

Document downloaded from:

<http://hdl.handle.net/10251/192669>

This paper must be cited as:

López Monfort, JJ.; Gutiérrez-Parera, P.; Savioja, L. (2020). Effects and applications of spatial acuity in advanced spatial audioreproduction systems with loudspeakers. *Applied Acoustics*. 161:1-8. <https://doi.org/10.1016/j.apacoust.2019.107179>



The final publication is available at

<https://doi.org/10.1016/j.apacoust.2019.107179>

Copyright Elsevier

Additional Information

The following article appeared in the Journal "Applied Acoustics" (Elsevier)
J. J. Lopez, P. Gutierrez-Parera, and L. Savioja, "Effects and applications of spatial acuity in advanced spatial audio reproduction systems with loudspeakers," Appl. Acoust., vol. 161, p. 107179, Apr. 2020.
doi: 10.1016/j.apacoust.2019.107179
and may be found at <https://doi.org/10.1016/j.apacoust.2019.107179>
© <2020>. This manuscript version is made available under the CC-BY-NC-ND 4.0 license
<https://creativecommons.org/licenses/by-nc-nd/4.0/>

Effects and Applications of Spatial Acuity in Advanced Spatial Audio Reproduction Systems with Loudspeakers

Jose Javier Lopez^{a,*}, Pablo Gutierrez-Parera^a, Lauri Savioja^b

^a*Universitat Politècnica de València, Spain*

^b*Aalto University, Espoo, Finland*

Abstract

Spatial audio reproduction systems using loudspeakers produce coloration effects at high frequencies due to spatial interference between loudspeakers, in both those based on panning and those based on field synthesis. As a response to this problem and in order to reduce coloration, this paper studies the feasibility of an alternative approach where high-frequencies are reproduced from a single loudspeaker, with a different direction from that of its panned low-frequency counterpart. Listening tests are conducted to investigate the localization and quality of the source in the case that frequencies higher than 1.5 kHz are reproduced from a different direction than the low frequencies. In this context, the human ability to discriminate the spatial direction of low/high frequency bands and the error in the perceived direction of arrival for different separation angles is evaluated and quantified. The resulting data has been analyzed with ANOVA, providing significant results that allow us to establish a threshold in the angular separation of the high and low frequency parts where subjects do not perceive source location artifacts. The term *just noticeable band splitting angle* (JNBSA) is defined and introduced. It represents the minimum angle of separation between high and low frequencies from which the listener starts to perceive artifacts in the reproduction of a sound source using loudspeakers.

Keywords: spatial audio; auralization; Wave-Field Synthesis; Vector Based Amplitude Panning (VBAP); loudspeakers; spatial acuity

1. Introduction and background

Spatial sound systems have been evolving in recent years due to public demand for better and more immersive audiovisual systems. We are witnessing the transition from channel-based to object-based systems [1, 2], at the same time that better and more sophisticated playback systems based on a large number of loudspeakers are becoming popular. Two main loudspeaker approaches

*Corresponding author

Email address: jjlopez@dccom.upv.es (Jose Javier Lopez)

are used to reproduce immersive sound; panning based and soundfield synthesis systems. The first group includes classic surround systems, being 5.1 [3] the most popular and widespread. The growing number of loudspeakers in surround systems generally results in better quality and localization, giving rise to 6.1, 7.1 systems and also systems that include elevation such as 10.2 and 22.2 [4, 5, 6] for the cinema. As a generalization, the Vector-Base Amplitude Panning (VBAP) [7] method provides a mathematical tool to apply the panning law for 3D loudspeakers distributed in any configuration. The second group of systems tries to synthesize a realistic soundfield based on the physical equations of sound propagation. The two main methods are Wave-Field Synthesis (WFS) [8] and Higher Order Ambisonics (HOA) [9]. As in the first group, the more loudspeakers employed, the better synthesis obtained. In the case of WFS, this is because the spatial aliasing is reduced if loudspeaker density is increased, and in the case of HOA, it is because a higher order raises the minimum amount of loudspeakers needed.

In both cases, the premise “the more, the better” (loudspeakers) is applied, but this approach obviously increases the cost and complexity of the systems. It would, then, be interesting to have more available studies to determine what number of loudspeakers should be adequate to achieve a quality experience in each system, as well as to define the minimum number of loudspeakers to preserve this. In response to this problem, this paper studies the feasibility of an alternative approach to reducing the number of loudspeakers in dense loudspeaker set-ups. The high frequencies are reproduced from a single loudspeaker, with a different direction from that of its panned counterpart low frequency. This approach may reduce coloration artifacts, but might, in turn, lead to misplacement or degraded quality of the sound source. Therefore, listening tests are conducted to investigate the location and quality of the source in the case where high frequencies are reproduced from a different direction than low frequencies.

Before addressing the research aspects, some background review is presented. First an introduction to panning systems and their coloration effects. In subsection 1.3, a previous study related to the perception of simultaneous sources from different directions is revisited. And after that, the motivation and hypothesis of this work are explained. Then, in section 2 the experiment approach is explained. Sections 3, 4 and 5 describe the three perceptual tests performed and show the results obtained. Section 6 summarizes the conclusions and outlines some possible applications and future work.

1.1. Panning Systems

Amplitude panning through multichannel loudspeaker systems is the most widely used method for spatial sound reproduction.

The simultaneous use of two separate loudspeakers allows the creation of a phantom source, which is perceived as a substitute for the real sources. This effect is called *tracking sum* [10] and is assumed to create binaural signals very similar to those created by the actual sources. There are objections to this explanation. In his *association model* [11], Theile argues that overlapping signals from different speakers do not create localization, but rather that the signals

from the two speakers give two different localization stimuli that melt together into a phantom source after a complex psychoacoustic process. Leaving aside the open questions about the nature of the phantom sources, it is a fact that the laws of stereo panning and recording techniques with stereo microphones have been widely used to achieve spatial localization in all kinds of good quality products. Practical experience and a variety of formal research claim that the optimal configuration for two speaker stereo is an equilateral triangle with the listener located at one of the vertexes. If the amplitudes of the two channels are controlled properly, it can produce a resultant phase and amplitude differences of continuous sounds that are very close to those experienced with natural sources, thus giving the impression of virtual or phantom images anywhere between the left and right speakers [12]. This is the basis of the Blumlein's stereo system, invented in 1931. It is often assumed that the mixing coefficients used in stereo synthesis relate to the angle θ_n of the virtual source perceived n by the panning law of the tangent [10]:

$$\frac{\tan \theta_n}{\tan \theta_0} = \frac{a_{1n} - a_{2n}}{a_{1n} + a_{2n}} \quad (1)$$

where θ_0 is the angle of separation of the speakers.

