13] M. Barron and A. H. Marshall, "Spatial impression due to early lateral reflections in concert halls: The derivation of a physical measure," *J. Sound Vib.*, vol. 77, no. 2, pp. 211–232, Jul. 1981.

[14] A. H. Marshall and M. Barron, "Spatial responsiveness in concert halls and the origins of spatial impression," *Appl. Acoust.*, vol. 62, no. 2, pp. 91–108, Feb. 2001.

[15] J. Käsbach, M. Marschall, B. Epp, and T. Dau, "The relation between perceived apparent source width and interaural cross-correlation in sound reproduction spaces with low reverberation," in *Proc. 39th German Annu. Conf. Acoust. (DAGA)*, Merano, Italy, Mar. 2013.

[16] N. Zacharov, "Subjective appraisal of loudspeaker directivity for multichannel reproduction," *J. Audio Eng. Soc.*, vol. 46, no. 4, pp. 288–303, Apr. 1998.

[17] H. Wittek, S. Kerber, F. Rumsey, and G. Theile, "Spatial perception in wave field synthesis rendered sound fields: Distance of real and virtual nearby sources," in *116th Audio Eng. Soc. Conv.*, Berlin, Germany, May 2004, paper 6000.

[18] P. Gutiérrez Parera, J. J. López Monfort, and E. Aguilera Martí, "On the distance perception in spatial audio system: A comparison between wave-field synthesis and panning systems," *Waves*, vol. 6, pp. 51–59, 2014.

[19] F. Völk, M. Straubinger, and H. Fastl, "Psychoacoustical experiments on loudness perception in wave field synthesis," in *Proc. 20th Int. Congr. Acoust.*, Sydney, Australia, Aug. 2010, paper 43.66.Cb.

[20] S. Spors, H. Wierstorf, M. Geier, and J. Ahrens, "Physical and perceptual properties of focused virtual sources in wave field synthesis," in *127th Audio Eng. Soc. Conv.*, New York, NY, USA, Oct. 2009, paper 7914.

[21] H. Wierstorf, A. Raake, M. Geier, and S. Spors, "Perception of focused sources in wave field synthesis," *J. Audio Eng. Soc.*, vol. 61, no. 1/2, pp. 5–16, Jan. 2013.

[22] N. Peters, J. Braasch, and S. McAdams, "Recording techniques and their effect on sound quality at off-center listening positions in 5.0 surround environments," *Can. Acoust.*, vol. 41, no. 3, pp. 37–49, Dec. 2013.

[23] X. Zhao and Z. Meng, "Perception of sound source localization in wave field synthesis," *J. Phys.: Conf. Ser.*, vol. 1739, Jan. 2021, Art. no. 012011.

[24] J. Goebel, "The empac high-resolution modular loudspeaker array for wave field synthesis," in *Proc. 23th Int. Congr. Acoust.*, Aachen, Germany, Sep. 2019, pp. 1148–1155.