Audio Engineering Society

Convention Paper 6370

Presented at the 118th Convention
2005 May 28–31 Barcelona, Spain

Interaction of source and reverberance spatial imagery in multichannel loudspeaker audio

John Usher[1], Wieslaw Woszczyk[2]

[1] *Multichannel Audio Research Laboratory, McGill University, Montréal, Canada*

[2] *Centre for Interdisciplinary Research in Music Media and Technology, McGill University*

Correspondence should be addressed to John Usher (`john.usher@mail.mcgill.ca`)

ABSTRACT

Sound imagery is discussed in many contexts in subjective loudspeaker audio evaluation. In this paper we investigate imagery in terms of the spatial properties of an auditory object. A general categorization of auditory images into either Source or Reverberance images is well established in the literature (also called ASW and LEV); here we discuss perceptual organization principles and physical factors which affect this distinction. The degree to which existing theories for controlling an S image direction can be applied to controlling an R image was investigated in a two-part experiment. With a loudspeaker arrangement according to ITU-R BS.775-1 (no centre speaker) and a graphical mapping system developed in previous work we investigated the spatial imagery associated with source and reverberance images around the listener. Stimuli used were a mono channel of recorded anechoic flute and a channel of artificial reverberation. The image direction was affected with either a real loudspeaker source at various locations or a phantom source created with pair-wise amplitude panning. We compare source and reverberance images in terms of perceived image width, distance and azimuth. We find that: the spatial location and width of S and R images can be independently described using a GUI; pair-wise amplitude panning of S and R images is possible using the side loudspeakers but the consistency of reported spatial image geometry is less than for frontal images; we gain an insight into spatial (un)masking effects of an S image by an R image.

1. INTRODUCTION

1.1. Source and Reverberance images in loudspeaker audio

1.1.1. Auditory objects and images

The choice of words to distinguish acoustic and perceptual sound is varied. A well-used couple are "sound source" and "auditory event" [7, pg. 102]. In the present work we call the "acoustic world" event *sound* (be it proximal or distal), and the "perceptual world" event the ***auditory object*** [20]. The relationship between sound and auditory objects is generally not a simple one-to-one correspondence. An example is given by Bregman [9, pg. 10]: "A series of footsteps, for instance, can form a single experienced event, despite the fact that each footstep is a single event." To reflect the active process which brings about the perception of the sound, Bregman uses the word "stream": "the perceptual unit that represents a single happening. ... The stream serves the purpose of clustering related qualities" [9, pg. 10]. The word *object* is often used synonymously with *stream* [29] but we find that the word "stream" is best used when the ecological relationship between acoustic and perceptual event are more abstract than a simple one-to-one mapping, such as with footsteps or musical voices.

Another term relating to auditory objects is an auditory ***image***. Letowsky [31] defines an auditory image as the representation of an auditory object in terms of timbre and space. Griffiths and Warren [20], on the other hand, use only the timbral dimensions of frequency and time. The distinction between these two representations of an auditory object reflects the theory that space in audition is not an *indispensable attribute* [28, 29]. "An attribute (or dimension) is defined as indispensable if and only if it is a prerequisite of perceptual numerosity"[29]. A simple example supporting this theory is given (again!) by Bregman [9, pg. 75]; that we can segregate (and count the number of) different voices (musical or human) even when they are all reproduced with a monophonic radio.

In the present paper we will investigate auditory objects in terms of the spatial dimension, restricted to the horizontal plane. We define an auditory image as "those properties of an auditory object which can be described with the dimensions of space". By "space" we mean the psychological construct which is equivalent to physical space. In this sense, it is tautological to say "perceived spatial imagery" instead of just "image".

In loudspeaker audio, a distinction is made between a **phantom**[1] image and a **real** sound image. A real image is perceived to exist at the same location as the acoustic source(s) whereas a phantom image is perceived to exist at a position generally (though not always [53]) between the loudspeakers creating the sound. Of course, in the phenomenological sense a phantom image is a real image too, but the word "real" is a good compliment to the word "phantom" and has been used in other studies (e.g. [7, 57]).

1.1.2. Source images

When recorded music is reproduced with two loudspeakers the listener will have a sense of location for the recorded instrument; the ***source image***. To describe how strong the sense of location of the source image is, Blauert [7] uses the term *locatedness*. If the sense of direction is very weak, the locatedness is said to be *diffuse*, if it is strong, then the locatedness is *sharp*, *concise*, *clear* or *defined* (e.g. [62, 54, 24, 66]). The term *definition* is also used to describe this sense of location, defined by Toole [62] as "the extent that different sources of sound are spatially separated and positionally defined". A source object may be spatially discontinuous, so a single auditory object may be described with more than one auditory image. For example, the perceived regions of space which contain the source object may be separated by regions of space which do not. An example of this in nature is the sound of a large but partly occluded sound-creating thing, such as the sea heard from behind a sand-dune.

1.1.3. Reverberance images

When describing or comparing a reproduced music listening experience, the most commonly used spatial adjectives are related to the reflected sound content in the recording (e.g. [35, 62, 5, 67, 21]). In acoustical terms, reflected sound can be thought of

[1] The word "virtual" is used interchangeably with "phantom".

as consisting of two parts: early reflections (ER's) and reverberation (reverb). ER's are defined as "those reflections which arrive at the ear via a predictable, non-stochastic directional path, generally within 80 ms of the direct sound" [4] whereas reverberation is generally considered to be sound reflections impinging on a point (e.g. microphone or ear drum) which can be modelled as a stochastic ergodic function [8].

The corresponding psychological term to describe the perception of reverberation is ***reverberance*** [42, 34, 45]. Morimoto and Asaoka [45] conducted a dissimilarity judgement experiment to rate simple sound stimuli in terms of the adjective "reverberance". A solo violin was reproduced from a centre loudspeaker in an anechoic chamber, and two uncorrelated reverberation channels were reproduced from the left and right with loudspeakers at various angles from the centre-speaker (10°, 75°, and 135°). The reverberation time was also varied (1.0, 1.4, and 2.0 seconds). An MDS analysis of the dissimilarity judgement ratings for the different scene configurations revealed that these two variables were represented approximately orthogonally on a two-dimensional space. The relationship between the psychological construct reverberance and the acoustical description of reverberation are clear: both are spatially distributed about the listener and have similar temporal patterns. The spatial distribution of reproduced reverberation needed for the perception of a diffuse sound field was investigated by Hiyama et al. [22], who found that the local acoustic wavefield around a listener does not have to be satisfy the acoustic definition of reverberation. In their study, it was found that regardless of frequency the perceived spatial homogeneity of reverberance using 24 loudspeakers equally spaced around the listener could be achieved using only 12, and that with only four loudspeakers (conveniently arranged according to the 2/2 ITU-R BS.775-1 format) it is possible to create a soundfield which is perceived to be nearly identical to the 24-speaker arrangement.

1.1.4. Summary

When listening to live or recorded musical instruments, according to Griesinger [19]: "The brain processes incoming sound into a foreground stream - the part which holds the information content of the signal - and a background stream... In a reverberant environment the background is the reverberation". This is related to the figure-ground analogy of what an auditory object is [29], where the figure in this case is the source or foreground stream and the ground is the perceived reverberation or background stream. In the ideal analytical listening case "Attention selects one putative object (or a small set of them) to become figure ... and relegates all other information to ground" [29]. In the present paper we shall only deal with one source and one reverberance object, so the foreground or background stream can be either the source or reverberance object.

Morimoto [46, 44] investigated temporal factors of the reflected sound which affected the perceptual distinction between S and R images. He showed that the factors were related to the precedence effect: early, high-level reflections (i.e. ER's) are fused with the direct sound to create the source stream and later low-level reflections (i.e. reverberation) are involved in the formation of the reverberance stream. Griesinger [19] compares this to a forward-masking effect, whereby the early-arriving source-related information masks the reverberation.

The formation of a reverberance stream *is influenced* by the perceived spatial distribution of reverberation but it is the *temporal* nature of this stream which distinguishes it from the source stream. For example, reproduction of reverberation from a single loudspeaker can still be recognized as reproduced reverberation and has the sound character of reverberance.

The combined effect of the source and reverberance objects in reproduced sound is generally associated with the holistic feature called (Auditory) Spatial Impression (ASI), or more simply *spaciousness* [7, pg. 348]. The bent of this paper is on imagery of perceived source and reverberance objects in a loudspeaker audio scene rather than a holistic description of the spatial listening experience, so we will avoid using multidimensional concepts such as ASI.

1.2. Control of spatial S and R imagery around a listener in loudspeaker audio using Pair-Wise Amplitude Panning (PWAP)

When two loudspeakers radiate a coherent signal, a phantom auditory image may appear from a direction and with an extent which is dependant on the location of each loudspeaker relative to the listener and the degree of coherence and relative gain between the signals fed to each loudspeaker [7]. This phenomenon is called *summing localization* [7], *intensity stereophony* [61] or (vector based) *amplitude panning* (VBAP) [48]. In this paper we will use the term Pair-Wise Amplitude Panning (PWAP).

1.2.1. Panning in front of the listener

When the loudspeaker pair are the front left and right loudspeakers in the standard ITU-R BS.775-1 arrangement (i.e. at $\pm 30°$) and the signal fed to each loudspeaker is coherent, the perceived direction of the image can be predicted for a variety of signals with an accuracy of a few degrees (e.g. [14, 10, 52]). The within-subject and between-subject consistency for reporting phantom image direction is related to the "localization blur" [7] and is expressed in terms of a standard deviation or inter-quartile range. In a study involving both real and phantom images created with PWAP, using 400ms octave-band-limited noise bursts Pulkki and Hirvonen [52] found the localization blur for a real source at 0° to be less then 2° (SD). For a phantom image panned at 15° (with channel gain coefficients according to the tangent-law [6]) the localization blur SD was $<4°$, regardless of frequency [52]. In a similar study using a variety of anechoic and noise stimuli, Choisel and Zimmer [13] found the between-subject variation in reported image azimuth to be between to 1° and 2° (SD) for both real and phantom sources located from 0° to 30°.