With the introduction of surround sound systems employing five or more speakers around the listener, the panning law has also been used to place the sources in the 360° azimuth range. In this case, the panning law is applied between each pair of speakers to create phantom sources. As commented before, more loudspeakers provide better localization and perception, but there is no precise mathematical model that relates them. Perhaps one of the most interesting works on the subject is the one carried out for the VBAP, thanks to the results of a study by Pulkki which analyzes the performance of panning methods when extended to 360° or even to 3D elevation [13, 14, 15].

The findings of these works support the idea that the panning law creates more stable phantom sources at the front, but performance is reduced with side, rear or elevated sources. These latest results are important in the design of spatial sound systems, especially for systems where listeners always look in the same direction (e.g. in cinemas or frontal stage shows) and lateral and rear concepts make sense. In this paper we use these prerequisites for the design of our experiments and extend them with our results.

1.2. Coloration Effects

One of the negative aspects of panning systems and other systems that use multiple speakers to recreate the sound field is the alteration of the spectrum of sources (coloration), especially at high frequencies. This effect is due to the sum of signals out of phase with different speakers. For example, although the listener of a stereo system aims to be centered on the axis between the two speakers symmetrically, the distance from each speaker to each ear is slightly different. This difference of acoustic paths causes the sum of the signals from the two speakers to be out of phase. Therefore, a pronounced effect of comb

filtering occurs beyond a certain frequency. These frequencies are above 1.5 kHz where the interaural distance begins to be comparable to its wavelength.

The coloration can be detected by the human ear. However, these effects have not been subjectively studied in depth, despite the existence of some works on the subject [16]. Depending on the reproduction system and the musical material employed, the effects may be more or less perceived by the listener. In any case, and regardless of the severity of these effects, it would be desirable to minimize or even remove them from the reproduction system.

The origin of the problem is the emission of high frequencies from two or more different points, therefore a direct solution would be not use panning techniques with conflicting frequencies and emit them solely from a single speaker. Obviously, this would result in a restriction of putting the virtual sources only in the places where we had a speaker, limiting the spatialization sound at first. This solution does not apply in the case of stereo playback set-ups (two speakers) or 5-channel surround systems, where the separation between speakers is very large [3]. However, in systems that use many distributed speakers, such as Wave-Field Synthesis [8], Higher-Order Ambisonics (HOA) [17] or VBAP systems with many channels, a solution of this kind might be approached. There would still, of course, be an error equivalent to the separation in the positioning of the speakers, but it would be less than in systems with a small number of speakers.

Given that the coloration effect is very small or negligible at lower frequencies, we might consider systems which would split the reproduction of low and high frequencies. Thus, low frequencies would be reproduced using panning or soundfield synthesis techniques (VBAP, HOA or WFS) and high frequencies, where the effects of coloration appear, would be reproduced from a single speaker. In this manner, some discrepancy could occur in the perception of the direction of arrival of the frequencies of the source, with the low and high frequencies coming from different positions. However, the hypothesis of this paper is based on the premise that if the difference is not very large, the sound source can be perceived as coming from a single point instead of two different ones.

1.3. Concurrent Minimal Audible Angle (CMAA)

There are few studies related to the resolution of the human ear to perceive two different but simultaneous sources over time from different directions. One of the most interesting studies on this “spatial acuity” was conducted by Perrott [18]. There was a previous concept of *minimum audible angle* (MAA) [19] used as a basis for his study, that identifies the ability of human hearing to detect an angular difference of the same source from two different directions when played sequentially. In his study, Perrott uses simultaneous acoustic events with the hypothesis that the new *minimal audible angle for concurrent events* (CMAA) is likely to be greater than the MAA because of the added difficulty to the human auditory system of having to separate sounds.

In the experiment, two speakers playing different signals were separated at different angular distances. The signals were two tones of different frequencies and the subject had to guess in each case whether the higher frequency tone was

on the right or left. The tones were reproduced randomly so that 50% of correct answers for a certain angle meant that the listener could not distinguish where each tone was. The smallest angle at which this phenomenon began to occur was called CMAA. Another parameter studied was the frequency separation between the tones. The closer the frequencies were, the more difficult it was to distinguish them. The minimum angle is dependent on the frequency difference between the tested tones. For frontal listening this angle is around 10 degrees and can decrease to 5 if the frequency difference is small. The position of the sound sources with respect to the median plane was also found to affect the perception of the CMAA. As expected, the angle was bigger in lateral positions than in frontal ones. At a lateral azimuth position of 40 degrees the minimum angle raises to 16 degrees and up to 40 degrees with a lateral source position of 67 degrees.

1.4. Work motivation

The previously described work suggests the possibility that sounds emitted from different positions with little separation in space, can be perceived by the human auditory system as coming from one single point. The signals employed in that study were two tones, but what would happen if more complex signals were used? for example, two people speaking from different points.

On the other hand, it might be interesting to study the effects of separating low and high frequencies at a certain angular distance. Not just two tones of different pitch but low and high frequency bands of a single sound. This could have an application in spatial sound systems where several speakers are usually employed. Besides, if the signals were related in some way, harmonically for example, we could hypothesize that distinguishing the direction of arrival of each sound would be even more difficult for human hearing. Therefore, a new term, *just noticeable band splitting angle* (JNBSA), can be defined and introduced. It represents the minimum angle of separation between the high and low frequencies from which the listener begins to perceive artifacts (Direction of Arrival (DOA) error, source width), in the reproduction of a sound source by means of loudspeakers.

