# SUBJECTIVE EVALUATION OF AUDITORY SPATIAL IMAGERY ASSOCIATED WITH DECORRELATED SUBWOOFER SIGNALS

*William L. Martens*

Multimedia Systems Lab
University of Aizu
Fukushima-ken 965-8580, Japan
`wlm@u-aizu.ac.jp`

## ABSTRACT

Although only a single subwoofer is typically used in two-channel and multichannel stereophonic sound reproduction, the use of two subwoofers enables manipulation of low-frequency interaural cross-correlation (IACC), and this manipulation is particularly effective in producing variation in auditory spatial imagery. In order to document this variation objectively, a series of listening experiments were executed using a set of stimuli generated at five correlation values and presented in two reproduction modes. Both modes used two subwoofers, but in one of the reproduction modes identical signals were applied to the two subwoofers. The results of both exploratory and confirmatory listening experiments showed that the range of variation in both perceived auditory source width (ASW) and perceived auditory source distance (ASD) is reduced when negatively correlated signals are not reproduced at low frequencies. Global dissimilarity judgments were made for this set of ten stimuli in an exploratory study designed to reveal the salient perceptual dimensions of the stimuli. A subsequent confirmatory study employed a two-alternative forced-choice task in order to determine how identifiably different the stimuli were with respect to the two perceptual attributes revealed in the exploratory study, those two attributes being ASW and ASD. The implications of these findings for loudspeaker-based spatial auditory display are discussed.

## 1. INTRODUCTION

This investigation asked the question, “Can reproduction of two-subwoofer signals enable increased variation in auditory spatial imagery?” But the goal of this work is perhaps more clearly stated in the converse inquiry, “Does the use of only one subwoofer signal in stereophonic sound reproduction truly provide an adequate range of auditory spatial images?” The results of a previous study by the author [1] showed that two-channel manipulation of sound reproduced at low-frequency (below 250 Hz) can have a strong impact upon the auditory spatial attributes such as perceived auditory source width (ASW) and perceived auditory source distance (ASD). The research results reported here provide an independent replication of this finding for two-subwoofer reproduction of low-frequency information limited to signals in the range 20 – 120 Hz, the standard range for the subwoofer channel published in the International Telecommunication Union (ITU) recommendation ITU-R BS 775-1 [2] for multichannel stereophonic sound systems. In addition to extending previously reported listening experiments to include alternative stimulus conditions, the present study also included more extensive psychophysical tests of the changes in auditory spatial imagery that are enabled by the reproduction of two subwoofer signals.

The research question could be narrowed to a very simple comparison, focusing upon the changes that occur when the low-frequency portion of a two-channel program is reproduced via a single channel. By listening to two otherwise identical audio reproductions, one with two subwoofer signals and one with one subwoofer signal, the question could be asked, “What changed?” But the answer to this question is far from simple, as there are many characteristics of auditory imagery that can change, including loudness, clarity, tone coloration, and a number of spatial attributes that are readily identified. Answering the question is further complicated by the choice of representative program material for the comparisons. Rather than attempting to answer the research question most broadly, the current study focussed upon a tightly constrained set of synthetic test stimuli. Consequently, a concern over the generality of the results presented herein is warranted. Nonetheless, the test material is representative of one class of sonic stimuli that has been important in film sound production and reproduction, the percussive blast exemplified by the crack of a snare drum or the report of canon fire. Varying the two-channel sound reproduction method while holding the input sound source constant facilitates reliable subjective reporting on the changes in resulting auditory spatial imagery.

The central idea of this research was to compare the auditory spatial imagery that is produced under two reproduction conditions that were matched to each other in every respect except for that one factor. Contrary to the implication of the initial question asked here, the factor chosen for differentiating between the two reproduction conditions was not whether one or two subwoofers were used; rather, the factor that differed between conditions was whether one or two low-frequency **signals** were applied to the inputs of an audio reproduction system that always employed two low-frequency **speakers**. Manipulating only the audio signals to be presented virtually eliminated potential confounding factors such as differences in loudness and tone coloration. Thus, the null and alternative hypotheses to be tested regarding attributes of auditory spatial imagery, when comparing between the one-subwoofer-signal and the two-subwoofer-signal listening conditions, can be stated most generally as follows: In comparison to listening to reproduction of a single low-frequency signal, reproduction of two different low-frequency signals …

**$H_0$:** makes no identifiable difference in spatial imagery.

**$H_A$:** creates readily identifiable differences in spatial imagery.

Of course, reproduction using two low-frequency signals versus a single low-frequency signal can create an easily detectable difference in auditory spatial imagery. That is not in question here. Rather, the question is whether the comparison reveals any substantial differences in perceptual attributes that are readily identified by human listeners. The consequence of using an identical reproduction system for comparing auditory spatial imagery under two reproduction conditions is that strong, valid conclusions can be drawn. The phenomena on which this paper reports have proven to be replicable for a variety of stimulus conditions and reproduction system configurations. Strong support can be provided for the primary conclusion, that two-subwoofer reproduction enables increased variation in the perceived sound *character* [3] of auditory spatial imagery. Accepting this conclusion does not, however, logically lead to a conclusion that the sound *quality* associated with two-subwoofer reproduction will be improved. The results of the experiments reported here do not bear directly on the sound *quality* issue related to subjective preference, which Letowski [3] suggested should be differentiated from sound *character* in that the former is more evaluative while the latter is purely descriptive. The relation between availability and usefulness of variation in auditory spatial imagery requires further investigation that could potentially build upon current findings. Nonetheless, the results have implications for the faithful transmission of imagery created via popular effects processing and audio production techniques.

Stereophonic reproduction using just two extended-range loudspeakers makes the decorrelated presentation of low-frequency information possible, which can produce interesting results using nothing more than common-practice production techniques. In contrast, common-practice audio production for multichannel stereophonic sound systems must target a wider variety of loudspeaker arrangements, many of which employ a subwoofer system to reproduce low-frequency signals at a relatively high level. For example, theatrical film sound reproduction assumes a 10-12 dB boost for the subwoofer signal [2]. In contrast, multichannel stereophonic sound media distributed for home systems cannot assume that the domestic consumer's sound system will include a subwoofer. Therefore, the main channels must carry normal-level, low-frequency sound, and the main-channel loudspeakers must be able to stand on their own in reproducing a satisfactory program without the inclusion of a subwoofer. The purpose of the subwoofer channel is then to extend the low frequency content of the program as an enhancement for low-frequency effects presented at higher levels than the other loudspeakers would be required to handle. Nonetheless, contrary to the recommendations of the International Telecommunication Union (ITU) [2], the combined low-frequency components contained in the main channels often end up in the subwoofer channel. It is possible, therefore, in tests of domestic multichannel sound systems that the audio programs selected from available commercial film titles will show no difference between one-subwoofer and two-subwoofer reproduction (this result was obtained in one admittedly preliminary investigation [4]).

Of course, if the single-channel subwoofer signal from a test audio program is processed to produce two somewhat decorrelated subwoofer signals, two-subwoofer reproduction of the program can produce an audible difference in spatial imagery. This in and of itself is not particularly interesting. As the use of two subwoofers enables the presentation of a very different stimulus, it is not surprising that the difference is detectable. On the other hand, there are two questions to ask about such detectable differences that are truly interesting. The most obvious question to ask is whether listeners prefer the two-subwoofer reproduction to the one-subwoofer reproduction. Another interesting question to ask is why listeners have such preferences, in effect, examining what forms the basis for one system being preferred over another. Answering the latter question requires an understanding of the perceptual distinctions that the human listener is able to make. Thus stated, the problem is one for which methods of psychological measurement may be applied to reveal the perceptual dimensions along which stimuli may differ.

Instead of focusing on subjective preferences for the spatial imagery created using systems capable of reproducing two low-frequency signals, this study was focused on the nature of the spatial imagery such systems make available. This choice is based upon the assumption that understanding the perceptual responses to the experimental stimuli has intrinsic value that goes beyond global qualitative preferences. It is also the case that investigations of how stimuli differ from one another are useful in the design of studies of preference. Indeed, the most effective studies of preference employ sets of stimuli that allow differences to be distinguished, rather than stimuli that differ only negligibly and therefore cannot generate differences in preference. Stimulus selection informed by the results of identification and discrimination experiments also has potential benefits for interpretation of preference data.

Thus this study attempted to identify the perceptual dimensions underlying judgments about stimulus differences, rather than attempting to observe differences in preference without mapping out the territory. The territory of interest is the range of auditory spatial percepts that can be elicited in two-subwoofer reproduction. In the larger context of overall evaluation of sound system performance, this information can be quite important, especially in dealing with difficulties due to the existence of large individual differences in preference for one spatial image over another [5]. Subjective preference studies can also be quite complicated, since preferences will depend upon the application to a great extent. In multichannel sound for the home theater, for example, some scenes will require big, broad spatial images while others will require an enclosed, intimate sound like that associated with small spaces, such as a hard-walled cafe. The possibility of creating a virtual surround effect in two-loudspeaker reproduction is also desirable, and benefits greatly from full-range loudspeakers that are able to reproduce a two-channel program at relatively low frequencies. In more exacting virtual acoustics applications, such as loudspeaker-based auralization, it is clearly a disadvantage to exclude the whole category of close-range sound sensations associated with negative cross correlations in low-frequency bands.