The width of a phantom source image (**ASW** [58], closely related to "spread" [50] or "focus" [38, 30]) when it is located between the front left and right loudspeakers is dependant on the reproduced signal [41, 65], e.g. 7° for anechoic bongos and 18° for anechoic speech . ASW of frontal phantom images also increase with loudness [26, 65].

In this paper we will not consider timbral artifacts introduced by amplitude panning, but it has been found that for PWAP with loudspeakers at $\pm 30°$, colouration is related to localization blur, being maximal for images panned at $\pm 15°$ [49] and minimal when there is a large amplitude difference (>20 dB) between coherent loudspeakers used in PWAP [37].

1.2.2. Panning to the side of the listener

When the loudspeaker pair is located to the side of the listener, such as the front-right and rear-right loudspeakers in ITU-R BS.775-1, the relationship between perceived image direction and inter-speaker signal gain is not a smooth function and the localization blur increases. Theile and Plenge [61] used a loudspeaker pair located to the left of the listener, with the front and rear speakers at 50° and 120° to the central axis- i.e. as if the front loudspeaker pair had been rotated by 80° to the left. The localization blur was largest when the inter-channel level difference was small- for a level difference of 0 to -6 dB (that is, with the front speaker softer than the rear), the inter-quartile range was approximately 50°. Different "angles of rotation" were also investigated: 40°, 60°, 80° and 90°. The pulling of the image towards the median plane away from the centre line of the loudspeaker pair was greater as the rotation angle increased (see table 1 for a summary). Similarly, Ratliff [55] found that for a quadraphonic "square" array (loudspeakers at $\pm 90°$ and $\pm 270°$), when panning between the front and rear loudspeakers with an interchannel level difference of 0 dB the image direction was reported at approximately ± 61 ° and described as "very diffuse" and "very jumpy".

Using the ITU-R BS.775-1 loudspeaker arrangement (rear loudspeakers at $\pm 120°$) Martin et al. [39] and Corey [16] found a similar forward-pulling effect as that found by Theile and Plenge [61]. Corey [16] investigated pair-wise panning of an anechoic source at a variety of intended directions between the front and rear loudspeakers 45°, 65°, 80° and 100°. The 45° and 65° sources were reported about 5° and 10° towards the median plane, whilst the 80° and 100° sources were pulled towards the rear loudspeakers by a similar amount [16]- the addition of simulated early reflections did not significantly affect the mean localization judgement.

Pulkki and Hirvonen [52] found that when the intended image direction was 75° (again, according to the tangent panning law), the mean reported image direction ranged from 55° (200 Hz) to 62° (1.6 kHz),

Rotation δ	Image direction φ	Detent (δ - φ)	Blur
0	0	0	0
40	38	2	8
60	55	5	15
80	68	12	40
90	77	13	45

Table 1: Data from the Theile and Plenge [61] study showing how both the image is pulled towards the median plane as a speaker-pair is rotated to the side of the listener, and how localization blur increases (blur is the between-subjects inter-quartile range, in degrees). The inter-speaker angle was 60° for all five loudspeaker configurations, so the configurations can be considered as a rotation of the front Left and Right loudspeakers in the ITU-R BS.775-1 arrangement. For the data shown, each loudspeaker signal had the same gain. δ is the rotation of the LR pair and φ the perceived image direction (median). In the present paper, we use the same azimuthal zero-degree reference as that point half way between the front left and right loudspeakers.

with a blur of 11° (SD). Also, the direction of side images in ITU-R BS.775-1 systems are reported with less certainty or confidence, with the lowest certainty for images reported in the direction of 80° [15] and 90° [32].

All of the discussed work used either noise [61, 52], or anechoic recordings of solo instruments [39, 16] or speech [61]. In the present paper we will investigate if reverberance images (both phantom and real) can be created to the side of a listener for a ITU-R BS.775-1 set-up, and how the reported direction of these reverberance images compares with that of phantom and real source images.

1.3. Hypotheses for interaction of Source and Reverberance images around a listener

The hypotheses apply to sound reproduced with loudspeakers arranged according to ITU-R BS.775-1. General hypotheses are stated here and later we describe a veridical method for investigating a particular hypothesis.

1.3.1. Source and Reverberance (S and R) image generalization

H 1 *Width and location (distance and direction) of S and R images can be described independently using a similar language. Graphical and verbal descriptions of S and R imagery show consistent trends for different audio scenes.*

H 2 *Data from S and R descriptions can be analysed using similar methods (e.g. geometrical description using circular or Euclidian geometry) to show that amplitude pair-wise panning is possible with both S and R images and direction of S and R images can be independently changed.*

1.3.2. Perceptual interaction of S and R images

H 3 *S and R images can cause mutual informational masking if they are perceived to originate from the same locations.*

H 4 *Increasing the perceived spatial separation and/or homogeneity of the R image will spatially unmask the S and R components.*

H 5 *Spatial homogeneity of reverberance is positively correlated with perceived naturalness and preference.*

Caveat to Hypothesis 4: *Increased levels of reverberant sound can be achieved by this spatial unmasking without affecting the readability of the audio scene.*

2. EXPERIMENT FOR INVESTIGATING S AND R SPATIAL IMAGERY OF REAL AND PHANTOM SOURCES AROUND A LISTENER

2.1. The GUI

The computer program used in the experiments is the result of a series of experiments which aimed to provide an intuitive tool for graphically describing the shape and location of perceived source and reverberance images [65, 64, 66]. The source image could be described as having either a stable or unstable definition. The idea of representing the overall extent of the source image as well as a centre of gravity or focus was independently developed by Ford et al. [18]. Their studies found that listeners spontaneously described a source image with an identifiable point where the source image seems to be unequivocally located (which could vary from a point to a region), though it was not always elicited. It is this "centre of gravity" which we call the stable region of a source image. The broadening of a source image as it nears a perceived room boundary into a less stable, more diffuse image was simulated by Corey [16] by processing an audio channel with an early-reflection simulator, which had the effect of smoothing the transient cues which are known to be used for concise source localization in anechoic environments[2]. The notion of a spatially unstable source image (or "fuzzy" source image) was described to the subject verbally (see Appendix B). Localization of a reverberance object is not a familiar task and in a pilot experiment [64] we found that listeners were generally unsure how to categorize such an image in terms of stability. This led us to use just a single category to describe the reverberance image; a category which encompassed both the stable and unstable image components.

An acoustically transparent yet visually opaque curtain surrounded the listener, as shown in Fig. 2. The inner curtain was placed close to the subject (0.9m) because in an earlier experiment [66], investigating perceived source distance of source images in a wave field synthesis system, we found that listeners never reported hearing a sound image in front of a curtain. The visual dominance in source localization and the plasticity of the source-localization properties of the auditory system when a person is presented with conflicting visual-auditory source cues was also demonstrated by Shinn-Cunningham et al. [59]. Numbered markers at 10° intervals were marked with suspended string, to help with spatial correspondence between the apparent "real-world" location of the auditory image and the "GUI-world" image description. Subjects were given a laser-pointer to mark the perceived image locations projected on to the curtain, and would then read-off the azimuth to help representing the image on the GUI. Choisel and Zimmer [13] found that using a laser pointer increased subject consistency in reported image directions for real and phantom source images.

[2] The relative importance of onset, offset, and ongoing disparities for bursts of noise has been investigated in a number of studies- a summary is given in [43, pg. 243-244].

2.2. Purpose of experiment

The general hypothesis outlined in section 1.3 are re-written in the form of testable hypotheses which can be investigated using the GUI:

To investigate hypotheses 1 and 2, we compare the spatial imagery properties of source and reverberance images around a listener. The control case is for real sources; a single channel of an anechoic recording or a single channel of recorded reverberation is reproduced with a single loudspeaker. This should give the perceptual correlate of a Source or Reverberance image. From the GUI-elicited image descriptions, we compare the following spatial properties: image extent (ASW), image distance and image direction (i.e. direction of the image "centre of gravity"). We then investigate the S and R spatial imagery for phantom sources created by pair-wise amplitude panning around the listener. The large body of data on phantom-image localization around a listener for source images (as discussed in section 1.2) will be compared with the data for R-image spatial properties found in the present experiment.

Regarding the interaction of the S and R images, hypotheses 3 and 4 will be investigated by affecting the spatial separation of the anechoic and reverberation channel. For the case with phantom images, this means independently changing the intended S and R image directions. Since the pioneering work by Cherry [12], studies on spatial (un)masking have generally concentrated on using speech as a target, and speech or speech-like noise as a masker, using speech intelligibility as a metric for the degree of unmasking. Plomp [47] considered reproduced reverberation as a masking signal, and found that when the reverberation is reproduced to the side of the listener the intelligibility of a speech target at 0° increases. In the present study we also consider the source and reverberance image to be either the target or masker signal, and the listener is engaged in analytic listening [17] to distinguish the two into separate auditory objects. For most sound scenes, the degree of similarity between the source and its complimentary reverberance image is very high; in spectral terms the two are nearly identical, so peripheral *energetic* masking occurs as well as *informational* masking due to the perceptual similarity [17]. Because of the similarity between the anechoic and reverberation signals, we might therefore expect to see some spatial fusion of the perceived source and reverberance images; for instance, sound components from one image may be "used" in the formation of the other, e.g. leading to a source image-spread in the direction of the reported reverberance image.

H5 is left as an intuition for the present paper; it will be addressed in a forthcoming thesis.

