DAFx'04 Invited paper

SPATIAL IMPULSE RESPONSE RENDERING

Juha Merimaa[1,2]

[1]Institute of Communication Acoustics
Ruhr-Universität Bochum, Germany
juha.merimaa@hut.fi

Ville Pulkki[2]

[2]Lab. of Acoustics and Audio Signal Processing
Helsinki University of Technology, Finland
ville.pulkki@hut.fi

ABSTRACT

Spatial Impulse Response Rendering (SIRR) is a recent technique for reproduction of room acoustics with a multichannel loudspeaker system. SIRR analyzes the direction of arrival and diffuseness of measured room responses within frequency bands. Based the analysis data, a multichannel response suitable for reproduction with any chosen surround loudspeaker setup is synthesized. When loaded to a convolving reverberator, the synthesized responses create a very natural perception of space corresponding to the measured room. In this paper, the SIRR method is described and listening test results are reviewed. The sound intensity based analysis is refined, and improvements for the synthesis of diffuse time-frequency components are discussed.

1. INTRODUCTION

In recent years, multichannel loudspeaker reproduction systems have become increasingly common. A standard 5.1 setup is able to produce a surrounding sound field with fair directional accuracy especially in front of the listener. By adding more channels, the precision can be further enhanced, or the reproduction can be extended to 3-D. However, due to limitations of microphone technology, current recording systems cannot achieve as high directional resolution as that available for the playback. Furthermore, with most recording techniques the loudspeaker setup needs to be known already at the time of recording, and conversion for other setups is very difficult if not impossible. Spatial Impulse Response Rendering (SIRR) [1]–[3] has been designed to overcome some of these problems.

In a typical recording scenario several spot microphones are placed close to sound sources to yield fairly "dry" source signals with ideally no audible room effect. An artificial scene is then constructed by positioning these signals in desired directions using, for instance, amplitude panning. Spatial impression is created by adding the signals of additional microphones placed further away from the sources in the recording room, or with the help of reverberators. With convolving reverberators it has recently become possible to use actual measured room responses to simulate a chosen acoustical environment. However, the problem is—as in any surround sound recording application—how to capture the responses so that the perceived spatial impression of the measured room or hall is accurately reproduced.

SIRR is primarily targeted for processing room responses to be used in convolving reverberators. The responses can be measured with commercially available SoundField or Microflown systems or with a suitable custom microphone array. The method yields multichannel impulse responses that can be tailored for an arbitrary surround loudspeaker system in the postprocessing phase. SIRR can also be applied to continuous sound but this part is still under development.

In this paper, the SIRR method and some refinements are described, and earlier listening test results are reviewed. The paper is organized as follows. Secs. 2 and 3 provide background related to conventional multichannel recording techniques and psychoacoustics of spatial hearing. Sec. 4 with description of the SIRR method forms the main part of the paper. Listening test results are reviewed in Sec. 5 and the paper is summarized in Sec. 6.

2. PROBLEMS WITH CONVENTIONAL TECHNIQUES

Spatial audio or multichannel impulse responses have been typically recorded using one microphone per loudspeaker. Several different microphone configurations have been proposed in the literature. It has been shown that coincident microphone techniques are able to produce sharpest virtual sources [4, 5]. In coincident microphone setups, directive microphones are positioned as close to each other as possible. The sound signal from a single sound source is thus captured in the same phase with all microphones. The microphones should have orientations and directivities corresponding to the loudspeaker configuration, so that sound from any specific direction would only be picked up by few microphones. Using more loudspeakers requires thus narrower directional patterns. However, with existing microphone technology, narrow enough broad band patterns cannot be achieved. Consequently, the sound from any direction is always picked up by several microphones, which results in a blurred and colored reproduction due to the crosstalk between loudspeaker channels.

Ambisonics [6] tries to solve the directivity problem by employing a spherical harmonic decomposition of the sound field. In theory it can accurately reproduce a directional sound field in a small sweet spot by the sympathetic operation of all loudspeakers in an arbitrary surround setup. In practice, however, microphone technology limits the order and thus the directional resolution of Ambisonics. The authors are only aware of first order commercial implementations, although higher order microphone systems have been recently proposed [7, 8]. Furthermore, the presence of the head of the listener further disrupts the ideal operation, and consequently the technique reduces to using a set of virtual coincident microphones that can be adjusted during playback. The problems are also similar to those discussed in the previous paragraph.

