DECORRELATION TECHNIQUES FOR THE RENDERING OF APPARENT SOUND SOURCE WIDTH IN 3D AUDIO DISPLAYS

Guillaume Potard, Ian Burnett

School of Electrical, Computer and Telecommunications Engineering
University of Wollongong, Wollongong, Australia
gp03@uow.edu.au

ABSTRACT

The aim of this paper is to give an overview of the techniques and principles for rendering the apparent source extent of sound sources in 3D audio displays. We mainly focus on techniques that use decorrelation as a mean to decrease the Interaural Cross-Correlation Coefficient (IACC) which has a direct impact on the perceived source extent. We then present techniques where decorrelation is varied in time and frequency, allowing to create temporal and spectral variations in the spatial extent of sound sources. Frequency dependant decorrelation can be employed to create an effect where a sound is spatially split in its different frequency bands, these having different positions and spatial extents. We finally present results of psychoacoustic experiments aimed at evaluating the effectiveness of decorrelation based techniques for the rendering of sound source extent. We found that the intended source extent matches well the mean perceived source extent by subjects.

1. INTRODUCTION

The spatial extent of a sound source is defined as the perceived spatial dimension or size of the sound source. The spatial extent of sound sources is a natural phenomenon (ie a beach front, wind blowing in trees, waterfall etc) and is an important perceptual cue of sound sources [1]. Therefore to improve the realism and immersion of 3D audio displays, it is important to also render spatial source extent. In MPEG-4 AudioBIFS [2][3] recent amendments have been proposed to allow the definition of spatial dimensions of sound sources [4].

Apparent source extent is tightly linked to the Inter Aural Cross Correlation coefficient (IACC) [1]. In concert halls for instance, a low IACC value improves the feeling of spaciousness and source width [5]. Various stereo techniques based on signal decorrelation have also been used to create wider stereo images [6]. Other authors [7][8] have studied the apparent width of noise presented on two speakers and headphones. They found that the amount of correlation between the two presented channels had a dramatic impact on the perception of wideness. When uncorrelated noise signals are presented on two speakers, the noise seems to fill the complete space between the speakers while correlated signals produce a narrow sound source placed in between the speakers [7].

In these previous studies, source extent is always considered as a one dimensional attribute of the sound sources. We broadened the concept of sound source extent by studying the perception of sound source 'shapes', that is, sound sources having particular 2D or 3D extents. We found that subjects could identify several shapes above statistical probability but below 40% of the time [9]. Experiment in this paper also study the differentiation between 1D (line) and 2D (rectangular) sound sources.

We first give a psychoacoustic overview of source extent perception. We then describe signal processing techniques for rendering source extent in 3D audio displays. We finally present results of a subjective experiment investigating the difference between intended source extent and the actual perceived source extent by subjects.

2. PSYCHOACOUSTICS OF SOUND SOURCE EXTENT

Source extent has been studied in a large amount of literature (see [1], [10] and [11] for a review) under the names of apparent source width, tonal volume and others. It has been shown that the perceived source extent depends on the value of the inter-aural cross correlation coefficient (IACC) [12], sound loudness [13], pitch and signal duration [14]. The IACC coefficient is a widely used parameter in acoustics [1], [15] to determine the spaciousness and envelopment of concert halls. An IACC value close to zero will introduce a sense of diffuseness and of spatially large sound source; in contrast, an IACC absolute value close to 1 will produce a narrow sound image.

The IACC coefficient is defined as the maximum absolute value of the normalised interaural cross correlation function in turn defined as:

$$IACC(\tau) = \frac{\int_{-\infty}^{+\infty} s_L(t-\tau)s_R(\tau)dt}{\sqrt{\int_{-\infty}^{+\infty} s_L^2 dt \int_{-\infty}^{+\infty} s_R^2 dt}} \qquad (1)$$

where $s_L(t)$ and $s_R(t)$ are the ear canal signals at the left and right ears. The normalised cross-correlation function is bounded between -1 and 1.

Surprisingly, the binaural system is able to compute IACC coefficients for different frequency bands [16] and is also sensitive to temporal fluctuations [17], [1] of the IACC coefficient. The Section 3.1 describes techniques to artificially produce temporal and spectral variations of the IACC coefficient, producing particular 3D sound effects.

We now review a list of techniques to render spatial sound source extent.

3. METHODS FOR RENDERING APPARENT SOURCE WIDTH

This Section describes the techniques employed to create and control the spatial extent of sound sources.