### 1.1. Negative Interaural Cross Correlation (IACC)

Negative cross correlation produces an interesting range of spatial effects that are particularly sensitive to differences in low frequency reproduction. In 1983 Kurozumi and Ohgushi [6] established that a direct relationship exists between the cross-correlation coefficient of two-channel audio signals and auditory spatial-image attributes width and distance. When the cross-correlation coefficient approached the extreme value of -1, they observed changes in the spatial image that included decreasing width and decreasing distance (or increasing intimacy). The research presented in this paper replicates some of their work, but introduces a new factor: a comparison of two low-frequency reproduction modes, one of which reproduces a high positive correlation in the subwoofer frequency range, the other of which faithfully reproduces a selected amount of decorrelation in this low frequency band (nominally, below 120 Hz). It is particularly at negative $CC$ values that differences between the reproduction modes are expected, since at positive cross-correlation values the two-subwoofer signals do not differ so greatly. If the sum of two different, but positively-correlated subwoofer signals is delivered as an identical subwoofer signal to both of two-subwoofers, no great change in the result is expected in comparison to the result given the two-subwoofer presentation of those somewhat decorrelated signals. On the other hand, the consequences of presenting positively correlated low-frequency signals are substantial for broadband stimuli exhibiting otherwise negative cross correlation. Because the low-frequency signals that reach the listener's ears exhibit a positive cross correlation in this case, the low-frequency components of a broadband stimulus will be strongly centered in the listener's perceived spatial image, despite the fact that the higher-frequency portion of program is spatially diffused. How substantial a difference this makes in human perception is the focus of the research described herein.

An early motivation for this study came from informal observation of differences between stereo loudspeaker reproduction before and after the introduction of a single subwoofer. The spatial imagery associated with reproduction using a pair of extended-range loudspeakers sounded broader and deeper than that associated with single-subwoofer-based reproduction. But this informal comparison relied too much upon memory for spatial imagery. To test the hypothesis that the use of a single subwoofer had a negative impact upon the spatial image, controlled listening experiments are required. Also, because the suggested use of two subwoofers is contrary to conventional practice, particularly careful tests should be made, with substantial attention to methodological details.

### 1.2. How Many Subwoofers are Needed?

Before beginning to address the question of whether two subwoofer signals should be used, there is an interesting related question about how many subwoofers are needed to provide good reproduction of a single low-frequency signal (i.e., when subwoofers are coherently driven), especially when considering coverage for a listening area that might be intended for more than one simultaneous listener. A recent study by Welti [7] investigated this question for a grid of possible listening positions, both through simulation and actual measurements of the response for one, two, or more coherent subwoofers. It was concluded that four subwoofers are enough to get the best results, but that two subwoofers located at wall midpoints are nearly as good as four, and also provide very good low frequency support. Of course, the metrics used in the Welti [7] study were based on strictly physical measurements. In contrast, the questions asked in the current study are focussed upon the perceptual differences that might be observed with one or more subwoofers, and in particular, whether one subwoofer signal is enough to satisfy perceptual criteria. What should be emphasized here, however, is that two-subwoofer reproduction may be recommended even when only a single subwoofer signal is presented, in order to excite room modes in a more "balanced" manner.

An argument based upon human perceptual limitations can be made supporting the conclusion that only a single subwoofer channel is needed in two-channel and multichannel stereophonic sound reproduction systems. Because it is difficult to identify the direction of a low-frequency sound source in a room, it has been thought that only one low-frequency signal need be delivered for adequate reproduction of auditory spatial imagery. Though there is support for the premise that directional perception can be confused in rooms (due to the room's low-frequency modal behavior, etc.), the premise does not lead necessarily to the conclusion that including two subwoofers in a system is without value. Indeed, the results of the author's recent research [1], in which low-frequency, inter-channel phase was systematically varied, show that having two channels of spatial information at low frequency enables better control over several subjective features of the resulting spatial image, but most particularly apparent source width (ASW) and apparent source distance (ASD).

Griesinger [8] has also emphasized the importance of decorrelated low-frequency sound (below 300 Hz) in creating optimal spaciousness and listener envelopment (LEV) in multichannel stereophonic sound. In particular, his research on identifying the critical features of the reproduced spatial image associated with an outstanding concert hall led to placement of two low-frequency drivers directly on either side of the listener (at ±90 deg. azimuth). He concluded that lateral separation at frequencies as low as 60 Hz is "vital to world class envelopment."

The dominance of lower-frequency decorrelation over higher-frequency decorrelation in generating increases in ASW has also been documented by Morimoto, et al. [9]. Though closely related, LEV and ASW must be clearly distinguished, as they describe distinct components of the auditory spatial image [10]. While the width of the direct sound image is properly associated with the temporally- and spatially-fused sound source, envelopment is the fullness of a sound image around a listener, excluding the spatial image relating to ASW. By definition, then, envelopment refers to the percept affected by the spatial distribution of indirect sound. It would not be appropriate, however, to regard envelopment as the perception of indirect sound, since indirect sound influences both LEV and ASW. Since a good portion of the early indirect sound (arriving within 25 ms of the direct sound) is typically perceptually fused with the direct sound, this portion will contribute more to widening the spatial image of the source (increasing ASW). Ultimately, the only proper definitions of the terms spaciousness, envelopment, and ASW include no reference to the physical stimulus at all. These terms describe perceptual attributes that may be predicted by measurements of the physical stimulus, but their characterization must be wholly psychological.[1]

### 1.3. Spatiotemporal Fusion of Decorrelated Signals

When questions are asked about the width of the temporally- and spatially-fused auditory image of a sound source, the listener must form an overall impression of ASW. If a broadband sound stimulus is not fused into a single auditory event, but rather is segregated into multiple images extending over different frequency ranges,

---

[1] Indeed, rigorous psychophysical study requires clarity regarding these perceptual dimensions of the stimuli, and such clarity has largely been missing in spatial hearing research (A notable exception to this observation is found in the work of Morimoto (see, for example [11]). Other work, well summarized in Rumsey's recent book [12], has provided a good foundation for continuing inquiry into this difficult problem, which includes, for example, differentiation between ASW and the apparent image widths for the ensemble and for the environment.

asking for a single report of ASW is not appropriate. One safeguard against this possibility is to use narrow band stimuli, such as the $\frac{1}{3}$ octave band noise stimuli employed in the Morimoto, et al. [13] study of the effects of sound pressure level (SPL) and the frequency characteristics of the IACC on ASW. As both IACC and SPL may be measured for each individual $\frac{1}{3}$ octave band, variation in ASW could be examined in relation to variations in these two parameters for a set of experimental stimuli that extend over two adjacent $\frac{1}{3}$ octave bands. Their conclusion was that ASW depends upon the frequency distribution of both SPL and cross-correlation. Though the overall impression of ASW can be predicted by the overall IACC, it is more adequately predicted by measures sensitive to variation in correlation and level over frequency.

A means for promoting spectral fusion of a synthetic stimulus into a single auditory event is to apply an amplitude envelope with a shape over time of the exponential decay of a percussive acoustical event. This was the strategy adopted in the author's [1] previous study of two-subwoofer spatial imagery, and was adopted again for the experiments reported here. When two sustained, decorrelated broadband-noise stimuli are reproduced by loudspeakers, spectral fusion is not guaranteed. In fact listeners typically report the segregation of the stimulus into multiple images with different tone coloration (see, for example, Plenge's 1972 study [14]). In the authors previous study of two-subwoofer imagery [1], an informal "picture-drawing" method of documenting changes in spatial imagery was employed in an attempt to determine whether single fused spatial images were experienced by experimental listeners. This method is not without precedent in related research [14] [15], and has been elaborated recently into what promises to be a useful non-verbal technique for summarizing what listeners find most salient in the auditory spatial imagery associated with reproduced sound [16].

So in a preliminary experimental listening session, the listeners in the current study were asked indicate the size, shape, number, and spatial locations of the auditory spatial imagery they heard in association with the sound stimuli presented. This informal method is admittedly sensitive to variation in task variables. For example, the experimenter's expectations can easily bias listeners toward one type of response or another, this bias being established while listeners receive instructions on what to draw, and also by the implied range of acceptable drawings communicated via the diagrammatic forms on which the listeners were asked to draw. Nonetheless, the results from this drawing method proved useful, and showed that single fused images were associated with virtually all stimuli presented to this study's listeners.