Using a playback system of this type, it would be possible to emit high and low frequencies from different points without any detection of difference by the listener, who would perceive the sound as coming from a point source. As commented above, the sum of the sound signals reproduced by different speakers produces coloration effects at high frequencies. If we could use such a reproduction system in which the low frequencies were reproduced by more than one speaker (using VBAP, WFS or HOA techniques) and the high frequencies were reproduced by a single speaker in a different position, we can hypothesize that the listener could perceive them as coming from one single point. If we can confirm this fact, it would open new paths to solve, or at least to improve, spatial sound reproduction systems.

2. Experiment approach and justification

To conduct the experiment, a set of speakers were arranged in a circular arc shape and angularly spaced at equal intervals. The array was made with 19 self-powered Genelec speakers with virtually flat response, with a 5" woofer and a 1" tweeter. Each of them was placed separated 5 degrees in an arc of 2 meters of radius, between -45° and 45° on either side of the listener (Figure 1). To evaluate the lateral sound and from behind, the listeners were rotated in the seat. Sound sources from all directions can then be evaluated but without the reflections that a complete circular array would produce.

Three signals were used in the experiment: brown noise, voice and trombone sound. These signals were split into two frequency bands, below and above 1.5 kHz. Third-order Butterworth low-pass and high-pass filters were used to make the division. This pair of filters sum correctly in amplitude and phase without producing any coloration in the transition band.

The speakers were placed in an acoustically equipped room to meet the criteria mentioned in ITU BS.1116-3 [20] for this type of sound reproduction experiments. Therefore, the acoustic conditions of reverberation time, background noise and setup arrangement were within the criteria of this standard recommendation.

The different tests making up the experiment were performed by 9 to 11 subjects (two women and the remaining men), between 22 and 42 years old. All of them were skilled researchers with experience in critical listening, and performed the listening tests in a double-blind manner.

The perceptual location of sound sources can be strongly influenced both by the position of the listener with respect to the source and by the movement of the listener and/or sources [21, 22, 23]. In this work the perception of static sound sources was studied. For this reason, the subjects who carried out the perceptual tests were instructed not to move during the execution of the tests, and in addition, this was controlled by a supervisor by means of visual inspection.

3. Experiment 1. Left/Right distinction

3.1. Description

Following Perrott's studies [18], the aim of this experiment was to find out whether listeners are able to locate where the high frequency is, when the reproduction of the low and high frequencies of a single sound source are separated from different angles.

The task proposed to the participants was to locate the high frequency and the question presented to them was: "Where is the high frequency?" with two possible answers: left or right. A graphical interface was developed to perform the test, as shown in Figure 2, which subjects used to record their responses to each stimulus. They were presented with their low and high frequency separated with 6 offset angles: 10, 15, 20, 25, 30 and 40 degrees. The setup allowed 5 degrees of separation, but in a preliminary test it was very obvious to the

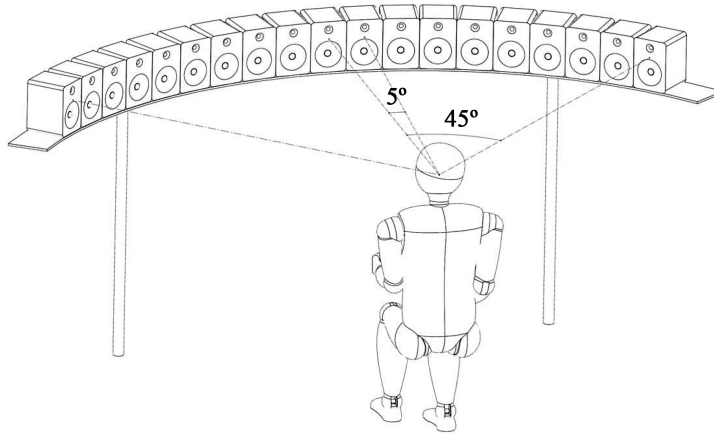


Figure 1: Loudspeakers set-up with 5° angular spacing.

researchers that it was impossible to distinguish, so the 5 degree separation was omitted to reduce the duration of the listening test.

The participants were placed with 3 different orientations with respect to the center of the array: frontal, rotated laterally 90 degrees and backward 180 degrees. (see top of Figure 3). The intention was to check the frequency separation effect of sound sources from all directions. Because of that, the possible directions of arrival of the sound, considering the middle point angle in between the low and high frequencies, were -30, 2.5, 5, 17.5, 20, 65, 90, 115, 155, 190 and 205 degrees in the horizontal plane. The participants had to indicate where they perceived the high frequency, on the left or on the right of the sound stimulus. In the case of a listening orientation rotated 90 degrees, the same coordinate system as the frontal orientation was used to indicate the answers, as an extension of the frontal orientation. In the case of a 180-degree backward listening orientation, the “left” and “right” positions were inverted to be consistent with the subject’s egocentric reference. Figure 4 shows a diagram with an example of a middle point angle and an offset angle. Three types of sound were used: brown noise, voice and trombone sound. Table 1 summarizes all the stimuli characteristics. Different combinations of the previous characteristics were employed, resulting in 39 stimuli presented three times and randomly to each participant. 9 participants performed the test. To avoid disorientation, reference sounds were available for participants to listen to at any time through buttons on the interface.

3.2. Results

In this test, participants were asked to indicate where they perceived the high frequency band of the sound, on the left or on the right. Figure 3 shows the percentage of correct answers, for each listening angle orientation and for each high and low frequency separation offset angle. It indicates the rate of correct

Table 1: Summary description of Test 1 Left/Right distinction

High-low frequency separated	Task: Locate the high frequency		
Three listening orientations:	frontal	lateral	back
11 middle point angles:	-30, 2.5, 5, 17.5, 20	65, 90, 115	155, 190, 205
6 offset angles:	10, 15, 20, 25, 30	10, 20, 30, 40	10, 20, 30, 40
3 types of sound:	Voice, Trombone, Brown noise		
39 stimuli total	9 participants		

localization of the high frequency band. Then, 50% means that the listener cannot distinguish the position of the high frequencies and therefore 50% can be considered as the baseline of correct answers. For frontal sources, listeners start to distinguish offset angles of 10 degrees but only slightly above the baseline (64%). This corresponds to the limit where it starts to be statistically significant for one person, according to [24]. As we are averaging the results of different people, it is even less significant. As the angular offset increases, the percentage of correct answers grows as expected, but surprisingly, even with the maximum separation angle (30 degrees) the detection rate is only about 80%. However, for lateral and back sources, high frequency localization is very bad and there is no clear tendency. Due to this lack of discrimination for lateral and back sources the second experiment was carried out just for frontal sources. No significant effect was found between the type of sound and the number of correct answers.