3.1. Uncorrelated point sources

A commonly used technique to render the extent of sound sources in virtual auditory displays relies on the observation that a physically broad sound source can be decomposed into several, spatially distinct, point sound sources (Figure 1a). However, for this effect to take place, the signals emitted by the point sources must be statistically uncorrelated from one another. This is due to the fact that if correlation is high between the point sources, the binaural system perceives them as a single auditory event [1]. This results in a summation phenomenon and consequently only a narrow sound source is perceived at the center of gravity (Figure 1b). The position of the center of gravity depends on the positions and intensity gains of the point sources. This equates to amplitude panning performed between several speakers. In contrast, if the signals generated by the point sources are weakly correlated, the binaural system perceives the point sources as distinct auditory streams. This results in the perception of a spatially wide sound source (Figure 1c). In reality however, if the point sources are densely distributed, it might not be possible to distinguish every single point sources as a different stream because the binaural system produces a final impression of a single, spatially large, sound source.

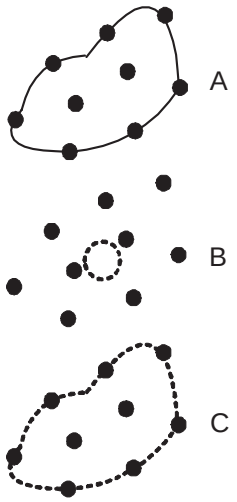

Figure 1: *a) Decomposition of a broad sound source into point sources, b) High correlation between point sources creating a narrow sound image, c) Low correlation creating a wide sound image.*

We now review different techniques to obtain decorrelated signals from a monaural sound signal.

3.1.1. Full band Decorrelation

The simplest way to obtain decorrelated signals is to introduce a small time delay between them. Although simple, this method can only produce a limited number of decorrelated signals as the upper permissible delay is restricted by the perception of an echo; this is typically around 40 ms. On speakers, this technique should however be avoided due to the possible comb-filtering effects caused by delays.

Decorrelation is most commonly achieved by filtering the input signal with all-pass filters having random, noise-like, phase responses [7]. Due to the ear instability to phase variations and the preservation of the signal amplitude spectrum (i.e. all-pass response), the obtained output signals are perceptually equal but statistically orthogonal.

Decorrelating all-pass filters can be implemented in FIR, IIR [7] or Feedback Delay Network (FDN) architectures.

This technique can be used to create only a finite and relatively small number of uncorrelated signals, as a high correlation value will eventually occur between a pair of signals, due to the finite length of the filters. Thus the filter phase responses also need to be maximally orthogonal and need to be obtained by a best performance selection process. With this technique, we were able to obtain only five to six totally decorrelated signals. The filter length used was typically 100 poles and zero.

In order to obtain further signals, time-varying or dynamic decorrelation [7] is introduced.

3.1.2. Dynamic Decorrelation

Time-varying or dynamic decorrelation can be defined by the use of time-varying all-pass filters. The advantage of dynamic decorrelation over fixed decorrelation is that a higher number of uncorrelated signals can be obtained. This is due to the fact that time-varying decorrelation will introduce time-varying levels of decorrelation, depending on the orthogonality of the filter phase responses, but if these variations are fast enough and cannot be tracked by the ear, the perceived mean correlation value is low.

With all-pass filters, dynamic decorrelation is obtained by calculating a new random phase response for every new time frame. FIR or IIR lattice filter structures are best suited for this task due to their resistance to the instabilities that can occur during frequent filter coefficient updates.

Dynamic decorrelation also generate special audible effects not obtained with fixed decorrelation: it has been said [7] that dynamic decorrelation creates micro-variations simulating the time-varying fluctuations caused by moving air.

However we found that dynamic decorrelation can also have a distracting effect and even creates fatigue due to noticeable changing positions of objects in a recorded scene. This is likely due to phase differences between point sources that produce an Interaural Time Difference (ITD). Therefore, it is left to the discretion of the sonification designer whether fix or dynamic decorrelation should be used.

3.1.3. Sub-band Decorrelation

So far we have only looked at decorrelation that is applied to the full signal spectrum. We now introduce a novel technique that allows us to alter decorrelation differently in each frequency band. Using this technique, a set of signals can be obtained where, for instance, their low-frequency components are uncorrelated while their high frequency components are left correlated. Using the point source method described above, this can lead to interesting effects where the spatial extent of a sound source varies in frequency . Therefore a sound source can be split into frequency bands having different spatial extents and positions. We call this effect the spatial Fourier decomposition effect. This effect can easily be noticed after some training.