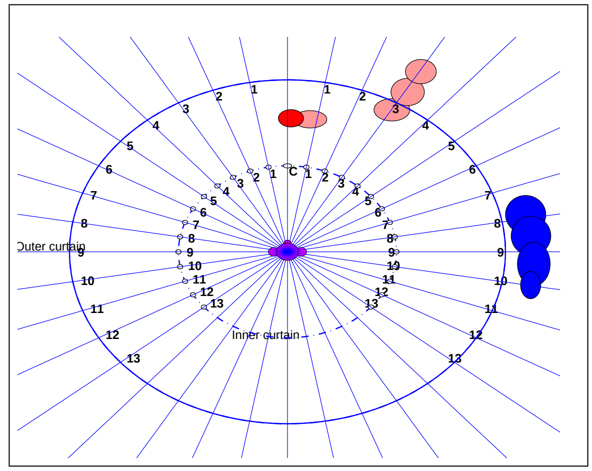

Fig. 1: Screen shot of the GUI. An interpretation of the graphical response shown here is: A source image was heard between -3° and +12°. The right-hand side of the source image (i.e. the right-most 9°) was spatially unstable; it may be that certain transients in the sound pulled the the source image to the right, or that when the listener rotated their head the source image "jumped" to the right. The listener also heard a source-image in the direction of the right-hand loudspeaker (+30°), but this image was unstable (maybe the listener heard an occasional resonance in this direction). The listener was less sure about the distance of this unstable image at 30°, hence they drew an image which had a larger depth than the stable image. A wide reverberance image was heard to the side of the listener from about 74° to 104°. This reverberance image was more distant than the source image.

2.3. Method

2.3.1. Stimuli

The original audio signal used was a monophonic, anechoic recording of a flute (Debussy's "Syrinx", 20 seconds excerpt, *legato*). In order to investigate the ability of listeners to describe reverberance imagery, we wanted to create a reverberation channel with the temporal properties of late-reverberation- that is, with the same temporal properties as the electrical output of a microphone in a diffuse sound field. We did therefore not want to simulate any early reflections. As can be seen from the IR measurements of the artificial reverberator (Fig. 15), only the first 10ms of different IR's was time invariant.
There are two general approaches to creating artificial reverberation for use in experiments concerning reverberance. They can be divided into the following categories for creating the impulse response (IR):

1. Stationary IR model (FIR filter).
2. Non-stationary IR model (IIR filter).

A similar distinction between the two methods is made by Blesser [8] in terms of the excitation signal employed for obtaining the IR (i.e. of a real or virtual enclosed space): the first class using a wide-band excitation signal, short or long in duration, and the other using a short pulse.

After exploring both approaches, we decided to use the non-stationary IR approach to create the reverberation from the anechoic recording. We used a commercially-available artificial digital reverberator (the M3000 manufactured by T.C. Electronics). Technical details of the IR's from this unit are shown in Appendix A.2. We also created a stationary IR using band-pass noise with frequency-dependant decay time, and convolved this IR with the anechoic signal. We found the reverberation from the M3000 to be more musically appropriate- the temporal and timbre balance sounded more like natural reverberance. As this experiment concerns summing localization with coherent sound sources [7], we only needed a single (mono) reverb channel. In accordance with the common idiom in sound recording practice, we call this channel the **Wet** channel and the anechoic channel the **Dry** channel.

2.3.2. Scene configurations

There are two separate experiments to investigate source and reverberance spatial imagery in loudspeaker audio. The first experiment concerns sound images created using only real sources from loudspeakers around the listener. The second experiment concerns the perception of phantom images created using a loudspeaker pair radiating a coherent signal. Each experiment investigates: (1) What the differences between the spatial imagery of source and reverberance objects are (2) How the two S and R images interact with each other when both S and R images are heard in the same scene. For investigating (1), only the dry or wet channel is active at any one time (called scenes **D** and **W**, respectively). For (2), both wet and dry sources are active together. For both we investigate the S and R images at different locations around the listener. This is summarized in table 2.

Scene Config.	Stimuli	Image directions
D	100% Dry	5 random directions
W	100% Wet	5 random directions
D+W	50% Dry 50% Wet	Dry from centre Wet: 5 random directions

Table 2: Scene configurations for experiment 1 and 2. In scene configuration **D+W**, the dry signal is reproduced from the centre channel only (experiment 1) or equally from the front loudspeakers (experiment 2). In other words, the intended direction of the source-image is at $0°$.

Control of image direction for the real-source experiment was by reproducing the audio signal from a loudspeaker incident at the listener from the required direction. For the phantom-image experiment, the only active loudspeakers were in the conventional ITU-R BS.775-1 configuration (no centre-speaker) as shown in Fig. 2. For the phantom-image experiment, the image direction was controlled by sending the dry or wet audio signal to a loudspeaker pair (i.e. the two loudspeakers closest to the intended image direction) with different gains. This is a form of pair-wise amplitude panning (PWAP). The five intended image angles we investigated are:

0°, 30°, 60°, 90° and 120°. The images at 30° and 120° were always real sources (i.e only a single channel was active). The loudspeaker gain coefficients were calculated using the tangent-panning law, as it had been found that the tangent law predicts image location better than Blumlein's classic stereophonic law of sines for mobile-head listeners [6]. The side angles were chosen because of the importance of lateral-incident sound reflections on subjective "spatial impression" [1, 2]. Furthermore, we wanted to investigate the spatial-unmasking effect of separating the source and reverberance image, and in the Plomp [47] study on spatial masking by reverberation, the unmasking effect was strongest when the reverberation was from 90° (RT = 1.4s). Using the target-masker analogy, we wondered if there would be any change in the reported spatial imagery of the target image (i.e. source image) if the masking image (i.e. reverberance image) was perceived to originate from a different spatial location.

2.3.3. Listening room set-up

The experiment was conducted in the MARLAB with a loudspeaker set-up is according to Fig. 2.

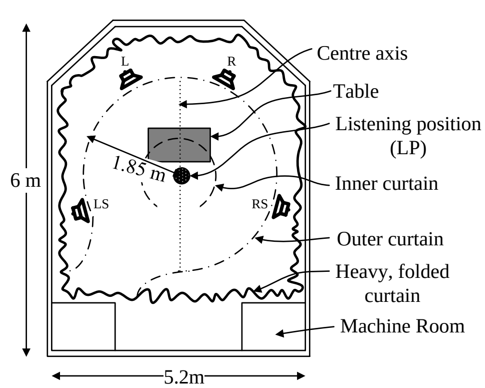

Fig. 2: Plan-view of listening room and loudspeaker arrangement, according to ITU-R BS.775-1 with rear-loudspeakers at ±120° to the centre axis, and no centre-speaker.

The listening room is acoustically heavily damped, with a RT_{20} of 100ms at 1kHz. The loudspeakers are manufactured by Bang and Olufsen (type Beolab 4000; active 2-way vented design). In both ex-

Trial#	Scene config.	Image direction: S image	R image
1	**D**	0	X
2	**D**	30	X
3	**D**	60	X
4	**D**	90	X
5	**D**	120	X
6	**W**	X	0
7	**W**	X	30
8	**W**	X	60
9	**W**	X	90
10	**W**	X	120
11	**D+W**	0	0
12	**D+W**	0	30
13	**D+W**	0	60
14	**D+W**	0	90
15	**D+W**	0	120

Table 3: Stimuli permutations. In experiment 1 the image direction is a real image direction, from a single loudspeaker source at the intended direction. In experiment 2 the direction is an intended direction: the image is a phantom source created by PWAP. In both experiments for the 30° and 120° image, only a single loudspeaker is active and it is therefore a real image not a phantom image. The symbol X indicates that either the source or reverberation channel is not active for this trial.

periments with real and phantom images, the loudspeakers are placed in the conventional 2/2 ITU-R BS.775-1 arrangement (rear loudspeakers at ±120°). However, in experiment 1, there are additional loudspeakers at 0°, 60° and 90°. All loudspeakers are at the same distance to the listening position (1.85 m), and are the same height (tweeter at 1.20 m). The angles were measured using a laser pointer on a tripod at the listening position (LP), and the distance fine-tuned by time-aligning impulses sent to each loudspeaker. Due to slight sensitivity differences between the loudspeakers, the level of the loudspeakers were equalized to 70 dBA, ±0.5 dB, slow-time-weighted, measured about the LP using pink noise (waving the sound level meter about the LP). Using pink-noise, the loudness was judged to be approximately equal for all loudspeakers- loudness is discussed in section 4.1.2.

2.3.4. Subject training and instructions

A training experiment was undertaken for both the real and phantom image experiment. The listeners were explicitly told whether there is dry, wet, or a mixture being reproduced, though they are not told the intended (panned) image direction. This enables the subjects to familiarize themselves with the stimuli and GUI. Whilst this training exercise is being undertaken, the experiment supervisor watches the subjects response (on a computer monitor) to see if there are any difficulties, e.g. hesitation, re-drawing, and also notes if the listener is seated correctly (e.g. not moving and beneath the LP).

The subjects were free to rotate their heads but told to keep it beneath the mark on the ceiling corresponding to the LP, as described in the instructions in Appendix B. The instructions were identical for the real and phantom-source experiments.

2.3.5. Summary of experiments

- Two experiments: the first with single loudspeakers, the second with pair-wise amplitude panning.
- Panned image locations (i.e. intended image direction): 0°, 30°, 60°, 90°, 120°. All on the right-hand side of the Listening Position (LP).
- Three scene configurations for each experiment:
  - Scene **D**: A mono anechoic channel of a flute recording is panned around the listener.
  - Scene **W**: A mono channel of artificial reverberation created from the flute recording is panned around the listener.
  - Scene **D+W**: The anechoic channel is panned at 0° and the reverberation channel is panned around the listener.
- "Expert" subjects took part in the experiments: all of whom graduate students in a *Tonmeister* sound recording program with at least three years critical listening experience. Subjects were paid $15 for each experiment. 6 subjects took part in the real-source experiment, and 5 subjects in the phantom-source experiment.
- 15 unique trials for each experiment (see table 3), repeated in 3 sessions with a 5-15 minute break between sessions. In each session the trial-order is randomized.
- Each trial plays continuously until the subject proceeds to the next trial.
- No feedback is provided to the subject at any time during the test.