In contrast to coincident techniques, spaced microphones are positioned at a considerable distance between each other. The sound signal from a single sound source is thus captured in different phases by different microphones. In a reverberant environment the resulting microphone signals will also be to a certain degree decorrelated. The noncoincident techniques are often said to cre-

ate a better feeling of "airiness" and "ambience", and the reproduction is less sensitive to the location of the listener. However, the directional accuracy is even lower than what can be achieved with a coincident microphone setup.

3. PSYCHOACOUSTICAL BACKGROUND

The goal of sound recording and reproduction is normally to relay a perception. However, in order to recreate the perceived spatial impression of an existing room or a hall, it is not necessary to perfectly reconstruct the original soundfield. Human sound localization is based on four frequency-dependent cues: (1) the interaural time difference (ITD) and (2) the interaural level difference (ILD), which resolve the left/right direction of a sound source, (3) monaural spectral cues, and (4) the effect of head rotation on the previous cues [9]. Additionally, human listeners are sensitive to the coherence of the left and right ear input signals (e.g. [10] and references therein), which has been proposed to be an important cue for localization in reverberant environments and multi-source scenarios [11]. In a room, reflections from several different directions affect these cues. For any nonstationary source signal, the summation of sound in different phase at the ears of the listener lowers the coherence and produces time-varying fluctuations in ITD, ILD, and the spectral cues. In SIRR it is assumed that these time and frequency dependent cues are what needs to be reproduced.

The limited resolution of human hearing has been studied extensively for monaural conditions (e.g. [12]). The frequency resolution of binaural hearing appears to be equal to that of monaural hearing [13, 14], although slightly larger analysis bandwidths have been found for some test signals [15]. This suggests that the monaurally derived ERB frequency resolution [16] is also appropriate for the analysis and synthesis of binaural cues. Determining the time resolution of binaural hearing is a little more complicated. A human listener is only capable of tracking in detail the spatial movements of sound sources corresponding to fluctuations of the ITD and ILD cues up to 2.4 and 3.1 Hz, respectively [17]. However, Grantham and Wightman [18] observed that listeners were able to detect ITD fluctuations up to 500 Hz, not based on movement but on perceptual widening of the sound sources.

As already mentioned, the interaural coherence and the time-variance of the other localization cues are related. Depending on the length of an analysis window, high frequency fluctuations transform into lowered coherence. If the fluctuations cannot be exactly recreated, it is important to reproduce the lowered coherence not only due to human sensitivity to it but also due to its stabilizing effect on the spatial sound image.

4. SPATIAL IMPULSE RESPONSE RENDERING

As discussed earlier, the current technology has shortcomings in recording and reproduction of spatial sound. The problems could be alleviated by designing microphones with higher directivity, but this is not an easy task. However, the previous psychoacoustical considerations suggest a different solution. In SIRR the direction of arrival and diffuseness of sound are analyzed at narrow frequency bands within short time windows. Based on an omnidirectional microphone signal and the analysis data, a perceptually similar sound field is then synthesized using a chosen reproduction system. The direction of arrival determines the localization cues appearing at each analysis band, and the diffuseness estimate is related to the interaural coherence. Hence, excluding the limitations of reproduction systems we assume that, if the frequency bands are narrow enough and the time windows are short enough, the reproduced spatial impression is very close to that of the recording room.

The analysis and synthesis can be implemented in several different ways. For the time-frequency processing we have adopted a short-time Fourier transform (STFT) based scheme common in audio coding applications. Similar processing could also be realized using an analysis-synthesis implementation of an auditory filter bank. However, Baumgarte and Faller [19] found the computationally more efficient FFT implementation to perform equally well with an auditory filter bank in their experiments with the Binaural Cue Coding (BCC) algorithm sharing some features with SIRR.

The analysis and synthesis parts of SIRR are illustrated in Figs. 1 and 2, respectively. The directional analysis dicussed in this paper is based on the concept of sound intensity as analyzed from SoundField microphone recordings. The synthesis for a multichannel loudspeaker system consists of spatialization of a recorded omnidirectional signal using amplitude panning and decorrelation techniques. The analysis will be described in more detail in Sec. 4.1 and the synthesis in Sec. 4.2.