## 2. METHODS

### 2.1. Subjective Evaluation of Auditory Spatial Imagery

Blauert [17] used the term "spatial impression" (in German, Raumeindruck) to describe the listener's concept of the type and size of an actual or simulated space. This is distinct from the term "spaciousness" (in German, Räumlichkeit), which is the single perceptual dimension modulated primarily by the ratio of indirect to direct sound levels [2]. The term "spatial impression" is intended as a broader concept to encompass both spaciousness and reverberance [19]. The author proposes to use the term "auditory spatial imagery" as the broadest, most comprehensive term to describe for a given sound stimulus the combined auditory response on all the perceptual dimensions of spatial hearing taken together. Subjective evaluation of reproduced sound is even broader than this. Toole [20] lists three overall ratings of sound system performance that should be formed by integrating over all perceptual dimensions to which the listener can attend. These include fidelity, pleasantness, and spatial quality. This study focused only on the perceptual dimensions relating to spatial quality. In the standard questionnaire for subjective evaluation developed by Toole [20], spatial quality ratings included:

1. definition of sound images
2. continuity of the sound stage
3. width of the sound stage
4. impression of distance/depth
5. abnormal effects
6. ambiance, spaciousness, and reverberation
7. perspective (eg., 'you are there,' 'close, but looking on,' and 'outside looking in').

In normative studies of auditory spatial imagery, standardized questionnaires of this sort have proven very useful especially when listeners are well trained [21]. In testing specific experimental hypotheses, however, bias-free measures of subjective response are valuable, such as those provided via forced-choice tasks. In the research reported here, pairs of spatial sound stimuli were presented to listeners in order to obtain two general types of judgments. For each pair of spatial images, both direct ratings of global stimulus dissimilarity and forced-choice judgments between stimuli were required of well-trained listeners.

The first formal psychophysical method required holistic rather than analytic listening. Here, pairwise comparisons of a set of experimental stimuli were presented under a number of conditions, and listeners were asked to judge the overall perceptual distance between the two stimuli making up each pair. Such responses are typically termed global dissimilarity judgments. Of course, individual subjects may differ in how they form judgments of global dissimilarity, and so a refined method for doing a *weighted* MDS analysis [22] that takes such individual differences into account is to be recommended. This paper teaches the use of individual differences scaling [23] analysis as a powerful means for deriving an interpretable representation of the dimensions underlying reported inter-stimulus dissimilarities obtained from a potentially inhomogeneous group of subjects, each of which may place a different *weight* upon each of the perceptual dimensions. As the first stage of this study was exploratory in nature, the dissimilarity analysis provided useful information about which perceptual attributes were most salient for the particular set of stimuli presented.

The second formal psychophysical method employed in this study, based upon forced-choice judgments, always required the selection of one out of two possible alternatives, hence the term two-alternative forced-choice (2AFC). In one 2AFC task, listeners were required to make an absolute discrimination between two stimuli, using whatever means enabled that discrimination. In another 2AFC task, listeners were required to make an identification of which of two stimuli dominated the other in terms of some perceptual attribute (either the perceived width or the perceived distance of the auditory image of the sound source, ASW and ASD).

---

[2] A direct relation between subjective spaciousness and indirect–to–direct ($i/d$) ratio is observed for the soundfields of large halls [18], but the relation is modulated by other factors, particularly strongly in smaller spaces in which a good deal of indirect sound arrives quickly (within the first 25 ms).

Note that questions about ASW were focussed upon the spatial extent of the auditory image on the left–right axis, while questions about ASD were not focussed upon spatial-image extent. Rather, questions about ASD related to the egocentric distance of the auditory image (i.e., the perceived range from the center of the listener's head to the center of the auditory image). An alternative attribute not investigated here focuses upon the spatial extent of the auditory image along the close–far, or radial axis, and is better termed auditory source *depth*.

### 2.2. Stimulus Preparation

The sound source used in stimulus preparation for this study was a percussive noise burst that was created by applying an exponentially decaying amplitude envelope to a band of white noise with a low-pass cut-off of 4 kHz. At a duration of one second, the sound of this source resembled that of a reverberated gunshot that might appear in a film soundtrack.

The manner in which cross correlation (CC) was manipulated for the experimental stimuli was a variant of the method described by Blauert and Lindemann [15] This method employed three independently generated noise sources that could be inverted and mixed in varying proportions to generate two-channel signals ranging from extreme negative CC values to CC values near zero. Plenge [14] also used a 180 deg. phase shift to generate similar stimuli, but by a method requiring only two noise sources. Because IACC is most validly measured at the listener's ears, some researchers use a different parameter for reporting the cross-correlation between the signals presented to the loudspeakers. For example, Plenge [14] used the ratio of the two summed noise levels as the parameter. In this paper, the maximum of the normalized cross-correlation function [24] calculated on the signals presented to the loudspeakers is used to provide a formal index for the similarity of the two signals before experimental manipulations. The IACC measured at the listener's ears will of course vary as a function of these manipulations. For this reason, the parameter used in the presented results is termed simply CC.

The conventional approach to connecting a single subwoofer into a stereophonic sound system is to sum the left- and right-channel signals below the subwoofer crossover frequency, and to deliver high-pass-filtered versions of these two signals to their respective loudspeakers. The alternative approach examined here keeps the left- and right-channel signals separate in the subwoofer frequency range by using separate subwoofers. Sound stimuli presented using the conventional approach will differ from those presented using an alternative approach in several ways that make a comparison between these two conditions less than straightforward. For example, the direction(s) from which the low-frequency signals arrive cannot be matched between the conditions, and the relative loudness from combined subwoofer output compared to that of a single subwoofer is difficult to control. Therefore, the conventional approach using a single subwoofer system was simulated using two subwoofers, each receiving an identical signal. The spatial imagery resulting from two laterally-positioned subwoofers of course can differ from that resulting from a single subwoofer centered on the listener's median plane; however, tighter experimental control was achieved using the two-subwoofer simulation of the single-subwoofer approach. The benefit for the experiment was that only one factor differed between the experimental condition and the control condition, that factor being the application of identical versus disparate signals to the subwoofer inputs.

In pilot studies, two types of stimulus generation for the control conditions were examined. In the first case, the left and right signals were summed to create a single signal to be applied to the two subwoofer inputs. This approach, though based upon conventional practice, presented a significant problem for reproduction level under stimulus conditions involving negative IACC values. In the extreme case, at IACC=-1, the two signals would cancel each other entirely, and under less extreme conditions, the output level was reduced considerably. Though transaural reproduction of binaural recordings can produce similar results [25], this situation does not mimic conventional production and reproduction practices, and therefore a second method of stimulus generation for the control condition was also examined.

In this second type of control stimulus preparation, only the (arbitrarily selected) right-channel signal was applied to inputs of the two subwoofers. In this case, the loudness of the subwoofer's outputs was matched between the experimental and control conditions, and the relatively flat magnitude response over frequency was also maintained (only differences in the relative phase of low-frequency components was manipulated by the decorrelation algorithm employed).[3] So, to summarize, the full-bandwidth, two-channel experimental condition had as its control condition the reproduction of a single low-frequency signal via the same two subwoofers that were used in the experimental condition. The higher-frequency signals applied to the satellite loudspeaker inputs did not differ between these two conditions. Note, however, that the digital synthesis of the experimental stimuli used filters that could vary the cutoff frequency for low-frequency stimulus preparation, and that the crossover frequency of the subwoofer systems was set to 120 Hz, the highest value within the range of values recommended for surround sound systems (that range being from 80 Hz to 120 Hz).

It should be asked whether negative decorrelation at such low frequencies is common in reverberant soundfields. Tohyama, et al. [27] showed that if the point-to-point cross correlation is measured between two microphones set about ear-distance apart in a reverberant room, the coefficient is typically near 1.0 at 100 Hz, but falls to near zero at around 1000 Hz Above this frequency, the typical cross correlation coefficient will fluctuate above and below zero to a roughly equal extent. If the cross correlation is measured in a diffuse field using the KEMAR coupler microphones, the first zero crossing in IACC is reached at a lower frequency than for the free-field microphone pair. However, relatively large positive IACC values are to be expected for diffuse reverberant soundfields below the 120 Hz cut-off frequency of the subwoofer signals used in this study. So the large negative IACC values presented here are quite uncommon. This is a problem for arguments based on realism, but not for standard stereophonic production techniques, which often create large phase differences at low frequency. The standard panoramic potentiometer (pan pot) can create large level differences between low frequency signals presented to a two-channel loudspeaker system, but such level differences cannot be observed at the listener's ears unless the loudspeakers are very close to the head (for example, a single sound source must be within a few inches of the head for a 20 dB interaural level difference to be observed at 50 Hz [28]). In the current study, the loudspeakers were placed 1.5 m from the listener, and so it is only phase differences that could actually be reproduced in the subwoofer signals arriving at the listener's ears.

---

[3] Minimal passband ripple was confirmed using a Type 2012 B&K Audio Analyzer, as suggested in [26].