3. RESULTS
3.1. Density Plots

3.1.1. Real-sources (experiment 1)

Intended image direction (in scene **D+W** this is the direction of the reverberation channel and the Dry channel is always from the centre speaker):

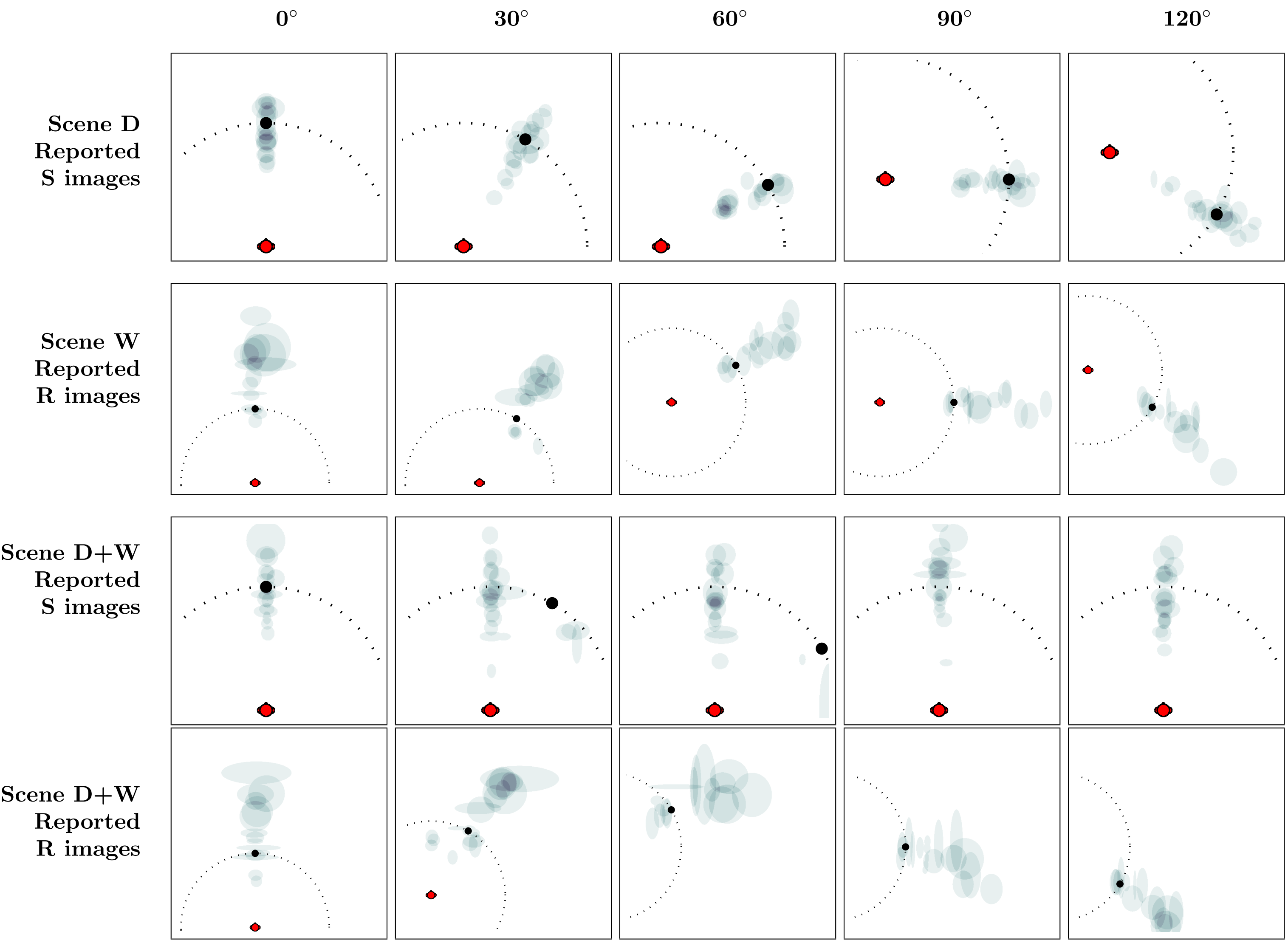

Fig. 3: Density plots for real source experiment. Responses summed from 6 subjects, 3 repeats. Arranged by scene configuration; **D**: Only Dry channel; **W**: Only Wet channel; **D+W**: Dry channel from centre speaker, wet channel from 1 of 5 locations. Black dot indicates active loudspeaker location. Spatial scale is the same for each scene (i.e. row) and density-scale the same for all plots.

3.1.2. Phantom-sources (experiment 2)

Intended image direction (in scene **D+W** this is the intended direction of the reverberance image and the intended source direction is always from the 0° location):

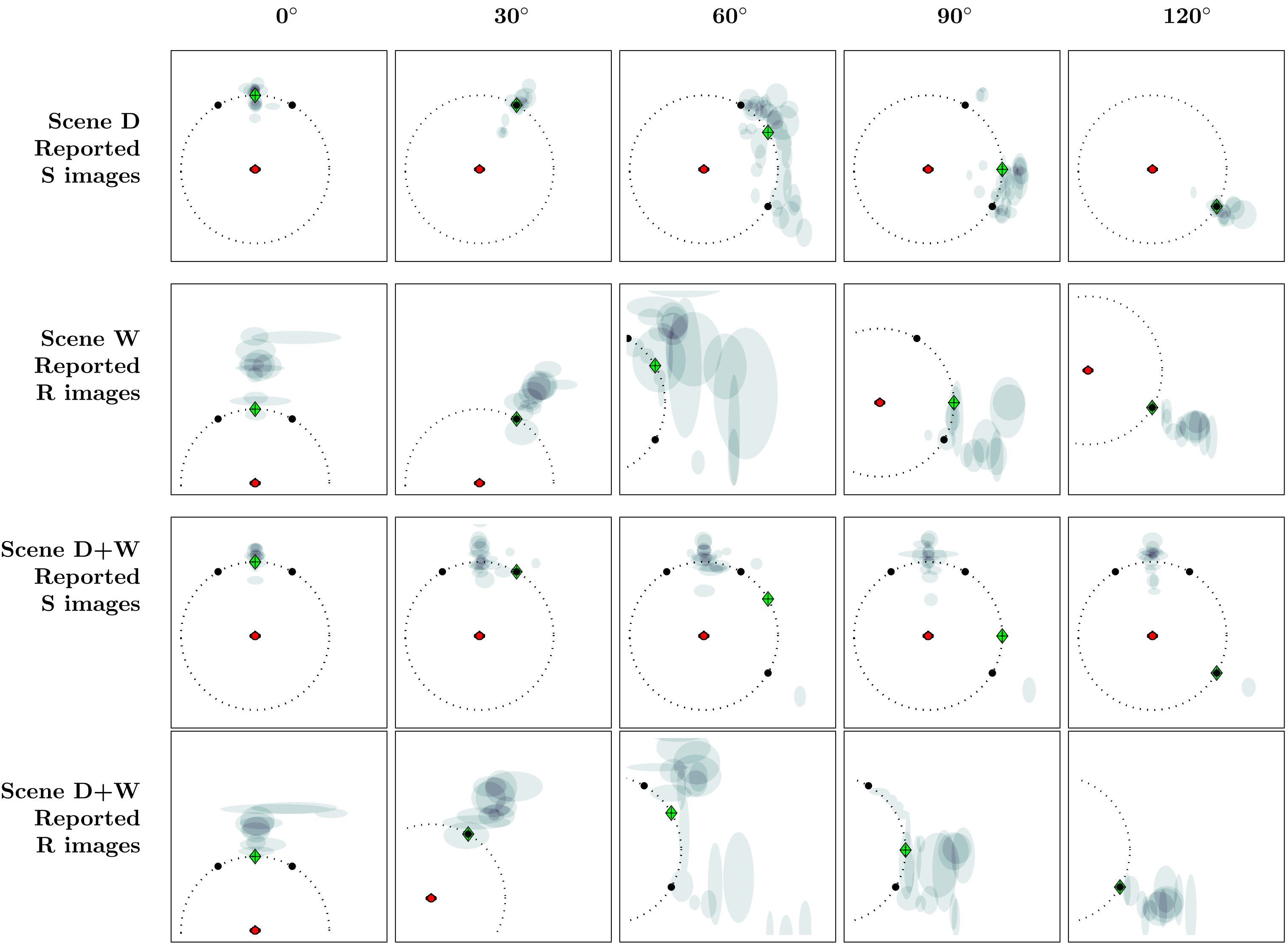

Fig. 4: Density Plots for phantom source experiment. Responses summed from 5 subjects, 3 repeats. Arranged by scene configuration; **D**: Only Dry channel; **W**: Only Wet channel; **D+W**: Dry channel from centre speaker, wet channel from 1 of 5 locations. Green diamond with cross indicates active source locations. For scenes **D+W**, the intended source direction is always from the centre location and the intended reverberance image from the direction shown by the diamond. Location of active loudspeakers shown with a block dot. Spatial scale is the same for each scene (i.e. row) and density-scale the same for all plots and the same as density plots for real-source experiment.

3.2. Analysis of Density Plots

In all the plots, means ± a single standard deviation are shown. 6 subjects took part in the real-source experiment and 5 for the phantom-source experiment. Both experiments had 3 runs (i.e. 2 repeats).