The sub-band decorrelation technique is depicted in Figure 2. The input is first split into different frequency bands by a decomposition filterbank made of high order low-pass, band-pass and high-pass filters. Each sub-band signal is then decorrelated using any of the decorrelation technique described above. Crossfader modules are then used to control the amount of correlation

in each frequency band by a decorrelation factor k. This works by re-injecting some common sub-band signal into each decorrelated signal. For example, if total decorrelation is wanted, k equal zero, then no common signal is injected. If k equals one, the cross-fader outputs only the common signal and no decorrelated signal, therefore the correlation coefficient is one. It is also possible to set k to any intermediate correlation value. A constant power cross-fading technique is preferable so that no change in signal level can be observed when k is changed. Finally the different sub-bands of the respective decorrelated signals are added together to form the final set of partially decorrelated signals.

We have implemented such a decorrelator on the MAX/MSP platform [18] with low (0–1 kHz), medium (1–4 kHz) and high (4 kHz–20kHz) sub-bands. A higher number of sub-bands could be employed in order to obtain a finer grain on the correlation spectrum.

Finally, we note that it is also possible to combine dynamic decorrelation and sub-band decorrelation to obtain time and frequency varying levels of signal correlation.

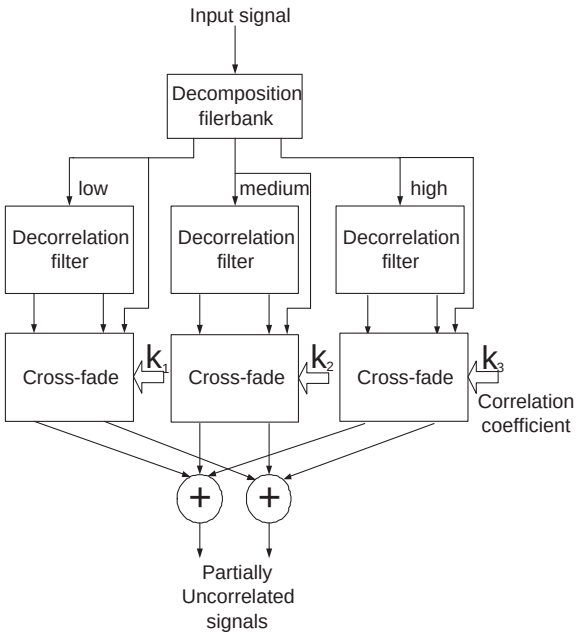

Figure 2: *A three sub-band decorrelator.*

3.1.4. Time-varying Decorrelation

Time-varying decorrelation is obtained by periodically re-injecting the original signal into the decorrelated signals. When decorrelation is changed at up to 10 Hz with correlation coefficients between sound sources varying between 0 and 1, this creates a sound source with a constantly varying spatial extent. Above 10Hz this effect is destroyed by the lag of the binaural system to derive an IACC coefficient.

3.1.5. Other decorrelation techniques

Other decorrelation techniques besides delay and all-pass filtering exist (see [19] for an overview), these are often used in echo-cancellation systems. However, we discarded these techniques for virtual auditory display applications because they either degrade the signal (artificial introduction of noise and distortion), create large source localisation shifts and a disturbing phasing effect (Hilbert transform based techniques), they destroy the signal (KLT transform) or do not generate a high enough number of decorrelated signals.

3.2. O-format

A different approach for reproducing the spatial extent of sound sources relies on the encoding of the sound source spatial dimensions and directivity into spherical harmonics impulse responses, these techniques known as O-format and W-panning [20], [21] are offspring of Ambisonics theory [22]. We have not yet experimented with these techniques but it seems that low IACC at the listeners ears could also be achieved if the convolution of the the monaural source signal with the spherical harmonics impulse responses creates enough decorrelation between parts of the broad sound source.

4. EVALUATION

The aim of this experiment was to study the shift between the intended source extent (ie the extent wished by the sound engineer or artist) and the actually perceived source extent by subjects.

The experiment was performed on the Configurable Hemispheric Environment for Spatialised sound (CHESS) [23] which uses fourth order Ambisonics spatialisation on a 16 speaker dome array. The space is not anechoic but has some acoustic proofing.

4.1. Stimuli

To create sound sources with various spatial extents, we employed six point sources and the technique described in Section 3. The point sources were spatialised using Ambisonics spatialisation and fed with independent white noise sequences having inter correlation coefficients of 0.