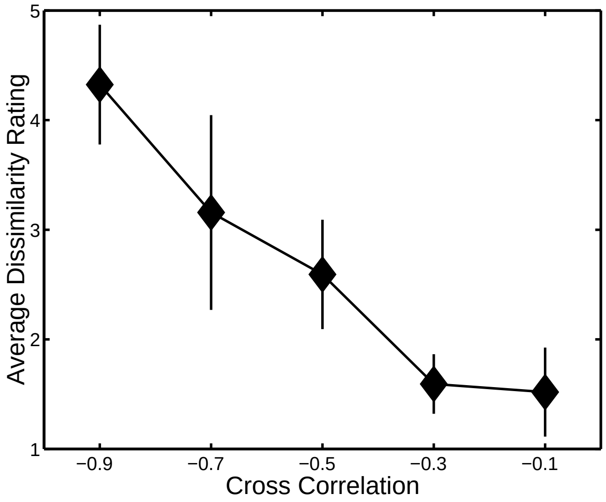

Figure 1: Average (constrained) dissimilarity ratings for the spatial imagery associated with the two-subwoofer signal versus that associated with the one-subwoofer signal for five CC values. The vertical bars through the plotting symbols extend one standard deviation above and below the average ratings for four listeners.

### 2.3. Stimulus Presentation

Sound stimuli were presented in a large anechoic chamber located at the University of Aizu, Japan. It was decided to evaluate the impact of decorrelated low-frequency reproduction on auditory spatial imagery under such controlled laboratory conditions in order to establish baseline performance for subsequent tests in more typical listening environments. The computer-processed audio samples were digitally transfered to a Sony DTC-ZE700 DAT recorder. The outputs of the DAT recorder were applied to the inputs of a Sony MU-A201 4-channel amplifier that powered two medium-sized studio monitors (JBL Control 8SR) situated in the anechoic chamber at a 30 deg. angle to the left and right of the listener's median plane. In contrast to the author's previously reported "two-subwoofer" experiments [1], however, the subwoofers in the current studies were not positioned at the same azimuth angle as the studio monitors; rather, the two large, active-servo subwoofers (Yamaha YST-SW500) were positioned further to the sides than the monitors, at ±90 deg. azimuth (as recommended by Griesinger [8] also for optimal listener envelopment). In addition, this lateral placement of subwoofers at wall midpoints was found by Welti [7] to provide nearly the best results for "balanced" reproduction of coherent subwoofers signals, and also to provide very good low frequency support. It should be pointed out as well that typical anechoic chambers are not strictly anechoic at the lowest frequencies, but rather present room modes based upon their size just as do more reverberant spaces.

### 2.4. Listeners

Four listeners (three male and one female) ranging in age from 22 to 41 years, participated in all exploratory and confirmatory experiments reported here. Each had participated in previous experiments that presented sound stimuli varying in spatial imagery.

## 3. LISTENING EXPERIMENTS

### 3.1. Dissimilarity Judgments

#### *3.1.1. Constrained Pairwise Comparisons*

The listener's task in this first exploratory experiment was to report on the overall perceptual dissimilarity between sound stimuli presented in pairs. They were instructed to rate on a scale from one to five the global difference between two spatial images, without respect to particular details. They were to give a rating of "1" if the spatial images associated with each of the two stimuli were virtually identical, and they were to give a rating of "5" when they heard spatial images that were most dissimilar. The stimuli were presented at five cross-correlation values ranging from $-.9$ to $-.1$. In each trial, a pair of signals were presented at the same negative CC value, but in different reproduction modes: one subwoofer signal vs. two subwoofer signals. A total of 18 ratings were collected for each of five pairs of stimuli in a single session of 90 trials (with randomized stimulus order). Each of four listeners completed two sessions of 90 trials. The average dissimilarity ratings at each CC value for stimuli in the two reproduction modes are are shown in Figure 1. Clearly, reproduction mode results in the biggest perceptual differences at the largest negative CC values. In order to get a more complete picture of the perceptual differences between the ten stimuli presented here, a comprehensive set of comparisons was made.

#### *3.1.2. All Pairwise Comparisons*

All pairwise comparisons of ten stimuli were rated in a single session of 90 trials. Each pair was presented twice in a randomized order, and the order within each pair was reversed upon the second presentation of that pair. The global dissimilarity judgments were averaged across the two presentations in an attempt to reduce the effect of stimulus order on the listener's ratings. The resulting matrix of averaged dissimilarity judgments was completed by the same four listeners who completed the above-described constrained pairwise comparisons, and so four distinct estimates of the perceptual distances between all ten stimuli could be made.

The obtained dissimilarity data were submitted to non-metric **IN**dividual **D**ifferences **SCAL**ing (INDSCAL) using the ALSCAL routine contained in the SPSS software [29]. The resulting two-dimensional configuration for four listeners that was obtained after 100 iterations (and at a stress value of less than 1%) is shown in Figure 2 (see [30] for a more detailed discussion of the application of such INDSCAL analyses in revealing dimensionality of spatial sound perception). The relative distances between the symbols plotted in this figure give an indication of perceptual distances between the ten stimuli presented. The position of each symbol in the plot indicates the relative magnitude of the stimulus on each of the two most salient perceptual dimensions underling the reported spatial image dissimilarities. The square symbols correspond to the reproduction mode employing two subwoofer signals, and the circular symbols correspond to the reproduction mode employing only one subwoofer signal. A line segment connects together all symbols plotting results for a given reproduction mode. The symbols were connected in an order determined by their position in the sequence of increasing stimulus CC value. The numbers labeling each symbol identify the negative CC value for each stimulus.

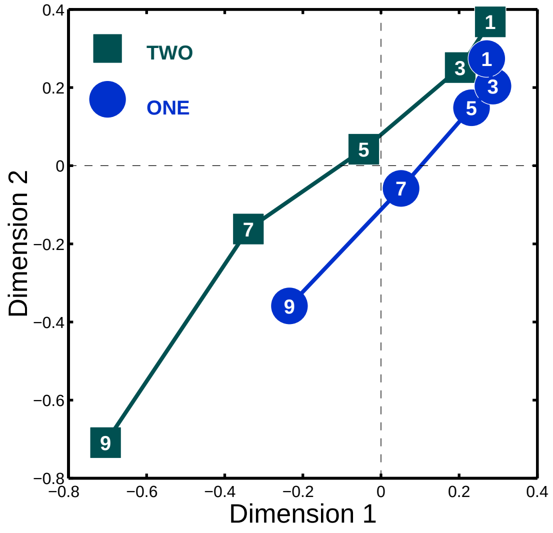

Figure 2: INDSCAL stimulus space results for four listeners showing the impact of negative cross correlation on auditory spatial imagery for two reproduction modes: one subwoofer signal (circular symbols) vs. two subwoofer signals (square symbols). These results are based upon all pairwise comparisons of the ten stimuli, rather than on dissimilarity judgments that had been constrained to pairs of stimuli sharing the same negative CC value. (Note that the negative sign and decimal point were omitted from the label on each symbol for graphical simplicity).

| Subject | Dim. 1 Weight | Dim. 2 Weight |
|---|---|---|
| 1 | 0.81 | 0.09 |
| 2 | 0.86 | 0.10 |
| 3 | 0.79 | 0.13 |
| 4 | 0.84 | 0.07 |

Table 1: INDSCAL subject space results for four listeners showing the weights placed upon each of the perceptual dimensions resulting from the analysis of inter-stimulus on dissimilarity judgments.

True to the exploratory nature of INDSCAL, the axes have been labeled "Dimension 1" and "Dimension 2." The reproduction mode has clearly made a difference in the coordinates found for each stimulus by the INDSCAL algorithm, but the interpretation of the configuration requires further study for the identification of the underlying dimensions. Perhaps the most striking feature of the spatial configuration shown in Figure 2 is that the two-subwoofer stimuli (square symbols) extend further on both Dimension 1 and Dimension 2 of the result than do the one-subwoofer stimuli (circular symbols). Note, however, that the five stimuli at the highest coordinate values on both dimensions are clustered together rather tightly, indicating that these stimuli were not so perceptually different from one another. It might also be true that the differences between these stimuli, while perhaps detectable, might be difficult to identify. This prediction is examined below via forced-choice psychophysical tests focussed upon particular perceptual attributes.

### 3.2. Attribute Identification Judgments

While INDSCAL is an analytical technique best described as exploratory, blind forced-choice identification and discrimination tasks are best described as confirmatory. That is, if there is an interest in accepting or rejecting a given hypothesis, forced-choice psychophysical methods provide one means of objective confirmation. The perceptual attributes of interest were those revealed as the two most salient dimensions in current and previous exploratory studies [1]: perceived auditory source width (ASW) and perceived auditory source distance (ASD). Two types of stimulus comparison schemes were executed for both of these attribute identifications: all pairwise comparisons and comparisons constrained to those pairs of stimuli matched in their $CC$ value.