4.1. Directional analysis based on sound intensity

The analysis data needed for SIRR consists of direction of arrival and diffuseness estimates as a function of time and frequency. Energetic analysis of a sound field can be used to obtain both of these estimates. In this Section, the sound intensity analysis is first introduced, followed by derivation of the required quantities from the B-format SoundField microphone signals.

The instantaneous sound intensity is defined as the product of the sound pressure $p(t)$ and the particle velocity vector $\mathbf{u}(t)$

$$\mathbf{I}(t) = p(t)\mathbf{u}(t) \tag{1}$$

[20]. The intensity describes the transfer of energy in the sound field, and the direction of arrival can be estimated simply as the opposite of the direction of $\mathbf{I}(t)$. Depending on the sound field, the direction of the instantanous intensity may vary as a function of time, which means that part of the sound energy oscillates locally and only part of it constitutes a net flow. The net flow can be characterized with the active intensity (or radiating intensity [21]) defined as the time average of the instantaneous intensity.

The proportion of sound energy contributing to the net transport of energy can be used to characterize the diffusness of the sound field. In earlier papers, we derived the total energy from the sound pressure signal as the active intensity of an ideal plane wave having the same sound pressure. However, this relation is valid only in monochromatic sound fields. The instantanous energy density of a general sound field can instead be calculated as

$$w(t) = \frac{1}{2}\rho\left[z^{-2}p^2(t) + \mathbf{u}^2(t)\right], \tag{2}$$

where ρ is the mean density and $z = \rho c$ is the impedance of the medium, where c denotes the speed of sound [22]. An average diffuseness estimate can now be written in the form

$$\psi = \frac{\|\langle \mathbf{I}(t)/c \rangle\|}{\langle w(t) \rangle} = \frac{2z\,\|\langle p(t)\mathrm{u}(t) \rangle\|}{\langle p^2(t) \rangle + z^2 \langle \mathbf{u}^2(t) \rangle}, \tag{3}$$

where $\|\cdot\|$ denotes the norm of a vector and $\langle\cdot\rangle$ denotes time averaging. This estimate equals the speed of energy transfer divided

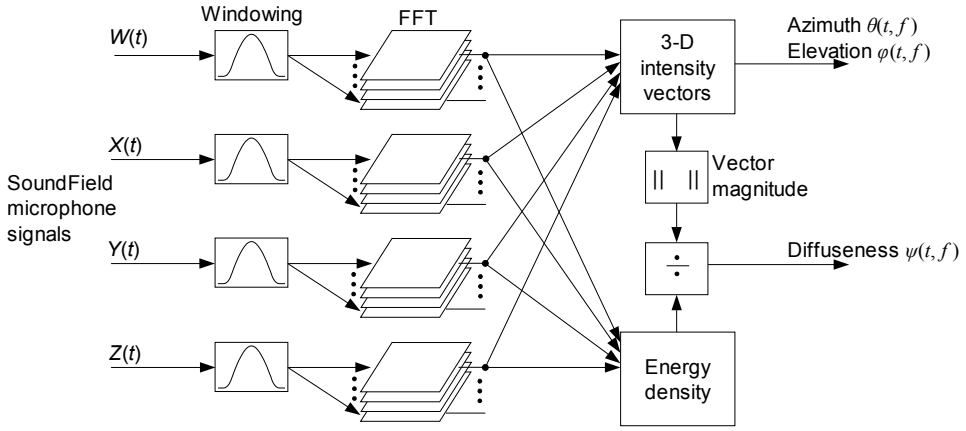

Figure 1: *Directional analysis of a B-format SoundField microphone signal using the concept of sound intensity.*

by the speed of sound, and it can be shown to be bound to values between $[0, 1]$, where 0 indicates an ideally diffuse sound field (no net transport of energy), whereas 1 signifies the absence of any locally oscillating sound energy [21]. Note that an instantaneous value $\psi(t)$ could also be defined but it would be of little use for synthesis purposes. A sound field with exact instantaneous properties according to Eqs. (1) and (2) is very difficult to synthesize, but a sound field with approximately similar time averages can, however, be created with the help of ψ.