3.2.1. Single channel: configurations D and W

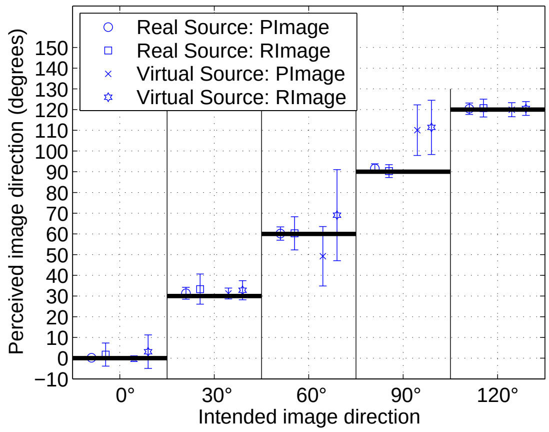

Fig. 5: Intended vs perceived image direction for real and phantom images.

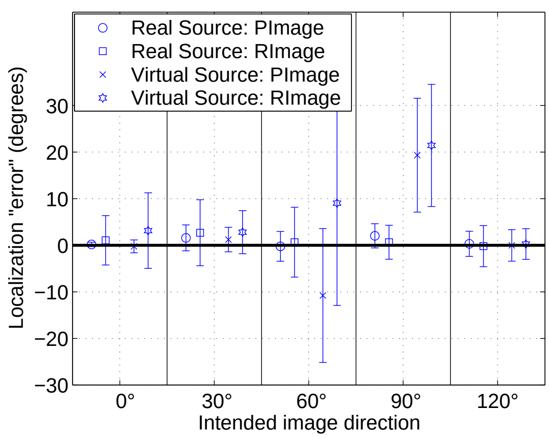

Fig. 6: Localization "error" of source and reverberance images (true veridicality is only possible for the real-source experiment).

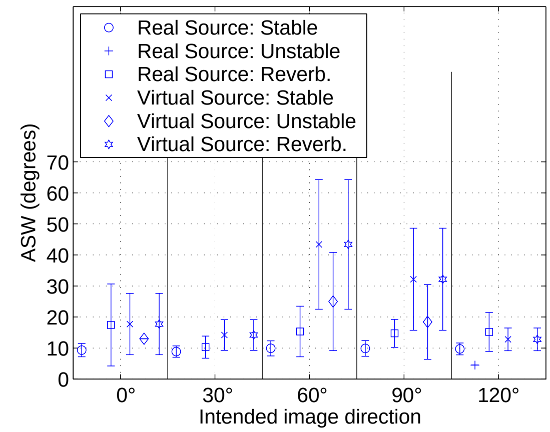

Fig. 7: ASW of source and reverberance images.

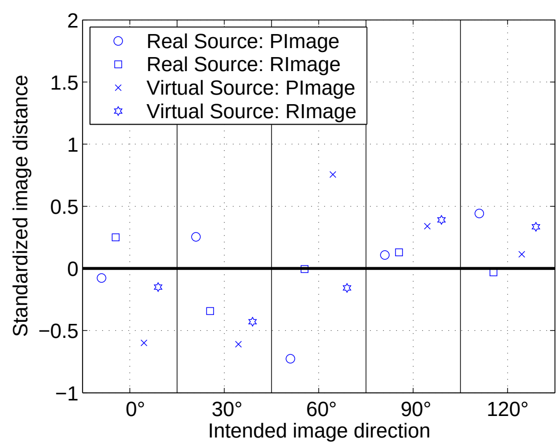

Fig. 8: Image distance. Standardized Z-score.

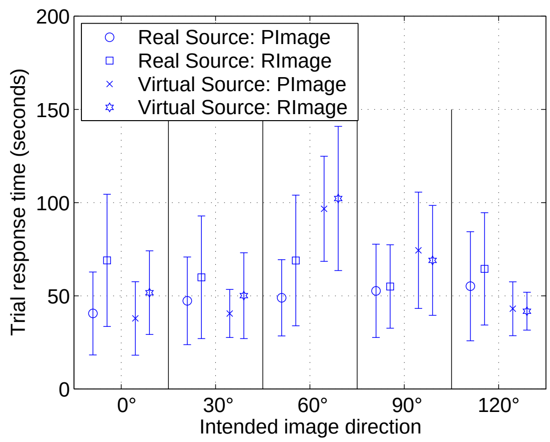

Fig. 9: Trial response time.

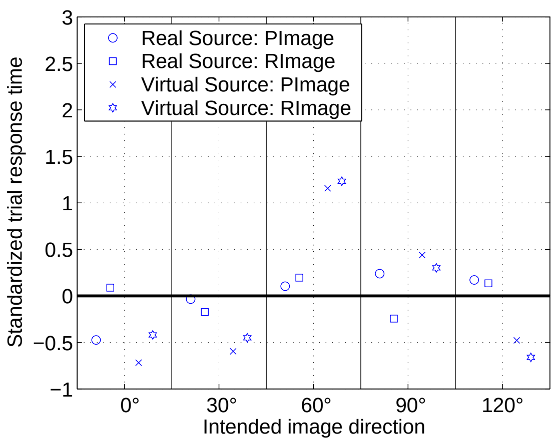

Fig. 10: Trial response time: standardized.

3.2.2. Configuration D+W

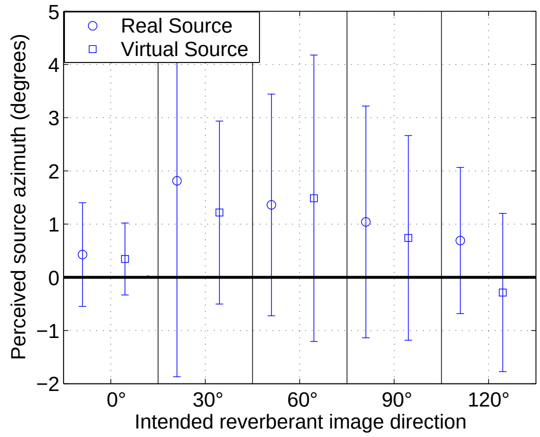

Fig. 11: Perceived stable-source image direction (intended source-image direction is zero degrees for all trials in scene configuration **D+W**).

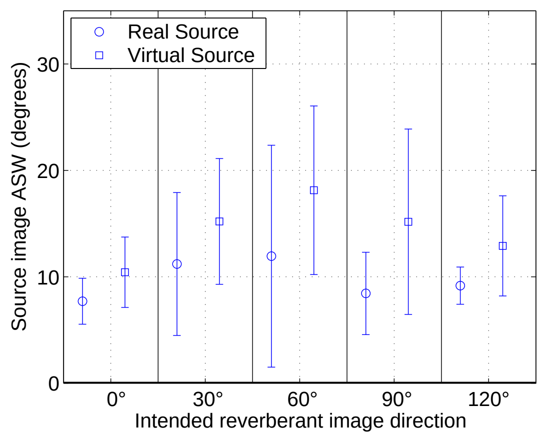

Fig. 12: Source image ASW. Scene configuration **D+W**.

4. DISCUSSION

4.1. Scenes D and W

This section will only consider the scenes when either just the dry channel (scene **D**) or just the wet channel (scene **W**) is reproduced as a function of the real and intended (panned) image source direction.

4.1.1. Localization "error"

For **real images**, subjects reported the direction of both source and reverberance images with a very high degree of accuracy. The maximum mean-error was $+3.1°$ for the reverberance image (at $30°$) and $+1.8°$ for the source image ($90°$), with error standard deviations of $8°$ and $3°$ for the reverberant and source images. The mean localization errors and standard-deviations were similar or smaller than found with other experiments which used single loudspeaker sources around the listener [33, 13, 52]. Fig. 5 shows how the reporting of phantom image direction is less consistent between subjects than for real sources, and that this variation is greater for reverberance images than for source images- though interestingly at $90°$ (intended direction) the reported source and reverberance image directions are very similar. S and R images panned at the front were reported with a similar consistency to real sources. However, side **phantom images** showed a consistent error bias, as shown by Fig. 6.

We find that for the phantom images panned at $60°$, the source image is reported about $10°$ (mean) closer to the front-right loudspeaker, and that the reverberance image is pulled by a similar amount in the direction of the rear-right loudspeaker. This is interesting, but we can not draw firm conclusions yet as the between-subject standard deviation for reported source and reverberance image azimuth was $15°$ and $20°$. This between-subjects variation is similar to that found in other studies for side phantom images [55, 51]. For the phantom images panned at $90°$ panned, both the source and reverberance images were reported at about $112°$; a detent of $-21°$. This is in agreement with Corey [16], who found that using a similar loudspeaker set-up and PWAP, images panned at $80°$ were also pulled towards the rear loudspeaker.

4.1.2. ASW

For both real and phantom images, the ASW of the reverberance images were generally larger than for the source images (see Fig. 7). As found in the study by Merimaa and Hess [41], the standard deviations of image ASW was proportional to the mean values. To investigate the effect of (intended) image direction on ASW, a type III sum of squares GLM procedure was used with factors SUBJECT and (panned) DIRECTION.

For the **real images**; the ASW of the source image was not significantly affected by the intended image direction, but was affected by the subject factor ($F = 19.386, p < 0.00$). The reverberance image ASW was smallest at $30°$ (mean of $12°$), and was least consistently reported at $0°$ (mean $17.6°$, SD $6.5°$). The relationship between real-image direction and ASW for reverberance images was also investigated as for the source image, and the relationship found to be "statistically interesting" ($F = 2.207, p = 0.079$). Again, the subject factor was significant ($F = 3.311, p = 0.010$), as was the subject*direction factor ($F = 2.607, p = 0.002$)- suggesting that subjects had different ideas about reverberance image ASW (or LEV [41]).