4. Experiment 2. Angle of arrival

4.1. Description

Since experiment 1 indicates that the offset angle between high and low frequencies should be very large to be perceived, it seems conceivable that the possible perceived error in the direction of arrival of the split source would be small. Therefore, a second experiment was prepared to check how much error or deviation angle is perceived in the direction of arrival of a sound source, when the high and low frequencies are separated.

The task of the listeners was to indicate from which direction they perceived the sound, aided by some marks on an acoustically transparent curtain placed in front of the loudspeakers. The answer was to be indicated by means of a graphical user interface (Figure 5). In this case, 6 offset angles of low-high frequency separation were employed: 0, 5, 10, 15, 20, 25 degrees, and just one single orientation of listening position with respect to the center of the array, that is, frontal orientation. The respective position of the high and low frequencies on the left or right was randomized. As in experiment 1, participants should listen to different stimuli sounds coming from different directions of arrival, in this case we considered directions from 9 middle point angles: -35, -32.5, -30, 0, 2.5, 5, 15, 17.5, 20 degrees. The same 3 types of sound were used: brown noise,

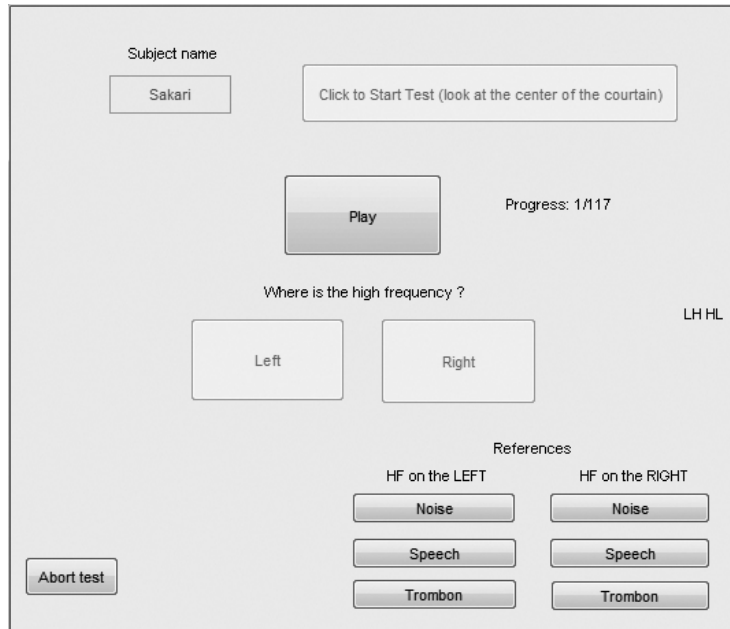


Figure 2: Experiment 1. MATLAB user interface for the subjective test left/right distinction.

voice and trombone sound. Thus, a total of 54 stimuli (18x3 types of sound) were presented to 10 participants, the same nine as in the previous test plus one other subject. Table 2 summarizes the characteristics of all stimuli.

Table 2: Summary description of Test 2 Angle source separated High-low frequency separated Task: Locate the source direction

One listening orientation:	frontal
9 middle point angles:	-35, -32.5, -30, 0, 2.5, 5, 15, 17.5, 20
6 offset angles:	0, 5, 10, 15, 20, 25
3 types of sound:	Voice, Trombone, Brown noise
54 stimuli total	10 participants

4.2. Results

The intention of this test was to identify the perceived error or deviation in the direction of arrival of the sound when the high and low frequencies are spatially split. These are shown in Figure 6, where the perceived deviation in degrees is indicated for each offset angle (also in degrees) of separation between low-high frequencies. The perceived deviation is considered as the absolute value of the difference between the answered angle and the middle point angle, meaning the deviations to the left (negative values) and to the right (positive values) together.

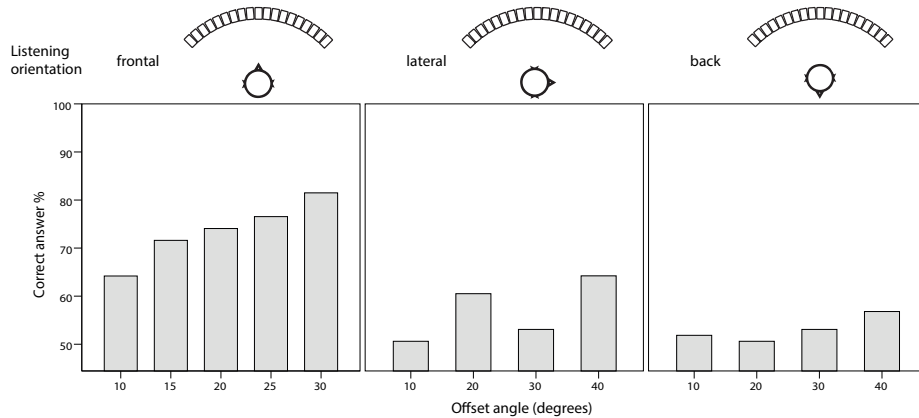


Figure 3: Experiment 1. Percentage of correct answers (left or right) for each of the listening orientation angles and offset angles between low-high frequencies.