The time-frequency analysis can be realized either by feeding the sound pressure and particle velocity signals through a filter bank and applying the equations above, or with a short-time Fourier transform (STFT) based scheme. In STFT implementation, the single-sided frequency distribution of the active intensity in an analysis window can be written as

$$\mathbf{I}_a(\omega) = 2Re\left\{P^*(\omega)\mathbf{U}(\omega)\right\}, \tag{4}$$

where $P(\omega)$ and $\mathrm{U}(\omega)$ are the Fourier transforms of the time windowed sound pressure and particle velocity, respectively, and * denotes complex conjugation [20]. Furthermore, the single-sided frequency distribution of the diffuseness estimate is given by

$$\psi(\omega) = \frac{\|\mathbf{I}_a(\omega)/c\|}{W(\omega)} = \frac{2z\,\|Re\left\{P^*(\omega)\mathbf{U}(\omega)\right\}\|}{|P(\omega)|^2 + z^2\,|\mathbf{U}(\omega)|^2}, \tag{5}$$

where $|\cdot|$ denotes the absolute value of a complex number.

The intensity and diffuseness estimates can be derived from the B-format output signals W, X, Y, and Z of an ideal SoundField microphone system as follows. The ideal omnidirectional signal W is proportional to the sound pressure p at the measurement position. Since we are not interested in the absolute values of the sound intensity and energy density, we define

$$p = W, \tag{6}$$

disregarding the sensitivity of the microphone. Furthermore, the orthogonal figure-of-eight signals X, Y, and Z are proportional to the components of the particle velocity in the corresponding directions of a cartesian coordinate system. X, Y, and Z are normalized such that a plane wave propagating in the direction of the corresponding coordinate axis yields twice the signal power of W. For a plane wave, the pressure p and the particle velocity $\mathbf{u}$ have the relation

$$|\mathbf{u}| = \frac{p}{z}. \tag{7}$$

The particle velocity is thus

$$\mathbf{u} = \frac{1}{\sqrt{2}z}\mathbf{X}', \tag{8}$$

where

$$\mathbf{X}' = \left(X\mathbf{e}_x + Y\mathbf{e}_y + Z\mathbf{e}_z\right), \tag{9}$$

and $\mathbf{e}_x$, $\mathbf{e}_y$, and $\mathbf{e}_z$ represent unit vectors in the directions of the corresponding cartesian coordinate axes. By substituting (6) and (8) in (4) and (5) we now have the frequency distributions of the active intensity and the diffuseness estimate

$$\mathbf{I}_a(\omega) = \frac{\sqrt{2}}{z}Re\left\{W^*(\omega)\mathbf{X}'(\omega)\right\}, \tag{10}$$

$$\psi(\omega) = \frac{\sqrt{2}\,\|Re\left\{W^*(\omega)\mathbf{X}'(\omega)\right\}\|}{|W(\omega)|^2 + |\mathbf{X}'(\omega)|^2/2}. \tag{11}$$

4.2. Synthesis with a multichannel loudspeaker system

Based on the analysis data, a sound field with similar energetic properties needs to be created. In SIRR, the synthesis is based on processing the omnidirectional signal W, which is analyzed with STFT using the same time-frequency resolution as in the directional analysis. Different spatialization methods are applied to the diffuse and non-diffuse parts of a room response.

An obvious method for synthesizing the non-diffuse part of a response with a multichannel loudspeaker system is to reproduce it as sharply as possible from the correct direction, for instance, with Vector Base Amplitude Panning (VBAP) [23]. Based on the proportion of non-diffuse sound energy, the frequency components within a time window are weighted by $\sqrt{\psi(\omega)}$ and panned to the direction opposite to the frequency-dependent intensity vector $\mathbf{I}(\omega)$. This step corresponds to deriving different linear phase filtered versions of the omnidirectional signal for each loudspeaker, with the filters changing from one time window to another.