#### *3.2.1. All Pairwise Comparisons*

Whereas the exploratory INDSCAL analysis required dissimilarity judgments for all pairwise comparisons of the ten experimental stimuli, the confirmatory analysis required attribute identification judgments for all pairwise comparisons of those experimental stimuli. In a given experimental session of 90 trials, responses were required with respect to only one of the two perceptual attributes. 90 pairs of stimuli were presented at all combinations of two reproduction modes, five negative $CC$ values (from $-.9$ to $-.1$), and two presentation orders. These 90 stimuli were presented in a single experimental session, and listeners completed two such sessions, one for each of the two attributes, ASW and ASD. The upper and lower panels of Figure 3 show the resulting "Width Dominance" and "Distance Dominance" functions, respectively.[4] In both of these graphs, symbols plotted above the dotted line that marks 0 on the y-axis correspond to stimuli which generally dominate others in the stimulus set with regard to the attribute in question, labeled "Width" or "Distance." So the stimulus with greatest ASW according to this analysis was the two-subwoofer-signal presentation at $CC = -.5$. The stimulus with lowest ASD was the two-subwoofer-signal presentation at $CC = -.9$.

#### *3.2.2. Constrained Pairwise Comparisons*

In a given experimental session of 90 trials, responses were required with respect to only one of the two perceptual attributes. Pairs of stimuli were presented at five negative $CC$ values ranging from $-.9$ to $-.1$. These five stimuli were presented 18 times in each experimental session, and in both reproduction modes in each trial. Listeners completed two such sessions for each of the two attribute identification tasks: judgments of ASW and ASD.

In each trial, two stimuli of identical cross correlation were presented first in one reproduction mode and then the other. Though the presentation order was randomly selected, the selection process was constrained so that the one-subwoofer stimulus would be presented first on exactly half of the trials. On a width-identification trial, the listener was required to choose whether the first or the second stimulus presented had a wider spatial image. On a distance-identification trial, the listener made the same forced choice of which interval contained the stimulus that was the farther of the

---

[4] A metric analysis of obtained dominance proportions has been relatively common ever since it was first introduced in 1927 by Thurstone [31]; for the current dataset, however, it is difficult to justify the assumptions of the underlying metric model (in which dominance data are related to *signed* differences between stimulus positions on a 1D scale).

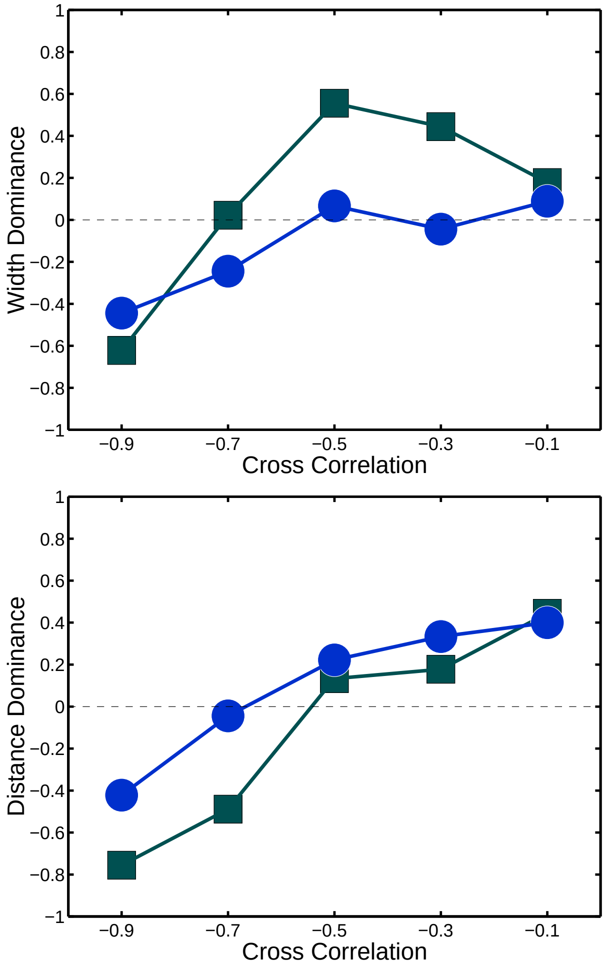

Figure 3: Two-alternative forced-choice (2AFC) identification of the temporal interval containing the stimulus of greater perceived auditory source width (upper panel), and 2AFC identification of the temporal interval containing the stimulus of greater perceived auditory source distance. The ordinate in both plots shows choice dominance as a function of stimulus $CC$ values. Dominance is defined as the proportion of trials within which a stimulus was judged to have greater ASW or ASD in pairwise comparisons with all nine of the other stimuli; hence the y-axis labels of “Width Dominance” and hence the y-axis label of “Distance Dominance.” Higher values on the y-axis indicate percepts that are more clearly wider or more clearly distant than others within the stimulus set. As in Figure 2, results for two-subwoofer-signal stimuli are plotted using square symbols, while those for one-subwoofer-signal stimuli are plotted using circular symbols. Note that, because the plotted dominance proportions are based upon a summation of choices for all pairwise comparisons, the values plotted for the two reproduction modes at each CC value will not necessarily sum to zero; rather, the dominance proportions will sum to zero across all stimuli in each plot.

two presented. The stimulus conditions under which these subjective differences were likely to be observed are summarized in Figure 4.

The upper panel of Figure 4 shows the relative dominance of perceived width associated with the two-subwoofer signal relative to that associated with the one-subwoofer signal. Therefore, high values indicate that the two-subwoofer signal created greater ASW. A width dominance proportion greater than .64 (upper dotted line) is regarded as significantly greater than chance variation would allow ($p < .05$). Width dominance less than .36 (lower dotted line) is regarded as showing that the width associated with the one-subwoofer signal significantly dominated that associated with the two-subwoofer signal (again, $p < .05$).

Sound signals presented with negative ($CC$) values via two subwoofers also produce images that extend further from the speakers into the personal space of the listener. The lower panel of Figure 4 shows that when two-subwoofer signals are included in the reproduction, one-subwoofer signals produce images that do not reach locations as close to the listener. Only at the lowest $CC$ magnitude tested (the nearly complete decorrelation of $CC = -.1$) did the identification of the dominant image width become difficult. In contrast, two-subwoofer signals presented at relatively small negative $CC$ values ($-.1$ to $-.5$) produce noticeably wider images than those produced using one-subwoofer signals. Only at a moderately large negative $CC$ value ($CC = -.7$) was the identification of the dominant image at chance level. At the largest negative $CC$ value tested ($CC = -.9$) the image was heard to be so close to the listener's head that the image width was judged to be extremely narrow, in fact narrower than images produced by one-subwoofer signals.

Because both ASW and ASD are modulated by the stimulus $CC$ value, it is not readily apparent from the INDSCAL solution which dimension of that solution corresponds to which perceptual attribute, width or distance. The two-signal presentation at a $CC$ value of $-.9$ created a spatial image that was both least distant and least wide. The spatial image associated with the two-signal presentation at a $CC$ value of $-.1$ lies at the opposite extreme in the spatial configuration, that being the most wide and most distant. The strong correlation between these two perceptual attributes creates a problem for straightforward interpretation of the derived perceptual space that can be addressed only through this subsequently executed forced-choice test.

Which dimension in the INDSCAL solution corresponds to which perceptual distinction can be determined by an examination of pairwise identification performance for the same set of stimuli. Pairs of stimuli which are located closer to each other on one dimension than the other should be more difficult to identify in terms of the perceptual attribute associated with that dimension. For example, if a pair of stimuli showed a similar resulting coordinate on the dimension corresponding to image width, but were more distant on the dimension corresponding to image distance, then those stimuli should be more identifiable in terms of image distance than image width. This is precisely what occurred at the second largest negative $CC$ values tested in the stimulus set ($-.7$).