The graph shows that although low and high frequencies are emitted from the same loudspeaker (offset angle=0) listeners have an underlying location error of 2.5 degrees. This can be considered as a reference value of the error (or absolute deviation) in the perceived direction of arrival, and be used to quantify the other values in relative terms. At 5° of offset angle there is a similar deviation error of about 2.5 degrees, therefore the effect of separating 5 degrees is virtually non-existent. For 10° and 15° of angular separation, listeners perceive the sound source in the middle of the low and high frequencies without significant distinction with respect to the point source case (0° offset angle). Furthermore, between 0° and 15° there is a difference of just 1.14 degrees (15° offset: 3.33° deviation, 0° offset: 2.19° deviation). Then, up to 15° of offset angle, the perceived error deviates by less than $\pm 0.6^\circ$ from the reference value, and it can be considered as very small. Only beyond an offset angle of 20 degrees of separation is there a more significant error deviation over 2 or 3 degrees above the reference value.

An Analysis of Variance (ANOVA) confirms that the offset angle is a significant factor in the perception of the direction ($F=7.577$, $df=10$, $p<0.05$). However, the type of sound is irrelevant in the evaluation of the direction of arrival, according to the significance value $p=0.170$, as well as the middle point angle between the real sources (from -35° to 20° azimuth), for the listening oriented towards the front studied here.

5. Experiment 3. Source width

5.1. Description

Results from experiments 1 and 2 can be complemented to further evaluate the spatial sensation of the listeners when the two frequency bands are separated. It would be interesting to find out if this split of frequency bands produces any

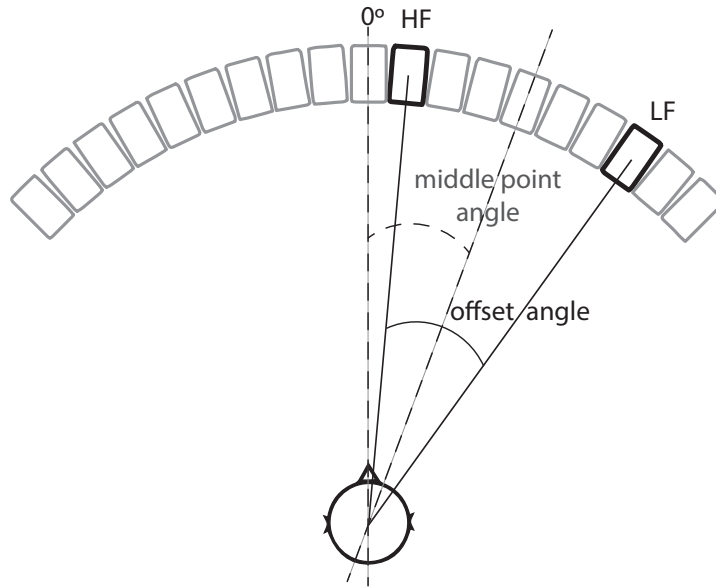


Figure 4: Example of middle point angle and offset angle, as used in the perceptual tests. HF and LF indicate the speakers that reproduce the high and low frequencies respectively.

noticeable defect, pursuant to the width or dispersion of the perceived sound source.

To study this, participants were asked to focus on the perceived width of the different sounds and to indicate their perceptions using a graphical interface (Figure 7), on a gradual scale from 1 (widest) to 5 (narrowest). The low and high frequencies were this time separated 7 offset angles: 0, 10, 15, 20, 25, 30 and 40 degrees. Participants listened oriented at three different angles: frontal, lateral and backward orientation, as in experiment 1 (see top of Figure 3). Then, the sounds coming from 18 directions of arrival (considering the middle point angles between the low-high frequencies) were presented: -32.5, -30, 2.5, 5, 17.5, 20, 60, 65, 85, 90, 110, 115, 150, 155, 185, 190, 200 and 205 degrees of the horizontal plane. The same three types of sound were used again: brown noise, voice and trombone sound. 144 stimuli were presented randomly to each of the 11 participants who performed the test, the same ten as in the previous test plus one other subject. Table 3 summarizes the different combinations of characteristics of the stimuli. Reference examples of the widest and narrowest sounds of each type, reproduced around the central loudspeaker of the array, were available to be listened to at any time.

5.2. Results

Using the graphical interface the participants evaluated the perceived width of the sound sources. The perceived width of the sources (1-wide to 5-narrow) for each offset angle (in degrees) is shown in Figure 8. There is a tendency in

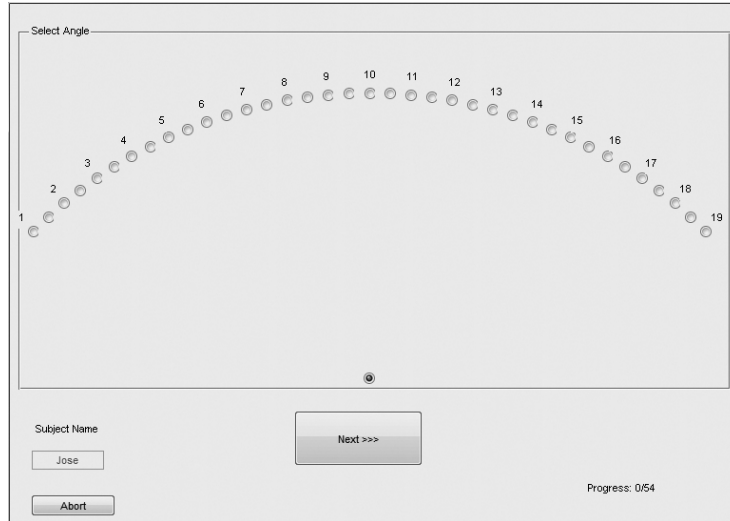


Figure 5: Experiment 2. MATLAB user interface for the subjective test angle of arrival.

Table 3: Summary description of Test 3 Perceived source width

High-low freq separated	Task: Rate source width		
Three listening orientations:	frontal	lateral	back
18 middle point angles:	-32.5, -30, 2.5, 5, 17.5, 20	60, 65, 85, 90, 110, 115	150, 155, 185, 190, 200, 205
7 offset angles:	0, 10, 15, 20, 25, 30	0, 10, 20, 30, 40	0, 10, 20, 30, 40
3 types of sound:	Voice, Trombone, Brown noise		
144 stimuli total	11 participants		

the perception of the width of the source that shows a little wider perception as the offset angle increases.