We constructed 49 sound sources having various spatial extents, locations and geometry (Figure 3). Firstly, horizontal lines were made with a spatial extent of 60 and 180 degrees (sequences 1-4 and 11-14 respectively). We then constructed vertical lines with 40 and 90 degree extents (sequences 5-8). We also created small and big square sound sources having spatial extents of 60 degrees horizontally and 30 degrees vertically (sequence 10) and 180 degrees horizontally and 40 degrees vertically (sequence 9).

Finally we investigated the perceived spatial extent of a single speaker (sequence 15-16).

4.2. Procedure

Subjects were asked to draw the spatial extent of the noise sequences they were listening to on an answer sheet that represented a top-down view of the dome speaker array. On the answer sheet, the center therefore represents the zenith of the dome. Subjects were placed at the center of the dome and facing the zero degree orientation. Head rotations were allowed.

Although not perfect and subject to transcription errors, this elicitation method seemed appropriate for the transcription of the sound source extents perceived by subjects.

Fifteen subjects with no particular experience or knowledge in the audio field participated in the experiment.

4.3. Results and discussion

Areas where subjects had drawn were counted, and from this, density graphs generated. Due to limited space, we only show sixteen sequences out of the obtained 49 (Figure 3).

The graphs show that, in general, the mean perceived source extent follow the intended sound source extent (thick line).

For sources with an horizontal extent of 60 degrees (sequences 1-4), the perceived source extent was narrower than intended. This is probably due to the source density being too high; this creates a narrower source extent. This effect has been observed in previous experiments that we have carried out [9][4]. For sources with an horizontal extent of 180 degrees (11-13), the perceived source extent matched the intended extent, however subjects perceived some elevation in the sound which was not actually present. Sequence 14, which is an horizontal sound source placed at 40 degrees elevation was perceived as being higher, but not with a great precision however. Sources with a vertical extent (sequences 5-8) can be seen as having been discriminated from the horizontal sources. The sources with a square extent (9-10) where perceived roughly like the horizontal sources, but with slightly more vertical extent. In general, we can also notice that the ability to assess source extent is diminished for sounds coming from behind. Finally we can see in sequences 15 and 16 that even a single speaker is not perceived as a point source and has some spatial extent.

In general we can conclude that localisation of the the wide sound sources were correct and that the mean perceived spatial source extent matches coarsely the intended extent. It can be seen however that subjects can discriminate horizontal from vertical sources with different extent.

The mean perceived source extent matches coarsely the intended extent but there can be a lot of variation on the perception of extent (or the elicitation error) between subjects.

Also, using spatially small speakers would also help in the sharpening of source extent.

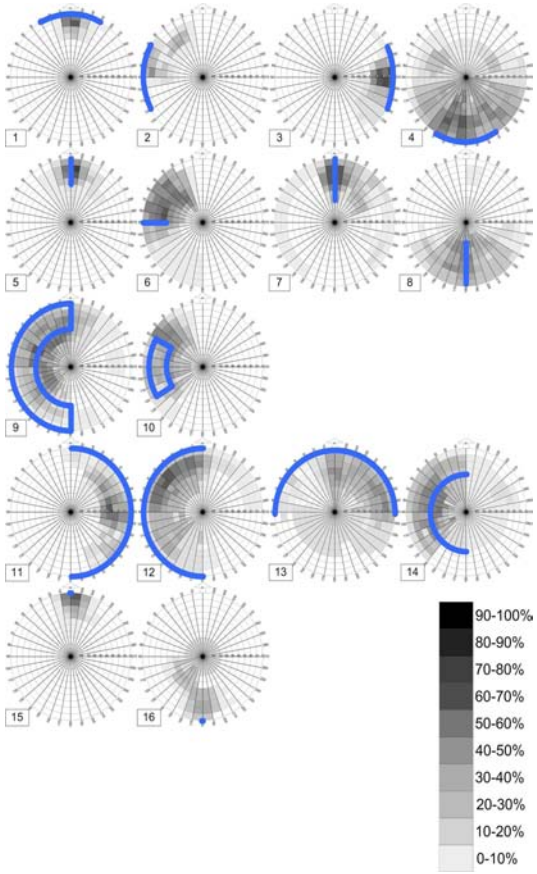

Figure 3: *Density plots of sound source extent perception.*

5. CONCLUSIONS

We gave an overview of the psychoacoustics concepts that control the perception of sound source extent. We then reviewed existing techniques to render the spatial extent of sound sources. We have then introduced a technique to alter the level of correlation of signals in different frequency bands. The resulting effect is that of a spatial Fourier decomposition where the different frequency bands of the signals are perceived in different positions with different spatial extents. This effect was clearly perceivable by the author but requires a substantial amount of training. We are planning to carry out experiments in order to assess further this effect on subjects.