For the diffuse part of sound a different method is required. The total diffuse energy $|W(\omega)|^2[1 - \psi(\omega)]$ is distributed uniformly around the listener by reproducing frequency weighted de-

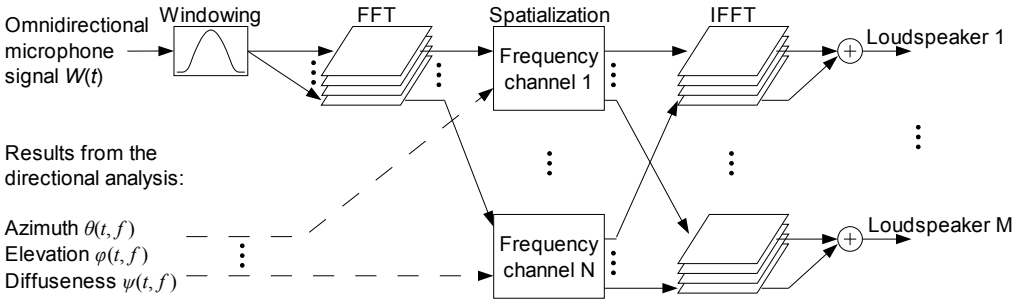

Figure 2: *Spatialization of the omnidirectional signal based on the directional analysis data.*

correlated versions of the current time window from all loudspeakers. Several methods can be used to implement the decorrelation. In earlier work, the phases were randomized by computing continuous uncorrelated noise for each loudspeaker, and by setting the magnitude spectrum of each channel in each time window equal to the magnitude spectrum of the omnidirectional microphone signal in the time window. This method can create highly decorrelated signals. However, the energy is spread over the whole analysis time window, which may produce audible pre-echo with long analysis windows. Furthermore, the frequency domain equalization of the magnitudes increases the time spreading and, if care is not taken, signal wrapping may occur even with a large amounts of zero padding. The latter problem can be alleviated by trading off the excessive time spreading to deviations in the magnitude response, which can be realized by windowing in the time domain or by smoothing the frequency response of the equalization filter. An alternative technique would be to design specific decorrelation filters, which would allow more precise control of the time spreading and the amplitude deviations [24]. We will return to this topic in future work.

4.3. Comparison of SIRR with existing techniques

Processing measured room responses with SIRR can be characterized as follows. In a large concert hall, the direct sound and early reflections are relatively sparse in time and they can usually be individually analyzed and synthesized. As non-diffuse sound, they are synthesized as point-like virtual sources using amplitude panning. The reproduction resembles coincident microphone techniques where SIRR can be thought to adaptively narrow the microphone beams in order to get the best possible directional accuracy. On the other hand, the late reverberant part of a room response is reproduced largely as decorrelated sound emanating from all loudspeakers. This is close to spaced microphone techniques and the pleasent "airiness" or "ambience" of the room should be preserved. In a smaller room the reflections are more dense, which means that fewer reflections can be individually processed and some of the directional resolution is thus lost. However, as will be seen in Sec. 5, the results are still preferred to Ambisonics.

The motivation of SIRR starts from the psychoacoustical principles discussed in Sec. 3. Interestingly, Farina and Ugolotti [25] have independently proposed an almost identical method based on theoretical considerations of sound energy analysis and the principles of Ambisonics. The difference is that Farina and Ugolotti did not divide the SoundField microphone signals into frequency bands, which has proven to be an important part of SIRR.

5. LISTENING TESTS

The perceptual quality of reproduction of room responses with SIRR has so far been evaluated in two formal listening tests [2, 3]. This Section reviews the results reported in [3].

5.1. Stimuli

Since a perfect spatial sound recording and reproduction method does not exist, it is not possible to to directly compare the perception in a real reference space to the perception of the reproduced sound in a controlled listening room. For this reason, a different approach was chosen. The evaluation was done by first creating as naturally-sounding virtual reality as possible. Recording of the virtual impulse responses with a SoundField microphone system was subsequently simulated, and the recordings were reproduced with the investigated techniques. In other words, the purpose of the test was to evaluate how close the reproduction can get to the (virtual) reference.

The virtual reference rooms were created with the DIVA software [26], which models the direct sound and early reflections with the image-source method, and late reverberation statistically. Two different room geometries were applied: a large room with a reverberation time of 1.5 s, and a class room with a reverberation time of 0.6 s. The direct sound and early reflections were applied to the nearest single loudspeakers, since using any spatialization method would have produced abnormal responses in recording the virtual environment. The same 16-channel 3-D loudspeaker system was used for reproduction of both the reference and the test samples.