With an ANOVA we can see that the type of sound is significant in the perception of the width ($F=11.80$, $df=4$, $p<0.05$) but this is mostly due to the perception of the brown noise, that is generally perceived slightly wider than the other type of sounds (Figures 9 and 10). Until 15° of angular separation between low-high frequencies, the voice is perceived narrower than the other sounds, and over an offset angle of 20° the voice sound is abruptly perceived as wider (Figure 9). In summary, the results indicate that with offset angles of up to 10° or 15° , even 20° depending on the type of signal, the perception of the source does not begin to become significantly wider and less punctual.

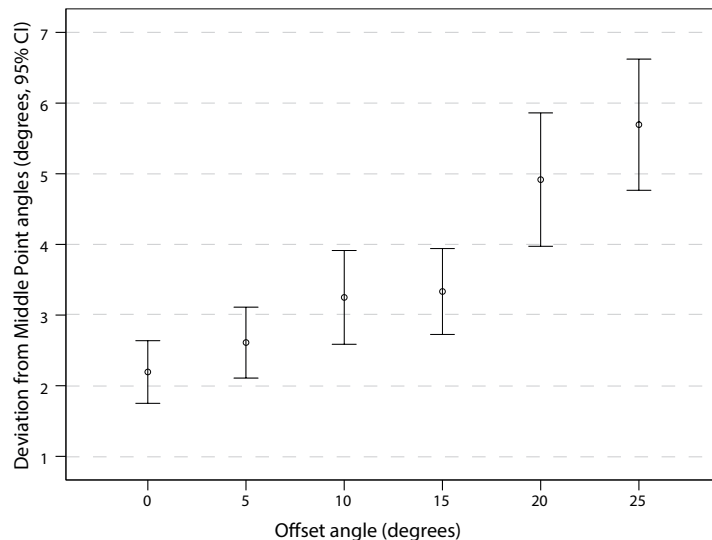


Figure 6: Experiment 2. Perceived deviation (in degrees) with respect to the middle point angles (between the two real sources) for each offset angle (separation of high and low frequencies).

6. Conclusions and applications

6.1. Conclusions

In this work we have studied the perceptual effects of reproducing high and low frequencies of one source from two different spatial positions. High frequencies were considered starting from 1.5 kHz and were reproduced in different loudspeaker positions than the low frequencies. In particular, it has been determined what the limit angles are in order that this separation is not appreciated by the listener and the sound source seems to be only one source and not two different sources. To this end, a series of subjective experiments have been carried out with an array of loudspeakers around the listener, allowing for a precise and feasible positioning.

In a first test a study has been made of what the limit angle is from which the listener is not able to discern on which side the high frequency is and where the low frequency is. It has been concluded that the orientation of the listening position is a key factor in the discrimination of the different frequencies. For frontal sources, listeners start to slightly distinguish separations of 10 degrees but just occasionally. As the offset angle increases, the percentage of correct answers increases as expected, but never reaches 100%. For lateral and backward oriented listening discrimination is very poor.

A second test has also been carried out to check the perceived direction of arrival when high and low frequencies are not reproduced from the same point in space. It has been verified that this angle of deviation is around the middle angle

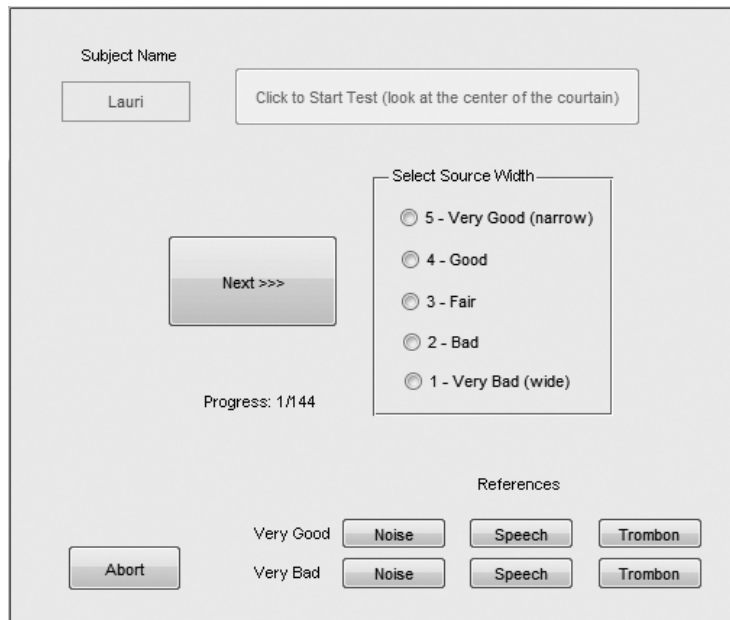


Figure 7: Experiment 3. MATLAB user interface for the subjective test source width perception.

between the high and low frequencies. Up to 15° of offset angle the perceived error is less than $\pm 0.6^\circ$, and it can be considered as very small. Only beyond 20 degrees of separation is there a significant error deviation.

In a final test a study has been carried out of how much the perception of source width artificially increases when separating the bands of high and low frequency. It has been found that with angular offsets of up to 10° or 15° , even 20° depending on the type of signal, the perception of the source does not begin to become significantly wider and less punctual.

As discussed in the introduction, these experiments focused on static listeners and/or sound sources. This is, then, a limitation of this work, given that to determine the dynamic angles of listening with movement, other experiments would be needed. In addition, the limited sound stimuli used in the tests may also constrain the results, especially due to their short duration and the plainness of the content.