In conclusion, we have seen that the apparent extent of sound sources can be controlled on several dimensions: spatial, temporal and spectral. We have, so far, only studied the spatial case.

Artificially produced spatial source extent based on temporal and spectral dependant decorrelation can be used to create special 3D audio effects that need to be further investigated and tested on subjects.

6. REFERENCES

[1] J. Blauert, *Spatial Hearing*, MIT Press, Revised edition, 1996.

[2] ISO/IEC standard, 14496-1 Information technology - Coding of audio-visual objects–Part 1: Systems, 2000.

[3] E. D. Scheirer, "AudioBIFS: Describing audio scenes with the MPEG-4 multimedia standard," *IEEE Trans. on Multimedia 1*, no. 3, pp. 237–250, 1999.

[4] G. Potard, J. Spille, "Study of Sound Source Shape and Wideness in Virtual and Real Auditory Displays," *presented at the 114th AES Convention*, Amsterdam, March 2003, preprint 5766.

[5] L. Beranek, "Concert and opera halls: How they sound," *J. Acoust. Soc. Am.*, 1996

[6] M. Gerzon, "Signal processing for simulating realistic stereo images," *93rd AES convention*, New York, 1-4 October, Preprint 3424, 1992.

[7] G. S. Kendall, "The Decorrelation of Audio Signals and Its Impact on Spatial Imagery," *Computer Music J.*, vol. 19, no. 4, pp. 71–87, 1995

[8] K. Kurozumi, K. Ohgushi, "The relationship between the cross-correlation coefficient of two-channel acoustic signals and sound image quality," *J. Acoust. Soc. Am.*, vol. 74, no. 6, pp. 1726–1733, Dec. 1983.

[9] G. Potard, I. Burnett, "A study on sound source apparent shape and wideness," *Proc. of Int. Conference on Auditory Displays (ICAD)*, Boston, USA, 6-9 July 2003, pp. 25–28.

[10] H. Lehnert, "Auditory Spatial Impression," in *Proc. of the 12th Audio Eng. Soc. Conf.*, Copenhagen, Denmark, June 1993.

[11] D. R. Begault, *3-D sound for virtual reality and multimedia*, Academic Press Professional, San Diego, USA, 1994.

[12] M. Morimoto, "The Relation between Spatial Impression and The Precedence Effect," in *Proc. of the ICAD Conf.*, Kyoto, Japan, July 2002.

[13] E. G. Boring, "Auditory theory with special reference to intensity, volume, and localization," *J. Acoust. Soc. Am.*, vol. 37, no. 2, pp. 157–188, 1926.

[14] D. R. Perrott and T. N. Buell, "Judgments of sound volume: effects of signal duration, level, and interaural characteristics on the perceived extensity of boradband noise," *J. Acoust. Soc. Am.*, vol. 72, no. 5, pp. 1413–1417, Nov. 1982.

[15] M. Tohyama, H. Susuki, Y. Ando, *The nature and technology of acoustic space*, Academic Press, 1995.

[16] J. M. Potter, F. A. Bilsen and J. Raatgever, "Frequency dependence of spaciousness," *Acta Acoustica*, vol. 3, pp. 417–427, Oct. 1995.

[17] R. Mason, *Elicitation and measuremnt of auditory spatial attibutes in reproduced sound*, Ph.D. Thesis, University of Surrey, Feb. 2002.

[18] `www.cycling74.com`

[19] Y. W. Liu, J. O. Smith III, "Perceptually similar orthogonal sounds and applications to multichannel acoustic echo cancelling," in *Proc. of the 22nd Audio Eng. Soc. Conf.*, Espoo, Finland, 2002.

[20] D. G. Malham, "Spherical harmonic coding of sound objects - the ambisonic 'O' format," in *Proc. of the 19th Audio Eng. Soc. Conf.*, Schloss Elmau, Germany, 1999.

[21] D. Menzies, "W-panning and O-format, tools for spatialisation," in *Proc. of ICAD Conf.*, Kyoto, Japan, July 2002.

[22] D. G. Malham, "3-D soound spatialization using Ambisonics techniques," *Computer Music J.*, vol. 19, no. 4, pp. 58–70, Winter 1995.

[23] CHESS Webpage, `http://www.whisper.elec.uow.edu.au/CHESS/`