For the **phantom images**; the ASW of the side images ($60°$ and $90°$) was significantly larger than for front and $120°$ images (mean of about $40°$ ASW). For the source image, the GLM analysis revealed that subject, intended direction, and interactions were significant ($p < 0.001$). The same was found for reverberance images. The ASW was greater than the directional spread predicted [50] by looking at the discrepancy between ITD and ILD-predicted source azimuth. Some subjects reported hearing "just-audible" tonal colouration from the right loudspeaker. This has not been investigated yet, but it may account for the slight bias towards the right-hand speaker for $0°$ phantom images, as can be seen for the R-image at $0°$ in the phantom-source density plots (scene configuration **W**, Fig. 4). Furthermore, the inverse relationship between interaural coherence and ASW [26, 11, 27, 36, 40] implies that any inter-speaker non-linearities would increase ASW by reducing the acoustical-output coherence of the two loudspeakers.

Although the SPL of the reproduced anechoic and reverberant audio channels was the same (i.e. for all real and phantom locations), the dry sound was consistently heard to be louder than the reverb. This could be explained by a forward-masking suppression-of-reverberation effect, similar to the precedence (or Haas) effect with discrete echoes [43]. A simple example of forward masking with reverberation is to record speech in an ordinary room and to replay the recording backwards[23]: the perceived reverberance is louder when the speech is played in reverse. ASW is positively correlated with loudness [26, 65], but as reverberance images were generally larger than source images it is difficult to interpret the loudness-ASW relationship from the data in the present work.

4.1.3. Image distance

An investigation [60, 68, 3] into the effect of source direction on perceived loudness found that although the direction-effect varied for different rooms, it was generally <1 dB (for example, in order for the the rear loudspeaker to have the same perceptual loudness of the centre loudspeaker, the rear audio channel would have to be amplified by about 1 dB). This study used various noise sources with differing "colours". Similar findings were reported by Ratliff [55], who used octave-band filtered noise and concluded that in a "normal" listening room (70m^3, RT=0.35s) "the average auditory response is almost equally sensitive around the full azimuthal circle, although there is a tendency for the back to be less sensitive than the front by about 1 dB". The frequency-dependencies of this relationship was investigated by Robinson and Whittle [56], who found that the loudness of low-frequencies (<4kHz) were perceptually attenuated for rear loudspeakers (being maximal of about 5 dB at 4 kHz) and boosted at higher frequencies (-7 dB at 8 kHz). This is perhaps a counter-intuitive finding, as one might expect head-shadowing to reduce the loudness of HF sound from behind the head. Due to the approximately inverse-square relationship of source distance with proximal level, loudness is a major cue for determining image distance for close sources. For instance, using speech stimuli Zahorik [69] found the loudness cue was used by listeners to determine source distance in preference to a conflicting direct:reverberation level cue. We would therefore expect the perceptually softer reverberance images to be heard further away than the source images. A quick visual inspection of the density plots shows this is the case for both real and phantom images.

The direction-dependance of image distance for **real images** is minimal for source images for all directions except at 60°, where it is closest (mean 1.5m). A type III sum of squares GLM procedure was used, with fixed-factors SUBJECT and DIRECTION, to investigate investigate the direction-distance relationship, and this difference was found to be statistically significant ($p = 0.042$). The reverberance image distance was not significantly affected by panned direction.

For **phantom images**, contrary to the previous finding the source image is heard farthest at 60° (about 2.1m) and closest at 0° (1.84m). it was found that source-image direction (intended) significantly affects elicited image distance ($F = 6.723, p = 0.000$), as does subject and the interaction. As with real sources, the reverberance image distance is not affected by (intended) direction in a statistically significant way ($F = 1.568, p = 0.198$), though the subject factor does affect reverberance image distance ($p = 0.000$).

The large subject effect on distance judgement of auditory images has been reported by other studies (e.g. [69, 41]) and z-transformation of the distance data was employed in these studies and the present study to reveal image distance trends with the subject-factor removed. These standardized trends can be seen in Fig. 8, which shows the general trend of front images being closer than rear images; a trend which is more pronounced for phantom images than for real images.

4.2. Scene D+W

This section will only consider the **D+W** scene configuration. In this scene configuration, the intended source image direction was 0° for both the real and phantom image experiment. The wet channel was reproduced from either real speakers at 0°, 30°, 60°, 90° or 120°, or was pair-wise amplitude panned so that the intended reverberance image direction (according to the tangent panning law) corresponded to these directions.

In order to investigate the interaction of S and R images as a function of spatial separation, i.e. hypotheses 3 and 4, we will look at how the elicited spatial properties of the source image change as the (intended) reverberance image direction is changed.

A type III sum of squares GLM procedure was used with two factors: subject and intended reverberance image direction. Three analysis with dependant variables corresponding to source-image spatial properties were investigated to see if they were affected by the change in intended reverberance image direction. Results are summarized in table 4.

Dependant variable:	Significance: Real Source	Phantom Source
S. Image Azimuth	$p = 0.213$	$p = 0.021$
S. Image ASW	$p = 0.011$	$p = 0.001$
S. Image Distance	$p = 0.203$	$p = 0.336$

Table 4: Statistical significance of changes in Source image spatial properties as the reverberance image is panned around the listener.

For the real-source experiment, the number of cases was between 88 and 89 (out of a total of 90- i.e. 6 subjects * 3 repeats * 5 intended reverberance image directions). For the phantom image experiment the number of cases were between 71 and 73 (out of a total of 75 for 5 subjects). The data was discarded if more than one unique stable source image was drawn. The subject factor was significant for all analyses ($p < 0.001$).

The source image was pulled in the direction of the reverberance image, but this detent affect was only significant for the phantom image experiment, reaching a maximum detent at 60° (mean of +1.5°) as shown in Fig. 11. This confirms the testable hypothesis discussed in section 2.2 about the spatial fusion of the source and reverberance images when spatial definition is unclear. A similar trend is observed with (stable) source-image ASW, which is also maximum when the intended reverberance image is at 60° for both the real image (14° source ASW) and the phantom image (18° source ASW). This trend is shown in figure 12. As found in the discussion for scene configurations **D** and **W**, the large inter-subject variability means there are no consistent trends in perceived source image distance. However, when we standardize (i.e. Z-transform) the data as a function of subject we see that when the R-image is panned at 60°, the source image is heard to be farther than at any other panned location.

4.3. Summary of findings

When a single anechoic or a single reverberation audio signal is reproduced from one or two loudspeakers around a listener:

- The high accuracy and low between-subject variation in reporting both image direction and width for S and R images reproduced from a single loudspeaker shows that the GUI can be reliably used to describe the perceived S and R spatial imagery in the horizontal plane.
- Reverberance images were consistently reported wider than source images for both real and phantom images.
- Using PWAP, both source and reverberance phantom images intended at 60° and 90° are wider than phantom images at 0° and real images.
- S and R images panned at 90° (using PWAP with two loudspeakers at 30° and 120°) are reported at 110° (mean).
- Using PWAP, the between-subject consistency for reporting perceived direction of phantom images panned at 60° and 90° are similar for both source and reverberance images (between 20° and 15° SD).

When a single channel of anechoic sound is panned at 0° with a loudspeaker pair at ±30° and a reverberation channel is panned around the listener at 0°, 30°, 60°, 90° and 120°:

- Spatial fusion of the source and reverberance image is strongest when the reverberance image is panned at 60°, as shown by a pulling of the source image in the direction of the reverberance image and an increase in the perceived source image width.
- Spatial fusion of S and R images is weaker for real images than for phantom images (real images are more spatially *defined* [62, 54, 24, 66] than phantom images).

5. CONCLUSION

We have investigated the changes in spatial imagery of a reverberance object as it is pair-wise amplitude panned around a listener with loudspeakers arranged according to ITU-R BS.775-1, with rear loudspeakers at 120° and no centre speaker. The spatial properties of both real and phantom source and reverberance images were investigated as a function of panned (i.e. intended) image direction. We find that image width and direction can be described using a graphical mapping system to reveal consistent bias affects associated with intended image direction.

The interaction of source and reverberance images was investigated by panning a channel of artificial reverberation around a listener in the presence of an anechoic channel panned at 0°. Increasing spatial fusion between the source and reverberance images is found as the reverberance image is panned towards 60°.

The two findings:

- that control of both reverberance image direction and width involves similar amplitude panning principles as for source images
- that spatial fusion of a source and a reverberance image is strongest when the reverberance image is panned at 60°, but reduces at 90°

leads us to develop a specification for a reverberance image enhancement system. The design of this new system is the focus of a forthcoming thesis by the first author.

6. ACKNOWLEDGEMENTS

The authors thank all who took part in the experiment. This work was sponsored with financial grants from the National Sciences and Engineering Research Council of Canada and Valorisation-Recherche Québec, to whom the authors are grateful.

References

[1] Barron, M. (1971). The subjective effect of first reflections in concert halls - The need for lateral reflections. *Journal of Sound and Vibration*, 15:475–494.

[2] Barron, M. and Marshall, A. H. (1981). Spatial impression due to early lateral reflections in concert halls: The derivation of a physical measure. *Journal of Sound and Vibration*, 77:211–232.

[3] Bech, S. and Zacharov, N. (1998). Multichannel level alignment, part III: The effects of loudspeaker directivity and reproduction bandwidth. In *106th Convention of the Audio Engineering Society*, Munich, Germany.

[4] Beranek, L. L. (1996). *Concert and Opera Halls: How They Sound*. Acoustical Society of America through the American Institute of Physics, Woodbury.

[5] Berg, J. and Rumsey, F. (1999). Spatial attribute identification and scaling by repertory grid technique and other methods. In *Proceedings of the AES 116th international convention*, Rovaniemi, Finland.

[6] Bernfeld, B. (1973). Attempts for better understanding of the directional sterephonic listening mechanism. In *Proceedings of the AES 44th international convention*, Rotterdam, Netherlands.

[7] Blauert, J. (1997). *Spatial hearing: The psychophysics of human sound localization*. MIT Press, Cambridge, Mass., revised edition.

[8] Blesser, B. (2001). Interdisciplinary synthesis of reverberation viewpoint. *Journal of the Audio Engineering Society*, 49(10):867–903.