Three different reproduction methods were tested: SIRR, diffusion, and Ambisonics. The loudness of each system in the reference listening position was equalized by monitoring the samples by ear. The SIRR method was implemented with 2.5 ms Hann windows with 2.5 ms zero padding. The reproduction of the diffuse time-frequency components utilized random noise equalized to have approximately the same magnitude spectrum as the omnidirectional signal in each time window, as described in Sec. 4.2 (for details see [3]). Note that the diffuseness estimate was not based on the expression of energy density for general sound fields and could have thus slightly lowered the quality of the reproduction.

The diffusion method was equivalent to SIRR, where the diffuseness indicator was set to a constant value of 0. Thus, the whole

response was always reproduced with all loudspeakers emitting decorrelated signals.

The Ambisonics decoding was realized using hypercardioid directivity as proposed in [27]. However, the reproduction was performed with only four loudspeakers in a standard quadraphonic setup, since in the previous study [2] the quality of Ambisonics was found inferior with 16 loudspeakers, and informal tests revealed that decreasing the number of loudspeakers to 4 gave considerably better results.

Once the reference and reproduced impulse responses were created, they were convolved with two different anechoic source signals to form the actual stimuli: a drum sample with four snare drum shots, and a male talker pronouncing the words "in language". It was assumed that the drum shots would reveal more differences in spatial perception, whereas the speech sample would be more sensitive to coloration due to the systems.

5.2. Procedure

The listening test method was a paired comparison with a hidden reference using the ITU impairment scale: 5.0 = imperceptible, 4.0 = perceptible but not annoying, 3.0 = slightly annoying, 2.0 = annoying, and 1.0 = very annoying [28]. Sixteen listeners, none of whom reported any hearing deficiencies, participated in the test. The listeners were instructed to pay attention to three aspects of the reproduction: sense of space, localization, and coloration, and to give a single overall rating for the difference between the reference and test samples.

Before the test, the subjects were allowed to listen to the samples for five minutes. The actual test was divided into three sessions, each consisting of two runs. Two different listening positions were utilized: the reference position ("sweet spot"), and a worst case position displaced approximately one meter from the reference position. The listening position was changed between the two runs in a session such that half of the subjects always started in the reference position and the other half in the worst case position. During each run, each of the 16 sample pairs was rated twice in a randomized order.

5.3. Results

The data from the two last sessions were taken to analysis, yielding four repetitions per each sample pair. The mean and variance of each listener in each session were normalized as recommended in [28]. The mean values and 95% confidence intervals were calculated over repetitions and over all subjects for the 32 different cases, resulting in mean opinion scores (MOS) of the listeners. The results are shown in Fig. 3. With both virtual acoustical environments, the listeners gave highest ratings for the reference-reference pair, although the values for the drum samples deviate considerably from 5.0.

Out of the reproduction methods, SIRR was always rated best. In the large room, the results for SIRR are indeed very promising. In the reference listening position, the speech samples were rated almost as high as the reference, indicating nearly transparent reproduction. With the drum sample the MOS is 4.0, which can also be considered a good result. According to the listeners' comments, there was a slight change in the pitch of the drum. One listener also reported that there were some artifacts in the reverberation tail. With the class room, the ratings for SIRR decrease on average by 0.3. This result was expected, since in the smaller room