Listeners readily distinguished changes in distance for images resulting from the two-signal versus the one-signal presentation at a $CC$ value of $-.7$ (the two-signal image was judged to be closer than the one-signal image on nearly all of the identification trials for all listeners). On the other hand, listeners did not reliably distinguish between the image widths associated with these two stimuli. This finding, taken together with the width and distance

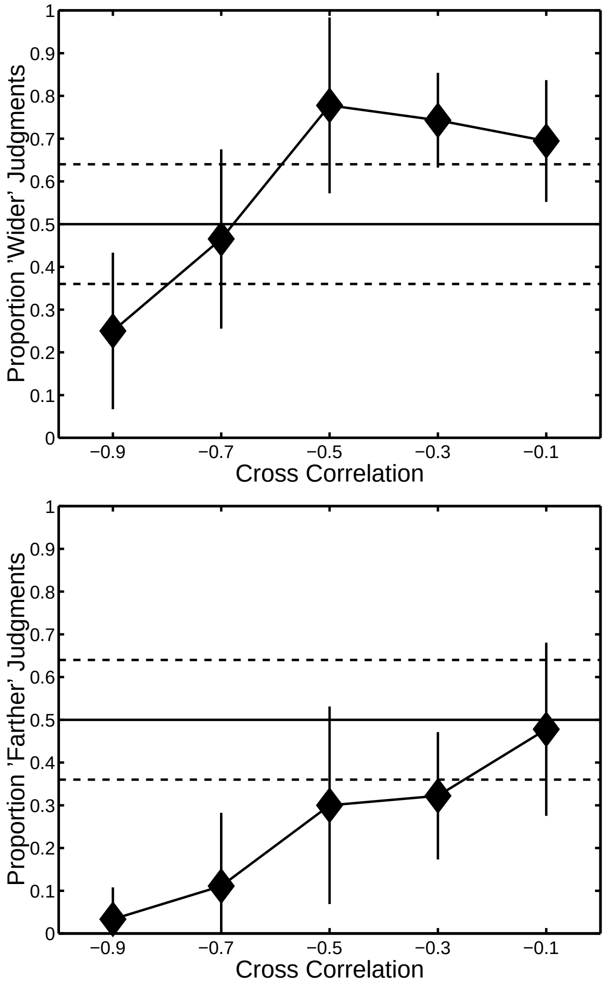

Figure 4: Proportion of 'Wider' judgments (upper panel) and proportion of 'Farther' judgments (lower panel) given for a two-subwoofer signal when presented in comparison with a one-subwoofer signal of the same $CC$ value (i.e., identification of whether the percept associated with the two-subwoofer signal dominates that associated with the one-subwoofer signal). The upper dotted line marks the criterion identification performance chosen as the indication that two-subwoofer signal was judged significantly wider than chance variation would likely allow. Width identification performance less than $.36$ (lower dotted line) is regarded as showing that the width associated with the one-subwoofer signal significantly dominated that associated with the two-subwoofer signal (again, $p < .05$). As in the upper panel of the figure, the dotted lines mark the criterion proportions enclosing the region of chance performance.

identification data for all pairwise comparisons of the ten stimuli, leads to the conclusion that Dimension 1 corresponds to image distance and Dimension 2 corresponds to image width.

### 3.3. Absolute Discrimination

In contrast to all other psychophysical tests reported here, listener responses in the absolute discrimination test could be scored as correct or incorrect. An ABX paradigm was used to test how well the one-subwoofer-signal stimulus could be discriminated from the two-subwoofer-signal stimulus on whatever basis enabled that discrimination. On each trial, three stimuli were presented. The first two stimuli, A and B, were always a pair of stimuli with matching $CC$ value, but differing in reproduction mode (one signal vs. two), but presented in a random order. The third stimulus, X, was always one of the first two stimuli presented yet again to the listener, whose task was to tell whether stimulus X was the same as stimulus A or stimulus B. Therefore, there was no required identification of the way in which the two stimuli differed.

Absolute discrimination performance was practically perfect (greater than 98% correct) at all $CC$ values more extreme than $CC = -.3$. The lowest obtained percentage of correct discriminations comparing over all listeners was around 81% at $CC = -.3$. For this one listener, such relatively poor performance is still high enough to warrant the conclusion that absolute discrimination was relatively easy, especially in comparison to the difficulty of identifying width and distance dominance between the two reproduction modes at the low $CC$ values (i.e., $-.1$ and $-.3$) [5].

## 4. DISCUSSION

The results of all listening experiments reported here indicated that extended control over auditory spatial imagery was provided by the use of two subwoofers to deliver decorrelated low-frequency signals. This result is summarized in Figure 5, in which the INDSCAL results have been re-plotted in order to make a better visualization of their interpretation. In that figure, the image width is represented by the width of a line segment that is plotted at the apparent distance of the image from the listener. In the interpretation of an obtained INDSCAL stimulus space, the integrity of the solution is violated only by a rotation of the axes; changes associated with translation, reflection, and permutation have no substantive importance with regard to the uniqueness of the spatial configuration (a comprehensive treatment of this issue may be found in [32]).

This plot may also be regarded as a view from above the listening space, in which the listener might be located at the origin $(0, 0)$, and the speakers might be located at the points $(\pm 1, 0)$. From this perspective it is easy to see that the two-subwoofer reproduction of a signals exhibiting $CC = -.9$ generates a spatial image that is the nearest and narrowest of all the spatial images associated with the 10 stimuli presented. At the smaller negative $CC$ values, however, ASD (associated with dimension 1 of the INDSCAL stimulus space) is not so different. These images only differ in ASD (associated with INDSCAL dimension 2).

---

[5] Discrimination at 81% corresponds to 29 correct 'A' responses out of 36 trials. If a listener truly cannot distinguish between stimulus A and stimulus B, and assuming that there is a 50% chance of guessing 'A' when stimulus X is 'A', then the probability of getting 29 correct 'A' responses out of 36 trials by chance alone is around one in 10,000.

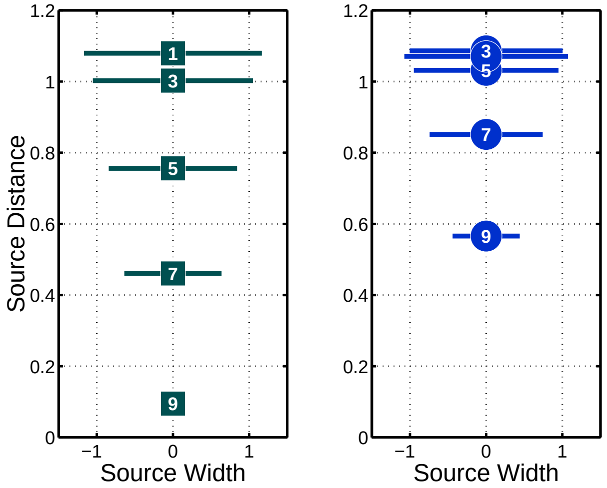

Figure 5: Interpretive representation of the INDSCAL results for two reproduction modes: one-subwoofer signal (circular symbols; right panel) vs. two-subwoofer signals (square symbols; left panel). All coordinates on Dimension 1 of the INDSCAL stimulus space configuration have been translated by a constant value in order to facilitate interpretation.

### 4.1. Implications for Spatial Auditory Display

A good starting point for discussing the implications of these results for spatial auditory display is to look at what will be missing when only a single low-frequency channel is included in stereophonic sound reproduction. One type of spatial sound effect that is largely missing from most two-channel loudspeaker reproduction is the sense intimacy that is associated with small spaces and sources located very near the listener (headphone reproduction of such auditory spatial imagery is another matter, addressed in recent papers by the author [33] [34]). One way to create this sense of intimacy using two-channel reproduction is to present low-frequency signals exhibiting a strongly negative cross-correlation (in the range from $-0.7$ to $-0.9$). The immersive spatial imagery that results from such processing is distinctly different from what can be created using five positively-correlated signals arriving from positions surrounding the listener. It remains to be discovered whether the sense of intimacy that results from this two-channel reproduction is preferred to that associated with the use of the split surround channels provided by the 5.1-channel "Dolby Digital" coding scheme. Certainly the "sweet spot" will be larger using the discrete-speaker multichannel system, but this issue is orthogonal to the issue of the well-situated listener's sense of immersion in the reproduced sound field. Of course, the selective use of negative decorrelation as a production technique allows for greater flexibility when program material is to be distributed using the discrete coding of 5.1-channel Dolby Digital. It is also possible that some increased flexibility in the coding scheme itself may allow for multichannel transmissions that include a second subwoofer channel, such as the 10.2-channel system proposed by Holman [35].

The historic "Dolby Pro Logic" scheme precluded the delivery of low-frequency signals exhibiting negative cross-correlation, since partial cancellation of the subwoofer signal would result. At frequencies above those channeled to the subwoofer, Dolby Pro Logic also precluded the discrete left- and right-channel delivery of mid-frequency signals exhibiting negative cross-correlation, since the difference signal associated with such negatively correlated signals would end up in the single-channel surround. Of course, this single-channel surround signal can be submitted to additional synthetic stereo processing (using more DSP than that required for Pro Logic), and delivered along with the Pro Logic left- and right-channel signals to a conventional two-channel reproduction system. This procedure creates a result akin to what Dolby has termed "Virtual Dolby Surround," and many companies are offering similar virtual surround technology[6]. Suffice it to say that increasing computational efficiency of this DSP application while maintaining perceptual quality is of continuing interest (see [36] for recent results in related signal processing).

Before further discussion of such two-channel virtual surround systems, there is an important caveat to be made. Though some full-range loudspeakers produce a significant response at frequencies as low as 50 Hz, such reproduction systems are rare. Especially if the target application for virtual surround is PC-based multimedia and television, we might expect to see relatively low-quality loudspeakers. Currently, the sound system upgrade for PC-based multimedia systems almost always includes a subwoofer. The consequence is that for decorrelated low-frequency signals cannot be reproduced by these systems. The following discussion of two-channel virtual surround and synthetic decorrelation effects should be interesting despite this unfortunate situation.