[9] Bregman, A. (1990). *Auditory Scene Analysis: The Perceptual Organization of Sound*. MIT Press, Cambridge, Mass.

[10] Cabot, R. C. (1977). Sound localization in 2 and 4 channel systems: A comparison of phantom image prediction equations and experimental data. In *Proceedings of the 58th Convention of the Audio Engineering Society*.

[11] Chernyak, R. and Dubrovsky, N. (1968). Pattern of the noise images and the binaural summation of loudness for the different interaural correlation of noise. In *Proceedings of the 6th International Congress on Acoustics*, pages A53–A56.

[12] Cherry, C. (1953). Some experiments on the recognition of speech, with one and two ears. *Journal of the Acoustical Society of America*, 25:554–559.

[13] Choisel, S. and Zimmer, K. (2003). A pointing-technique with visual feedback for sound-source localization experiments. In *Proceedings of the 115th Convention of the Audio Engineering Society*, New York.

[14] Clark, H., Dutton, G., and Vanderlyn, P. (1958). The "stereosonic" recording and reproduction system: A two-channel system for domestic tape records. *Journal of the Audio Engineering Society*, 6(2):102–117.

[15] Corey, J. and Woszczyk, W. (2002). Localization of lateral phantom images in a 5-channel system with and without simulated early reflections. In *113th Conference of the Audio Engineering Society*, Los Angeles.

[16] Corey, J. A. (2002). *An integrated system for dynamic control of auditory perspective in a multichannel sound field*. PhD thesis, Department of sound recording, McGill university.

[17] Durlach, N. I., Mason, C. R., Kidd, G., Arbogast, T. L., Colburn, H. S., and Shinn-

Cunningham, B. G. (2003). Note on informational masking. *Journal of the Acoustical Society of America*, 113:2984–2987.

[18] Ford, N., Rumsey, F., and Nind, T. (2003). Creating a universal graphical assessment language for describing and evaluating spatial attributes of reproduced audio events. In *Proceedings of the AES 115th international convention*, New York.

[19] Griesinger, D. (1996). Spaciousness and envelopment in musical acoustics. In *Proceedings of the AES 101st international convention*, Los Angeles.

[20] Griffiths, T. and Warren, J. (2004). What is an auditory object? *Nature Reviews Neuroscience*, 5:887–892.

[21] Guastavino, C. and Katz, B. (2004). Perceptual evaluation of multi-dimensional spatial audio reproduction. *Journal of the Acoustical Society of America*, 116(2):1105–1115.

[22] Hiyama, K., Komiyama, S., and Hamasaki., K. (2002). The minimum number of loudspeakers and its arrangement for reproducing the spatial impression of diffuse sound field. In *Proceedings of the Audio Engineering 113th Convention*.

[23] Houtsma, A., Rossing, T., and Wagenaars, W. (1987). Auditory demonstrations CD [track 35]. IPO and ASA.

[24] ITU-R BS. 1116 (1994). Methods for the subjective assessment of small impairments in audio systems including multichannel sound systems. Recommendation BS 1116, International Telecommunication Union Radiocommunication Assembly.

[25] ITU-R BS.775-1 (1994). Multichannel stereophonic sound system with and without accompanying picture. Recommendation BS.775-1, International Telecommunication Union Radiocommunication Assembly.

[26] Keet, W. d. V. (1968). The influence of early lateral reflections on the spatial impression. In *Proceedings of the 6th International Congress on Acoustics*, pages E–2–4, Tokyo.

[27] Kendall, G. (1995). The decorrelation of audio signals and its impact on spatial imagery. *Computer Music Journal*, 19(4):72–87.

[28] Kubovy, M. (1981). Concurrent-pitch segregation and the theory of indispensable attributes. In Kubovy, M. and Pomerantz, J., editors, *Perceptual organization*, pages 55–99. Lawrence Erlbaum, Hillsdale, NJ.

[29] Kubovy, M. and Valkenburg, D. V. (2001). Auditory and visual objects. *Cognition*, 80:97–126.

[30] Lee, H.-K. and Rumsey, F. (2004). Elicitation and grading of subjective attributes of 2-channel phantom images. In *Proceedings of the AES 116th international convention*, Berlin.

[31] Letowsky, T. (1989). Sound quality assessment: concepts and criteria. In *Proceedings of the AES 87th international convention*, New York.

[32] Lund, T. (2000). Enhanced localization in 5.1 production. In *Proceedings of the 109th Convention of the Audio Engineering Society*.

[33] Makous, J. and Middlebrooks, J. (1990). Two-dimensional sound localization by human listeners. *Journal of the Acoustical Society of America*, 87:2188–2200.

[34] Marshal, A. and Barron, M. (2001). Spatial responsiveness in concert halls and the origins of spatial impression. *Applied Acoustics*, 62:91–108.

[35] Marshall, A. H. (1967). A note of the importance of room cross-section in concert halls. *Journal of Sound and Vibration*, 5(1):100–112.

[36] Martens, W. L. (2001). Two-subwoofer reproduction enables increased variation in auditory spatial imagery. In *Proceedings of the 2001 International Workshop on Spatial Media*, Aizu-Wakamatsu, Japan.

[37] Martin, G. (2002). Interchannel interference at the listening position in a five-channel loudspeaker configuration. In *Proceedings of the AES 113th international convention*, Los Angeles.

[38] Martin, G., Woszczyk, W., Corey, J., and Quesnel, R. (1999a). Controlling phantom image focus

in a multichannel reproduction system. In *Proceedings of the AES 107th international convention*, New York.

[39] Martin, G., Woszczyk, W., Corey, J., and Quesnel, R. (1999b). Sound source localization in a five-channel surround sound reproduction system. In *Proceedings of the AES 107th international convention*, New York.

[40] Mason, R., Brookes, T., and Rumsey, F. (2004). Development of the interaural cross-correlation coeffcient into a more complete auditory width prediction model. In *Proceedings of the 18th ICA*, pages 2453–2456, Kyoto, Japan.

[41] Merimaa, J. and Hess, W. (2004). Training of listeners for evaluation of spatial attributes of sound. In *Proceedings of the AES 117th international convention*, San Francisco.

[42] Meyer, E. and Schodder, G. R. (1952). Über den Einfluss von Schallrückwürfen auf Richtungslohalisation und Lautastärke bei Sprache. *Nach. Akad. Wiss. Gottingen, Math. Phys. Klasse IIa*, 6:31–42.

[43] Moore, B. C. J. (1997). *An introduction to the psychology of hearing*. Academic Press, San Diego, Calif., 4th edition. Brian C.J. Moore.

[44] Morimoto, M. (2001). How can auditory spatial impression be generated and controlled? In *Proceedings of the 2001 International Workshop on Spatial Media*, Aizu-Wakamatsu, Japan.

[45] Morimoto, M. and Asaoka, A. (2004). Multidimensional analysis of "reverberance". In *Proceedings of the 18th International Congress on Acoustics*, Kyoto, Japan.

[46] Morimoto, M. and Maekawa, Z. (1989). Auditory spaciousness and envelopement. In *Proceedings of the 13th International Congress on Acoustics*, volume 2, pages 215–218, Belgrade.

[47] Plomp, R. (1976). Binaural and monaural speech intelligibility of con-nected discourse in reverberation as a function of azimuth of a singlecompeting sound wave speech or noise. *Acustica*, 31:200–211.

[48] Pulkki, V. (1997). Virtual sound source positioning using vector base amplitude panning. *Journal of the Audio Engineering Society*, 45:456–466.

[49] Pulkki, V. (1999a). Coloration of amplitude-panned virtual sources. In *Proceedings of the AES 110th international convention*, Amsterdam. The Netherlands.

[50] Pulkki, V. (1999b). Uniform spreading of amplitude panned virtual sources. In *Proceedings of IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*.

[51] Pulkki, V. (2001). Localization of amplitude-panned virtual sources I: Stereophonic panning. *Journal of the Audio Engineering Society*, 49:739–752.

[52] Pulkki, V. and Hirvonen, T. (2005). Localization of virtual sources in multi-channel audio reproduction. *IEEE Transactions on Speech and Audio Processing*, 13:105–119.

[53] Queen, D. (1979). The effect of loudspeaker radiation patterns on stereo imaging and clarity. *Journal of the Audio Engineering Society*, 27:368–379.

[54] Rasch, R. A. and Plomp, R. (1984). The listener and the acoustic environment. In Deutsch, D., editor, *The Psychology of Music*, pages 135–147. Academic Press.

[55] Ratliff, P. A. (1974). Properties of hearing related to quadraphonic reproduction. Technical report, BBC Research Department, Report BBC RD 1974/38.

[56] Robinson, D. W. and Whittle, L. S. (1960). The loudness of directional sound fields. *Acustica*, 10:74–80.

[57] Rumsey, F. (2001). *Spatial Audio*. Focal press.

[58] Rumsey, F. (2002). Spatial quality evaluation for reproduced sound: Terminology, meaning, and a scene-based paradigm. *Journal of the Audio Engineering Society*, 50(9):652–666.

[59] Shinn-Cunningham, B., Durlach, N., and Held, R. (1998). Adapting to supernormal auditory localization cues I: Bias and resolution. *Journal of the Acoustical Society of America*, 103(6):3656–3666.

[60] Suokuisma, P., Zacharov, N., and Bech, S. (1998). Multichannel level alignment, part I: Signals and methods. In *Proceedings of the Audio Engineering Society 105th Convention*, San Francisco.

[61] Theile, G. and Plenge, G. (1977). Localization of lateral phantom sources. *Journal of the Audio Engineering Society*, 25(4):196–200.

[62] Toole, F. E. (1983). Subjective measurements of loudspeakers - A comparison of stereo and mono listening. In *Proceedings of Audio Engineering Society 74th Convention*, New York.