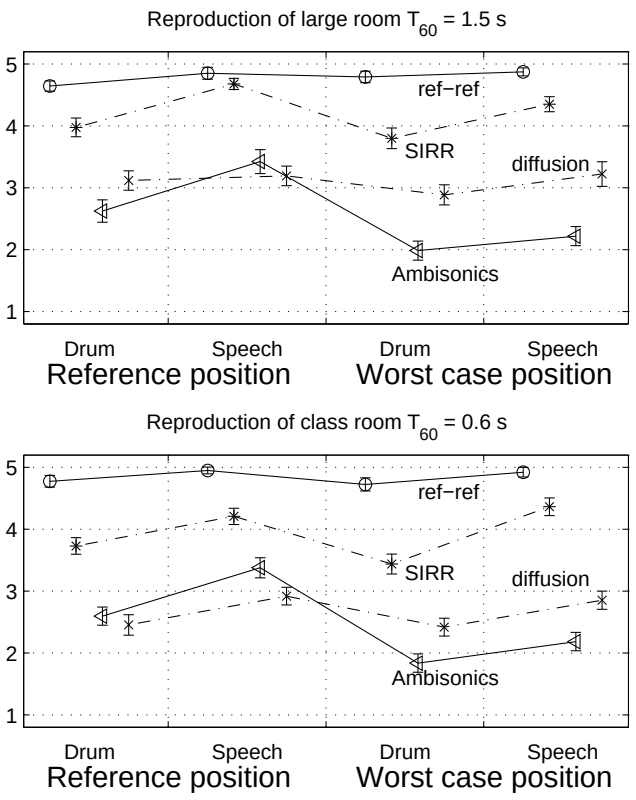

Figure 3: *MOS and 95% confidence intervals for the investigated spatial room impulse response reproduction methods for different stimuli and listening positions.*

the reflections arrive closer to each other. It is thus likely that the sound is more often interpreted as diffuse, although it consists of many discrete reflections. When moving from the reference listening position to the worst case position, the MOS values are reduced on average only by 0.2.

The results for Ambisonics are almost identical in the two virtual rooms. In the reference listening position, the MOS values are 2.6 for the drum sample, and 3.4 for the speech sample. However, in the worst case position, the MOS values are close to 2.0 with both sound samples, which means that the listeners have perceived the samples annoyingly different from the reference. According to their comments, the sound was localized to the nearmost loudspeaker, which completely changed the perception of directions and the envelopment created by the virtual room.

In the large room, the MOS of the diffusion method has a nearly constant value of approximately 3.0 for both stimuli and listening positions. In the class room, the MOS drops with the drum sample to 2.5. The listeners reported that with the diffusion method the sound was not colored and that the room size remained the same as in the reference. However, in the reference listening position, the localization of the sound sources was lost, and in the worst case position the sources were localized mainly to the nearmost loudspeaker. Neverhteless, the envelopment by the virtual room did somewhat remain.

6. SUMMARY

The Spatial Impulse Response Rendering (SIRR) method for reproduction of room acoustics was described. SIRR is motivated by psychoacoustics, and it utilizes energy analysis of sound fields to obtain the necessary data to synthesize room responses suitable for reproduction with arbitrary surround loudspeaker systems. More specifically, the direction of arrival and diffuseness of the sound field are analyzed within frequency bands. The discussed synthesis method spatializes an omnidirectional response using amplitude panning and a decorrelation technique. The reviewed listening test data indicate that the perceptual quality of SIRR is superior compared to Ambisonics and the tested diffusion method, providing at best almost transparent reproduction of the spatial impression of a measured room or hall.

7. ACKNOWLEDGMENTS

The authors would like to thank Angelo Farina for pointing out the correct form for energy density of general sound fields and Tapio Lokki for his contribution in the evaluation of SIRR. The work of Juha Merimaa has been supported by the research training network for Hearing Organisation And Recognition of Speech in Europe (HOARSE, HPRN-CP-2002-00276). Ville Pulkki has received funding from the Academy of Finland (101339).

8. REFERENCES

[1] J. Merimaa and V. Pulkki, "Perceptually-based processing of directional room responses for multichannel loudspeaker reproduction," in *IEEE Workshop on Appl. of Sig. Proc. to Audio and Acoust.*, New Paltz, NY, USA, 2003, pp. 51–54.

[2] V. Pulkki, J. Merimaa, and T. Lokki, "Multi-channel reproduction of measured room responses," in *International Congress on Acoustics*, Kyoto, Japan, 2004, pp. II 1273–1276.

[3] V. Pulkki, J. Merimaa, and T. Lokki, "Reproduction of reverberation with spatial impulse response rendering," in *AES 116th Convention*, Berlin, Germany, 2004, Preprint 6057.

[4] S. P. Lipshitz, "Stereo microphone techniques... Are the purists wrong?," *J. Audio Eng. Soc.*, vol. 34, no. 9, pp. 716–744, 1986.

[5] V. Pulkki, "Microphone techniques and directional quality of sound reproduction," in *AES 112th Convention*, Munich, Germany, 2002, Preprint 5500.

[6] M. A. Gerzon, "Periphony: With-height sound reproduction," *J. Audio Eng. Soc.*, vol. 21, no. 1, pp. 2–10, 1973.

[7] A. Laborie, R. Bruno, and S. Montoya, "A new comprehensive approach of surround sound recording," in *AES 114th Convention*, Amsterdam, The Netherlands, 2003, Preprint 5717.