The results of decorrelation-based synthetic stereo processing to create a virtual surround from a single-channel signal often sound unrealistic. The resulting spatial image might sound wide and have great depth, but may sound somewhat unnatural. If we take as a standard of comparison the sense of surround that comes as part of the realistic spatial imagery available from well-calibrated, cross-talk cancellation systems delivering binaural recordings, then the results of synthetic decorrelation will fall short of the mark. But for most applications this is not a practical standard of comparison.

Outside of the laboratory, spatial sound effects must be robust enough to withstand variation in head size and listening position, and hence cannot stand up to the binaural standard described above. Rather, it should be able to meet the standard of comparison associated with the best program material delivered over the best discrete-speaker multichannel system. This means that the goal of realism is not as critical as the goal of well-controlled expressive and powerful spatial imagery. Surreal imagery (greater than reality, or matching some imagined other reality) is perhaps more important here (cf. [37]).

Effects processing systems that produce unrealistic but exceptionally good-sounding results should be supported by virtual surround systems in a manner that mimics the performance of the more popular reproduction target, the discrete-speaker multichannel system. By definition, unrealistic imagery does not require realistic head-related transfer functions (HRTFs) to succeed in producing good results. The extended variation in spatial imagery provided by controlled two-channel decorrelation is an example. Though reproduction systems providing low IACC produce spatial imagery that is almost always preferred to that of systems providing only high IACC [5], a virtual surround system supporting this feature should also be preferred. Note again that there is support from the current results only for the conclusion regarding extended

---

[6] AC-3, Dolby Digital, Dolby Pro Logic, and Dolby Virtual Surround are trademarks of Dolby Laboratories. For more information see their website regarding professional products: `www.dolby.com/pro`

variation of spatial imagery, and not for the conclusion that such imagery will necessarily be prefered. As preferences were not investigated in this study, no further speculation about preferences will be made here. Rather, it need only be pointed out that interesting spatial sound effects are available that are not necessarily natural sounding as high-quality binaural recording and reproduction, but might nonetheless be preferred. Indeed, it has been found that when several spatial sound capture systems are used to simultaneously transmit a common set of acoustic events, the classic Blumlein stereo system [38] is frequently preferred over high-quality binaural transmission [39]. Suffice it to say that the most natural sounding imagery is not necessarily most preferred.

The AC-3 multichannel audio format allows discrete storage and transmission of surround channel information that has the producer's chosen amount of decorrelation, and this artistic choice should be reproduced if at all possible. Although simulating a natural phenomenon that can be observed in the acoustical response of actual spaces, the spatial imagery associated with synthetic decorrelation is more easily controlled parametrically, and extends over a range of perceptual responses that is hard to cover using binaural recording or other technology designed to capture naturalistic spatial soundfields [40]. This observation should be contrasted, however, with observations about actual binaural soundfields in concert halls, and the perceptual relevance of the spatial distribution of reflections (e.g., see [41] [42]), which is not addressed here.

### 4.2. Implications for common-practice production

Prior to conducting the controlled studies reported herein, it was noted that the common-practice techniques used in audio production for entertainment (such as music and sound effects for film) produced spatial imagery that sounded more powerful when two subwoofers were used. That is to say, in comparison to the spatial imagery that results when a single subwoofer is used, applying the left- and right-channel signals to the respective inputs of two spatially-separated subwoofers results in imagery that is distinctly different, and preferred by many listeners. This is despite the fact that the targeted reproduction system has only a single subwoofer. Why should the result be so powerful when production techniques target a different reproduction system? The reason is that common-practice techniques can produce decorrelation at low-frequency, and that this decorrelation is "lost in the mix" before it can reach the average listener.

One such common-practice audio-production technique is to record a vocalist (or instrumentalist) twice, and to use the two similar (but decorrelated) recordings for the left- and right-channel signals in the final mix. Though the pitch and amplitude contours are very similar, the inter-channel phase variation at the fundamental frequency will be great enough to produce a widening of the spatial image that is greatest when two subwoofers are used (assuming that the fundamental frequency is within this low-frequency range). Another common practice is the use of two-channel delay effects offered by many consumer and professional effects processors such as reverberation units. Such two-channel delay effects may be the most popular tools for audio production in current and historical practice. The simplest version of this technique is to deliver a dry sound via one channel (targeting the left loudspeaker, for instance), and to deliver a delayed copy of that sound via the other channel (targeting the right loudspeaker). As Kendall [43] pointed out, this produces little decorrelation (unless the delay is extremely long), and has tone coloration as the dominant perceptual result. Another common variant, however, is much more effective in producing decorrelation. This popular technique delivers a dry sound to one loudspeaker and a delayed, wet (reverberant) version of that sound to the other loudspeaker. Even if the delay is not large in this case, the reverberation processing can produce phase differences at low frequencies and the resulting spatial imagery will benefit from the use of two subwoofers.

A skeptic might point out that the phase shift at extreme low frequencies will be very small unless the inter-channel delay is very large. While this is true, it does not lead to the conclusion that the resulting subjective effects are insignificant. Though the subjective experience associated with extreme negative CC values is somewhat strange (sometimes termed "phasey" but perhaps more aptly described as "penetrating" [39]), such imagery can also be experienced in everyday listening if a single strong reflection arrives at the proper delay relative to the direct sound. A simple sound demonstration of Barron's [44] classic results varying the delay and level of single lateral reflections (described in the author's previous report on this topic [1]), will confirm that the subjective effects of including two subwoofers in the reproduction of two-channel delay effects are substantial.

## 5. CONCLUSIONS

This paper asked the question, "Can reproduction of two-subwoofer signals enable increased variation in auditory spatial imagery?" The results of exploratory experiments showed that presenting two low-frequency signals rather than one had an impact upon spatial image quality, and revealed the two most salient perceptual dimensions underlying the subjective differences between the two reproduction modes. In order to determine whether differences along these two dimensions could be interpreted as differences in the apparent width and apparent distance of the auditory spatial imagery, a set of confirmatory experiments were run. The results showed that differences between the presented auditory spatial images was well described in terms of these two perceptual dimensions, and furthermore, that the two reproduction modes were associated with predictable differences in spatial imagery that could be consistently identified as differences in the perceptual attributes of width and distance.

An alternative statement of the goal of this paper was, "Does the use of only one subwoofer signal in stereophonic sound reproduction truly provide an adequate range of auditory spatial images?" The results of this research show quite clearly that the two most salient perceptual attributes of auditory spatial imagery that can be manipulated via cross correlation are extended when two-subwoofer signals are presented. It is concluded that the variation in auditory spatial imagery provided by the use of two subwoofers is reduced when negative correlations are not reproduced at low-frequency. Of course, the generality of this conclusion is limited by a number of factors, most notably, that stimuli were presented under anechoic reproduction conditions. A follow-up study in a standard listening room might not show such clear differences. Furthermore, the stimuli themselves were very special, consisting of gated noise with autocorrelation functions not typical of musical instruments or speech, and so again it is difficult to generalize the results. Also, it should be reiterated that no investigation of subjective preference was done in the current study. Nonetheless, with these caveats in mind, a clear summary of this study and its results can be proffered.

Two-channel sound presentations were made under controlled conditions in which stimuli differed only with respect to the presence or absence of decorrelation in a low-frequency band. Subjective features of the reproduced spatial image were manipulated via the powerful psychoacoustic parameter, inter-aural cross correlation (IACC). Three listening experiments were executed to test the hypothesis that extended control over such spatial imagery is provided by the use of two subwoofers. Some of the judgments made by listeners in this study focussed their attention upon particular features of the presented spatial imagery, such as auditory source width and auditory source distance. Other judgments required them to report on the global perceptual difference between two spatial images, without respect to particular details. Results for all of these experiments, taken together, lead to the conclusion that extended control over spatial imagery is provided by the use of two subwoofers, and that results using a single subwoofer will be particularly degraded for source material containing signals exhibiting negative cross correlation values.

## 6. ACKNOWLEDGMENTS

The author would like to thank Francis Rumsey for his review of an early draft of this paper, and also Michael Cohen and Jens Herder for their comments on earlier drafts, and for their input throughout the course of this project. The experimental listening subjects who voluntarily participated also gave valuable feedback. This research was supported in part by the Fukushima Prefectural Foundation for the Advancement of Science and Education.

## 7. REFERENCES

[1] W. L. Martens, "The impact of decorrelated low-frequency reproduction on auditory spatial imagery: Are two subwoofers better than one?," in *Proceedings of the Audio Engineering Society 16th Int. Conf. on Spatial Sound Reproduction*, Rovaniemi, Finland, 1999, pp. 67–77.