[63] Ueda, Y. and Ando, Y. (1997). Effects of air conditioning on sound propagation in a large space. *Journal of the Acoustical Society of America*, 102:2771–2775.

[64] Usher, J., Martens, W., and Woszczyk, W. (2004). The influence of the presence of multiple sources on auditory spatial imagery. In *Proceedings of the 18th International Congress on Acoustics*, Kyoto, Japan.

[65] Usher, J. and Woszczyk, W. (2003). Design and testing of a graphical mapping tool for analyzing spatial audio scenes. In *Proceedings of the AES 24th International Conference on Multichannel Audio*, Banff, Canada.

[66] Usher, J. and Woszczyk, W. (2004). Visualizing auditory spatial imagery of multi-channel audio. In *Proceedings of the AES 116th international convention*, Berlin, Germany.

[67] Zacharov, N. (2001). Understanding spatial sound reproduction: Perception and preference. In *Proceedings of the International Workshop on Spatial Media*, pages 43–61, Aizu-Wakamatsu, Japan.

[68] Zacharov, N., Bech, S., and Suokuisma, P. (1998). Multichannel Level Alignment, Part II: The Influence of Signals and Loudspeaker Placement. In *Proceedings of the AES 105th international convention*, San Francisco.

[69] Zahorik, P. (2002). Assessing auditory distance perception using virtual acoustics. *Journal of the Acoustical Society of America*, 111(4):1832–1846.

Appendix A. STIMULI

Appendix A.1. Anechoic flute recording

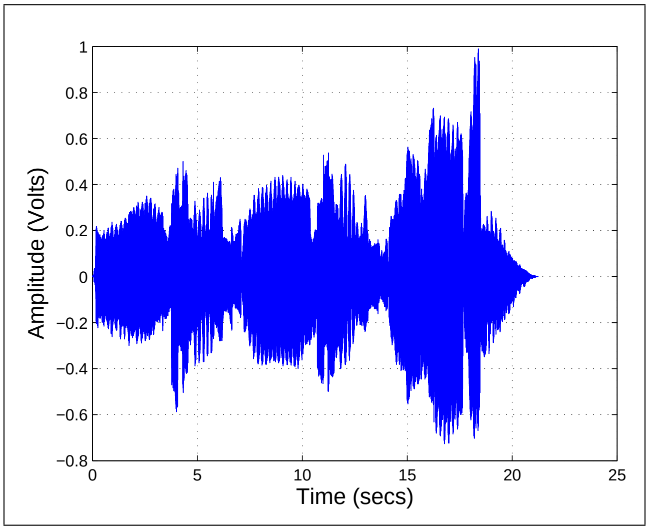

Fig. 13: Time-domain plot of anechoic flute recording used in experiments.

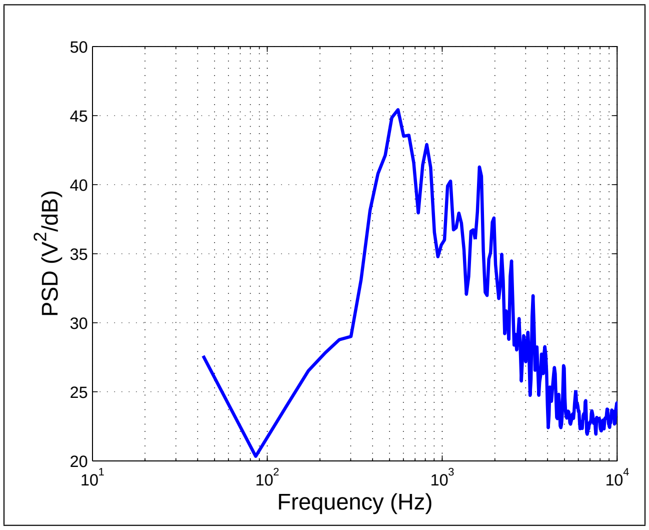

Fig. 14: Power spectrum of anechoic flute.

Appendix A.2. Artificial Reverberation

Scene:	#127 "Church piano"
RT	4.0s
Early Reflection level	-100dB
Wet mix	100%
Pre delay	12 ms
Modulation width	90%

Table 5: Configuration of tc electronics M3000 artificial reverberator for generation of single reverb channel used in the experiments.
.

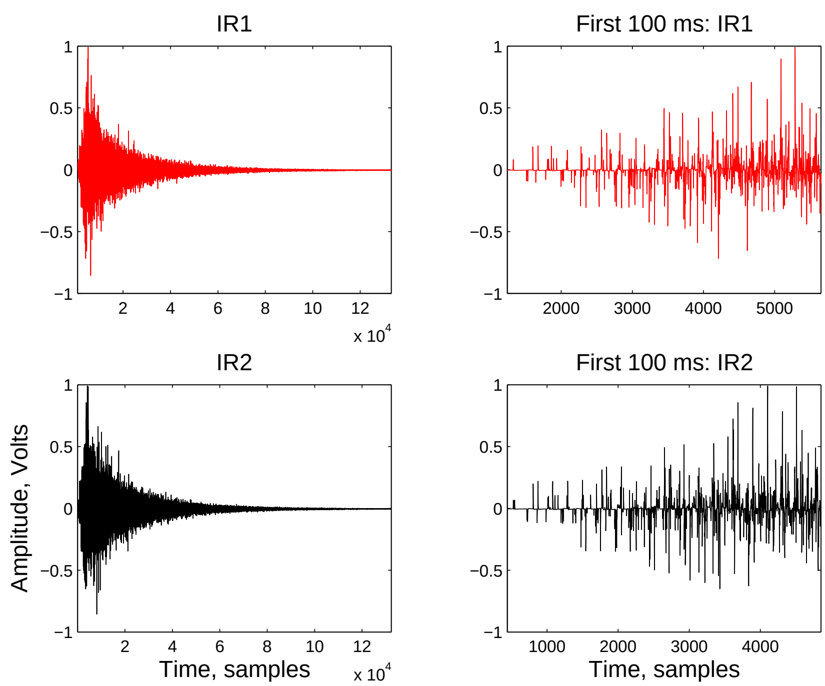

Fig. 15: Consecutive Impulse Responses of tc electronics M3000 artificial reverberator (as used in experiments). Created from a single Dirac function in the digital domain (44.1 kHz fs, 16-bit signal). IR is time variant for T>10 ms, as would be expected in a normal large room where air convection causes changes in the late part of the IR [63, 8]. All x-axis refer to sample time (Fs = 44.1 kHz), and y-axis to Voltage. Note from the time-energy envelope that the maximum energy is not reached until about 50ms, so there is minimal transient localization cue for the reverb channel (62ms with the reverb pre-delay).

Appendix B. SUBJECT INSTRUCTIONS

Dearest Subject,
This experiment is about where you hear phantom images in a multi-channel, "surround sound" audio experience. You will hear a flute reproduced from one or more loudspeakers behind the curtains around you. The loudspeakers are spaced around you, but sound will generally only come from speakers on your right-hand side. A computer program has been developed to allow you to graphically describe the sound images you hear. Using this program (a Graphical User Interface, or GUI), you will be able to describe the perceived location of the phantom image and describe how large it is by drawing ellipses. Drawing these ellipses is like drawing ellipses using other drawing software such as Photoshop, Paint, GIMP etc. You will see a top-down, plan-view of the listening room, with the curtains and other numbered reference markers indicated. Use the reference markers on the curtain and the GUI to help match where you hear the sound images to the GUI. Use the laser-pointer to mark the image direction and width- when you do this try to ignore the number-markers, as these may bias where you hear the image. Just use these numbers after you have decided where the image seems to be. **It is the direction, width and distance of the image that you should describe.** Remember, you are describing where you hear the phantom sound image, not where you think the loudspeakers are! Therefore, if you think the image sounds like it is coming from in front of the curtains when you close your eyes, then draw the image there on the GUI. You may use as many ellipses as you wish to describe the image shape.
The images are described in two ways: either as a **source image** or a **reverberance image**. A source image is a sound image which seems to be where the (phantom) source exists. In this experiment, this means where the flute seems to exist in the sound scene. You can describe the source image as being either stable or unstable. Unstable applies if the source image seems jumpy or fuzzy in a certain region of space, or if it moves when you rotate your head. A reverberance image is a sound image which sounds like live-reverberation (reverberance is the perceptual equivalent of acoustical reverberation). If you listen to the output of an artificial reverberator, set to "100% wet", then you will hear a reverberance sound image. You can use as many ellipses as you wish to describe where you hear the source and reverberance sound images. Sometimes you will hear only the source image, sometimes only the reverberance image, and sometimes both. In the case when you hear both, try and describe each sound image separately (it doesn't matter which order you describe them)- the two sound images are not necessarily reproduced by the same loudspeakers!
To reiterate, there are three things you have to describe:
1. The image direction.
2. The image width.
3. The image distance.

The pointer should be used for the first two. When describing the image direction (i.e. the centre of the image), you should do this when facing forward, with your head in line with the centre ("C") and number 9 markers- this is the "sweet-spot". If the image direction does not change when you move your head, then you draw the image with the **stable** source image ellipse button. If the image moves when you turn your head, mark all the directions over which the image is heard with the **unstable** ellipse button. When describing the image width and distance, you can rotate your head and upper-body in the same way you freely move your head when mixing music in the studio- please don't leave or twist the chair though and remember to keep your head underneath the marked cross on the ceiling. There is only one recorded sound- an anechoic flute recording- which has been processed with an artificial reverberator. You will have a training session at first to get used to the GUI and the sounds, which consists of 15 different sound-scenes. In the main test, you will hear the same sounds, and the main test is repeated twice (i.e. there are 3 test "sessions", i.e. 2 repeats). The sound is controlled by another computer (a Mac) so when you have finished describing the images with the GUI you must hit the "next scene" button on the other machine.
Cheers, -jHon