[8] A. Laborie, R. Bruno, and S. Montoya, "High spatial resolution multichannel recording," in *AES 116th Convention*, Berlin, Germany, 2004, Preprint 6116.

[9] J. Blauert, *Spatial Hearing*, The MIT Press, Cambridge, MA, USA, revised edition, 1997.

[10] S. E. Boehnke, S. E. Hall, and T. Marquadt, "Detection of static and dynamic changes in interaural correlation," *J. Acoust. Soc. Am.*, vol. 112, no. 4, pp. 1617–1626, 2002.

[11] C. Faller and J. Merimaa, "Source localization in complex listening situations: Selection of binaural cues based on interaural coherence," *J. Acoust. Soc. Am.*, 2004, Accepted for publication.

[12] B. C. J. Moore, *An Introduction to the Psychology of Hearing*, Academic Press, London, UK, 4th edition, 1997.

[13] A. Kohlrausch, "Auditory filter shape derived from binaural masking experiments," *J. Acoust. Soc. Am.*, vol. 84, no. 2, pp. 573–583, 1988.

[14] M. van der Hejden and C. Trahiotis, "Binaural detection as a function of interaural correlation and bandwidth of masking noise: Implications for estimates of spectral resolution," *J. Acoust. Soc. Am.*, vol. 103, no. 3, pp. 1609–1614, 1998.

[15] I. Holube, M. Kinkel, and B. Kollmeier, "Binaural and monaural auditory filter bandwidths and time constants in probe tone detection experiments," *J. Acoust. Soc. Am.*, vol. 104, no. 4, pp. 2412–2425, 1998.

[16] B. R. Glasberg and B. C. J. Moore, "Derivation of auditory filter shapes from notched-noise data," *Hear. Res.*, vol. 47, pp. 103–138, 1990.

[17] J. Blauert, "On the lag of lateralization caused by interaural time and intensity differences," *Audiology*, vol. 11, pp. 265–270, 1972.

[18] D. W. Grantham and F. L. Wightman, "Detectability of varying interaural temporal differences," *J. Acoust. Soc. Am.*, vol. 63, no. 2, pp. 511–523, 1978.

[19] F. Baumgarte and C. Faller, "Binaural Cue Coding. Part I: Psychoacoustic fundamentals and design principles," *IEEE Trans. Speech Audio Proc.*, vol. 11, no. 6, pp. 509–519, 2003.

[20] F. J. Fahy, *Sound Intensity*, Elsevier Science Publishers Ltd., Essex, England, 1989.

[21] D. Stanzial, N. Prodi, and G. Schiffrer, "Reactive acoustic intensity for general fields and energy polarization," *J. Acoust. Soc. Am.*, vol. 99, no. 4, pp. 1868–1876, 1996.

[22] G. Schiffrer and D. Stanzial, "Energetic properties of acoustic fields," *J. Acoust. Soc. Am.*, vol. 96, no. 4, pp. 3645–3653, 1994.

[23] V. Pulkki, "Virtual sound source positioning using vector base amplitude panning," *J. Audio Eng. Soc.*, vol. 45, no. 6, pp. 456–466, 1997.

[24] M. O. J. Hawksford and N. Harris, "Diffuse signal processing and acoustic source characterization for applications in synthetic loudspeaker arrays," in *AES 112th Convention*, Munich, Germany, 2002, Preprint 5612.

[25] A. Farina and E. Ugolotti, "Subjective comparison between stereo dipole and 3D ambisonic surround systems for automotive applications," in *Proc. AES 16th International Conference*, Rovaniemi, Finland, 1999, pp. 532–543.

[26] L. Savioja, J. Huopaniemi, T. Lokki, and R. Väänänen, "Creating interactive virtual acoustic environments," *J. Audio Eng. Soc.*, vol. 47, no. 9, pp. 675–705, 1999.

[27] G. Monro, "In-phase corrections for Ambisonics," in *Proc. Int. Computer Music Conf.*, Berlin, Germany, 2000, pp. 292–295.

[28] ITU-R, "Recommendation BS.1284-1, General methods for the subjective assessment of sound quality," International Telecommunication Union Radiocomminication Assembly, 2003.