[2] ITU-R, *Recommendation BS.775-1, Multichannel stereophonic sound systems with and without accompanying picture*, Int. Telecommunications Union Radiocommunication Assembly, 1994.

[3] T. Letowski, "Sound quality assessment: concepts and criteria," in *Proc. Audio Engineering Society 87th Int. Conv.*, 1989, Preprint 2825.

[4] N. Zacharov, S. Bech, and D. Meares, "The use of subwoofers in the context of multichannel surround sound reproduction," **J. A***udio* **E***ng.* **S***oc.*, vol. 46, no. 4, pp. 276–287, 1998.

[5] M. Tohyama, H. Suzuki, and Y. Ando, *The Nature and Technology of Acoustic Space*, Academic Press, London, 1995, ISBN 0-12-692590-9.

[6] K. Kurozumi and K. Ohgushi, "The relationship between the cross-correlation coeffiecient of two-channel acoustical signals and sound image quality," **J. A***cous.* **S***oc.* **A***mer.*, vol. 74, pp. 1726–1733, 1983.

[7] T. Welti, "How many subwoofers are enough," in *Proc. Audio Engineering Society 112th Int. Conv.*, 2002, Preprint 5602.

[8] D. Griesinger, "Objective measures of spaciousness and envelopment," in *Proc. Audio Engineering Society 16th Int. Conf. on Spatial Sound Reproduction*, Rovaniemi, Finland, 1999, pp. 27–41.

[9] M. Morimoto, K. Ueda, and M. Kiyama, "Effects of frequency characteristics of degree of interaural cross-correlation and sound pressure level on auditory source width," *Acustica*, vol. 81, pp. 20–25, 1995.

[10] M. Morimoto and Z. Maekawa, "Auditory spaciousness and envelopment," in *Proc. 13th Int. Congress on Acoustics*, Belgrade, Yugoslavia, 1989, pp. 215–218.

[11] M. Morimoto, "How can auditory spatial impression be generated and controlled?," in *Proc.* **I***nternational* **W***orkshop on* **S***patial* **M***edia*, Aizu-Wakamatsu, Japan, Oct. 2001, pp. 1–15.

[12] F. Rumsey, *Spatial Audio*, Focal Press, Oxford, 2001.

[13] M. Morimoto and M. Iida, "Appropriate bandwidth and integration time for measurements of the degree of interaural cross correlation as a measure of apparent source width in concert halls," *ACUSTICA - acta acustica*, vol. 85, pp. S120, January/February 1999, Supplement 1.

[14] G. Plenge, "On the problem of inside-the-head locatedness," *Akustica*, vol. 26, pp. 421–424, 1972.

[15] J. Blauert and W. Lindemann, "Spatial mapping of intracranial auditory events for various degrees of interaural coherence," **J. A***cous.* **S***oc.* **A***mer.*, vol. 79, pp. 806–813, 1986.

[16] R. Mason, N. Ford, F. Rumsey, and B. de Bruyn, "Verbal and non-verbal elicitation techniques in subjective assessment of the spatial attributes of reproduced sound," **J. A***udio* **E***ng.* **S***oc.*, vol. 49, no. 5, pp. 366–384, 2001.

[17] J. Blauert, *Spatial Hearing: The Psychophysics of Human Sound Localization, (Revised Edition)*, MIT Press, Cambridge, Massachusetts, 1997.

[18] W. Reichardt and W. Schmidt, "Die hörbaren Stufen des Raumeindrucks bei Musik [The audible steps of spatial impression in music]," *Acustica*, vol. 17, pp. 175–179, 1966, (in German).

[19] W. Reichardt and U. Lehmann, "Raumeindruck als Oberbegriff von Räumlichkeit und Halligkeit. Erläuterung der Raumeindrucksmasses $r$ [Spatial impression as a broad concept including spaciousness and reverberance. Defining the index of spatial impression $r$]," *Acustica*, vol. 40, pp. 277–289, 1978, (in German).

[20] F. E. Toole, "Subjective measurements of loudspeaker sound quality and listener performance," **J. A***udio* **E***ng.* **S***oc.*, vol. 33, 1985.

[21] S. Bech, "Selection and training of subjects for listening tests on sound-reproducing equipment," **J. A***udio* **E***ng.* **S***oc.*, vol. 40, no. 7/8, pp. 590–610, 1992.

[22] S. S. Shiffman, M. L. Reynolds, and F. W. Young, *Introduction of multidimensional scaling*, Academic press, 1981.

[23] J. D. Carroll and J. J. Chang, "Analysis of individual differences in multidimensional scaling via an n-way generalization of "Eckart-Young" decomposition," *Psychometrika*, vol. 35, pp. 283–319, 1970.

[24] Y. L. Lee, *Statistical theory of communication*, John Wiley, New York, 1995.

[25] D. H. Cooper and J. L. Bauck, "Prospects for transaural recording," **J**. **A***udio* **E***ng.* **S***oc.*, vol. 37, pp. 29–40, 1989.

[26] J. A. Pedersen and A. Mäkivirta, "Quantification of subwoofer requirements, part I: Generation of stimuli and listening system equalization," in *Proc. Audio Engineering Society 108*th *Int. Conf.*, Paris, Feb. 2000.

[27] M. Tohyama and H. Suzuki, "Interaural cross-correlation coefficients in stereo-reproduced sound fields," **J**. **A***cous.* **S***oc.* **A***mer.*, vol. 85, pp. 780–786, 1989.

[28] R. O. Duda and W. L. Martens, "Range dependence of the response of an ideal rigid sphere to a point sound source," **J**. **A***cous.* **S***oc.* **A***mer.*, vol. 105, no. 5, pp. 3048–3058, 1998.

[29] SPSS Inc., Chicago, WA, *SPSS Base 10.0 User's Guide*, 1999.

[30] W. L. Martens and N. Zacharov, "Multidimensional perceptual unfolding of spatially processed speech I: Deriving stimulus space using INDSCAL," in *Proc. 109*th *Conv. of the Audio Engineering Society*, Los Angeles, Sept. 2000, Preprint 5224.

[31] L. L. Thurstone, "A law of comparative judgment," *Psychological review*, vol. 35, pp. 273–289, 1927.

[32] P. Arabie, J. D. Carroll, and W. S. DeSarbo, *Three-Way Scaling and Clustering*, Sage University Paper Series on Quantitative Applications in the Social Sciences, series no. 07-065. Sage Publications, Newbury Park, CA, USA, 1987.

[33] W. L. Martens, "Efficient auralization of small, cluttered spaces: Simulating sonic obstructions at close range," in *Proc. 7*th *Western Pacific Regional Acoustics Conf.*, S. Kuwano and T. Kato, Eds., Kumamoto, Japan, Oct. 2000, **A**cous. **S**oc. **J**apan, pp. 317–320.

[34] W. L. Martens, "Psychophysical calibration for controlling the range of a virtual sound source: Multidimensional complexity in spatial auditory display," in *Proc. Int. Conf. on Auditory Display*, Espoo, Finland, 2001, ICAD, pp. 197–207.

[35] T. Holman, *Sound for Film and Television*, Focal Press, Boston, 1997.

[36] T. Hayashi and W. L. Martens, "The synthesis of low-peak cross-correlation sequences using trigonometric function aliasing," *IEICE Transactions Fundamentals*, vol. E82-A (8), pp. 1402–1411, 1999.

[37] G. Theile, "On the naturalness of two-channel stereo sound," **J**. **A***udio* **E***ng.* **S***oc.*, vol. 39, no. 10, pp. 761–767, Dec. 1991.

[38] A. D. Blumlein, "Improvements in and relating to sound-transmission, sound-recording, and sound-reproducing systems," U.K. Patent 394,325, 1931.

[39] N. Zacharov, "Understanding spatial sound reproduction: Perception and preference," in *Proc.* **I***nternational* **W***orkshop on* **S***patial* **M***edia*, Aizu-Wakamatsu, Japan, Oct. 2001, pp. 43–61.

[40] M. Morimoto, K. Iida, and K. Sakagami, "The role of reflections from behind the listener in spatial impression," *Applied Acoustics*, vol. 62, no. 2, pp. 109–124, 2001.

[41] M. Barron, "Late lateral energy fractions and the envelopment question in concert halls," *Applied Acoustics*, vol. 62, no. 2, pp. 185–202, 2001.

[42] H. Furuya, K. Fujimoto, C. Y. Ji, and N. Higa, "Arrival direction of late sound and listener envelopment," *Applied Acoustics*, vol. 62, no. 2, pp. 125–136, 2001.

[43] G. S. Kendall, "The decorrelation of audio signals and its impact on spatial imagery," *Computer Music Journal*, vol. 19, 1995.

[44] M. F. E. Barron, "The subjective effects of first reflections in concert halls - the need for lateral reflections," *Journal of Sound and Vibration*, vol. 15, pp. 475–94, 1971.