In [18], the term *concurrent minimum audible angle* (CMAA) was introduced by Perrott. In this work a new term, *just noticeable band splitting angle* (JNBSA) is defined and introduced. It represents the minimum angle of separation between high and low frequencies from which the listener starts perceiving artifacts (DOA error, source width), in the reproduction of a sound source using loudspeakers. Considering the overall results of the three tests, a JNBSA of 15° can be taken as a conservative value for frontal sources and much bigger for lateral and back sources. Due to the resolution of the set-up and the results

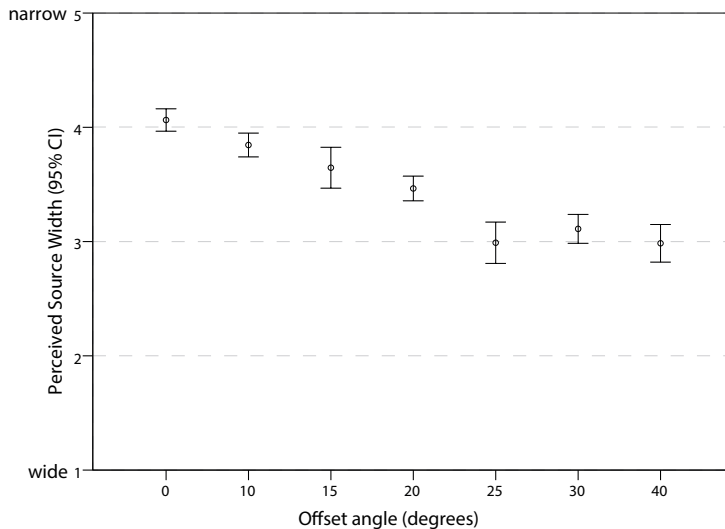


Figure 8: Experiment 3. Perceived width of the sources (1-wide to 5-narrow) for each offset angle (between low-high frequencies).

of the perceptual tests, 15° of separation between high and low frequencies is the highest offset angle value that does not yet produce significantly incorrect spatial locations.

6.2. Applications and future work

The results of this work therefore allow us to study in the future the feasibility of an alternative approach for the reproduction of spatial sound, in which high frequencies are reproduced from a single loudspeaker coming from a different direction than that of the low-frequency counterpart, which can be reproduced with more sophisticated systems. This approach can be used to develop advanced sound systems using multiple loudspeakers with simplified set-ups according to perceptual considerations.

In particular, these findings can be applied in the improvement of WFS reproduction, especially in the reduction of staining artifacts that occur in high frequencies when issued from more than one speaker.

One of the drawbacks of the WFS is the spatial aliasing that occurs at high frequencies because of the separation between loudspeakers [25]. In [26] the OPSI method based on phantom sources was presented, which despite reducing aliasing adds other problems. In [27] a sub-band WFS system is proposed, where the low frequencies are reproduced by means of field synthesis, but the high frequencies are emitted by a single loudspeaker. This system presents the problem that if the listener is not in the center of the array there may be an angular error between the low and high frequency parts. The effect of this error was not evaluated and neither was the possibility of it confusing the perceived

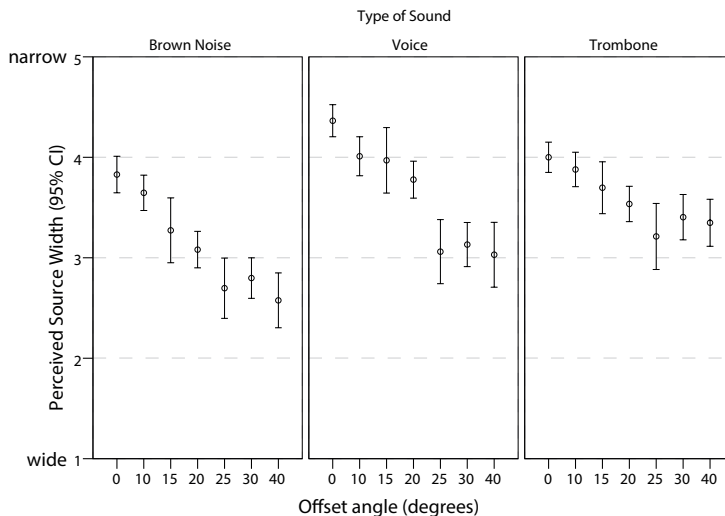


Figure 9: Experiment 3. Perceived width of the sources (1-wide to 5-narrow) for each offset angle and type of sound.

position of the source. Thanks to the study introduced here, systems such as the one presented in [27] can be tested to see if they exceed detectability limits.

In addition, due to the large number of loudspeakers that the WFS needs if a high aliasing frequency wants to be achieved, these systems are not widely used in the field of professional sound reinforcement. The current trend, partly reinforced by the introduction of object-based sound systems, is to install multiple loudspeaker arrays but with more affordable separations between 1 and 3 meters, depending on the size of the listening area. In these systems (related to different brands), the reproduction of the sources in each location is achieved by combining aspects of panning and VBAP such as the treatment of the amplitude of the signal that is sent to each speaker, with others that are characteristic of the WFS such as the delay. Hybrid systems are created empirically without a convenient objective evaluation of the result. These systems necessarily create spatial aliasing above a certain frequency. By applying amplitude panning and delays only at low frequencies and emitting the high frequency from a single loudspeaker, these aliasing effects could be reduced while keeping the perceived position of the source in the proper place, if the criteria calculated in this work are met.

Future work plans are to set up one of these hybrid systems (halfway between WFS and VBAP) and subjectively re-evaluate the advantages achieved and the angle error obtained. In addition, with such a system, it would be interesting to test a wider range of realistic sound stimuli with complex characteristics of time and frequency, as well as to test several sound sources simultaneously. This would reveal whether masking effects could condition the listening and detection of the JNBSA. Moreover, 15 degrees was determined as a conservative value of

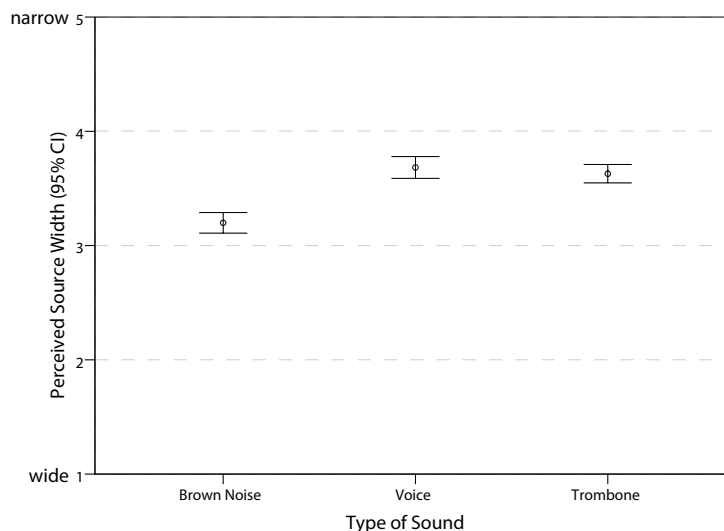


Figure 10: Experiment 3. Perceived width of the sources (1-wide to 5-narrow) for each type of sound.

JNBSA, since the experiments have been limited to the resolution of the array set-up, which has 5 degrees separation of loudspeakers. Considering that 20 degrees of angular offset has been found to be significantly detectable, the zone between 15 and 20 degrees could be explored with more precision using an array with closer loudspeakers, and even obtain some kind of threshold bounds within this zone.

Acknowledgments

The Spanish Ministry of Economy, Industry and Competitiveness supported this work within the framework of the TEC2015-68076-R and RTI2018-097045-B projects, as well as the Aalto University with a visiting professor grant. The authors also thank Sakari Tervo for the support in the experimental procedure, and Philip W. Robinson for the suggestion of bibliographic references.

References

- [1] European Broadcasting Union EBU, ETSI TS 103 223 - V1.1.1 - MDA; Object-Based Audio Immersive Sound Metadata and Bitstream (2015).
- [2] T. Sporer, J. Liebetrau, T. Clauss, Evaluation of object-based audio. What is the reference?, *The Journal of the Acoustical Society of America* 141 (5) (2017) 3464–3464. doi:10.1121/1.4987191.
- [3] ITU-R BS. 775-3, Multichannel stereophonic sound system with and without accompanying picture (2012).

- [4] T. Holman, *Surround Sound: Up and Running*, Second Edition, Focal Press, 2008.
- [5] K. Hamasaki, K. Hiyama, R. Okumura, The 22.2 multichannel sound system and its application, in: *Audio Engineering Society 118th Convention*, Barcelona, Spain, 2005, p. Paper 6406.
- [6] ITU-R BS.2051-2, *Advanced sound system for programme production* (2018).
- [7] V. Pulkki, Virtual Sound Source Positioning Using Vector Base Amplitude Panning, *Journal of Audio Engineering Society* 45 (6) (1997) 456–466.
- [8] A. J. Berkhout, D. de Vries, P. Vogel, Acoustic control by wave field synthesis, *The Journal of the Acoustical Society of America* 93 (5) (1993) 2764–2778. doi:10.1121/1.405852.
- [9] M. A. Gerzon, *Periphony: With-Height Sound Reproduction*, *Journal of Audio Engineering Society* 21 (1) (1973) 2–10.
- [10] A. Blumlein, *Improvements in and relating to sound transmission, sound recording and sound reproduction system* (British Patent Specification 394325) (1933).
- [11] G. Theile, *Über die Lokalisation im überlagerten Schallfeld (Localization in the Superposed Sound Field)*, Ph.D. thesis, Technical University of Berlin (1980).
- [12] D. M. Leakey, *Some Measurements on the Effects of Interchannel Intensity and Time Differences in Two Channel Sound Systems*, *The Journal of the Acoustical Society of America* 31 (7) (1959) 977–986. doi:10.1121/1.1907824.
- [13] V. Pulkki, M. Karjalainen, *Localization of Amplitude-Panned Virtual Sources I: Stereophonic Panning*, *Journal of the Audio Engineering Society* 49 (9) (2001) 739–752.
- [14] V. Pulkki, *Localization of Amplitude-Panned Virtual Sources II: Two- and Three-Dimensional Panning*, *Journal of the Audio Engineering Society* 49 (9) (2001) 753–767.
- [15] O. Santala, V. Pulkki, *Directional perception of distributed sound sources*, *The Journal of the Acoustical Society of America* 129 (3) (2011) 1522–1530. doi:10.1121/1.3533727.
- [16] V. Pulkki, *Coloration of Amplitude-Panned Virtual Sources*, in: *Audio Engineering Society 110th Convention*, Audio Engineering Society, Amsterdam, The Netherlands, 2001, p. Paper 5402.

- [17] D. Malham, Higher order ambisonic systems for the spatialisation of sound, in: International Computer Music Conference Proceedings (ICMC 1999), 1999, pp. 484–487.
- [18] D. R. Perrott, Concurrent minimum audible angle: a re-examination of the concept of auditory spatial acuity., *The Journal of the Acoustical Society of America* 75 (4) (1984) 1201–1206. doi:10.1121/1.390771.
- [19] A. W. Mills, On the Minimum Audible Angle, *The Journal of the Acoustical Society of America* 30 (4) (1958) 237–246. doi:10.1121/1.1909553.
- [20] ITU-R BS.1116-3, Methods for the subjective assessment of small impairments in audio systems (2015).
- [21] J. Blauert, *Spatial hearing : the psychophysics of human sound localization*, 2nd Edition, MIT Press, Cambridge, 1997.
- [22] H. Sato, H. Sato, M. Morimoto, Y. Nakai, Localization of intermittent sound with head movement: Basic study on optimum temporal characteristics of acoustic guide signals, *Applied Acoustics* 101 (2016) 58–63. doi:10.1016/J.APACOUST.2015.08.003.
- [23] J. J. Lopez, A. Gonzalez, 3-D audio with dynamic tracking for multimedia environments, in: 2nd COST-G6 Workshop on Digital Audio Effects, 1999.
- [24] H. Burstein, Approximation Formulas for Error Risk and Sample Size in ABX Testing, *Journal of the Audio Engineering Society* 36 (11) (1988) 879–883.
- [25] G. Theile, H. Wittek, M. Reisinger, Potential wavefield synthesis applications in the multichannel stereophonic world, in: AES 24th International Conference on Multichannel Audio, 2003, p. Paper 35.
- [26] H. Wittek, OPSI: Optimised Phantom Source Imaging of the high frequency content of virtual sources in Wave Field Synthesis, Tech. rep., Internal CARROUSO Paper (2002).
- [27] J. J. Lopez, S. Bleda, B. Pueo, J. Escolano, A Sub-band Approach to Wave-Field Synthesis Rendering, in: Audio Engineering Society 118th Convention, Barcelona, Spain, 2005, p. Paper 6403.