CONTROL AND MEASUREMENT OF APPARENT SOUND SOURCE WIDTH AND ITS APPLICATIONS TO SONIFICATION AND VIRTUAL AUDITORY DISPLAYS

Guillaume Potard, Ian Burnett

University of Wollongong
School of Electrical, Computer and Telecommunications Engineering
Faculty of Informatics, Wollongong, NSW, Australia
gp03@uow.edu.au, i.burnett@elec.uow.edu.au

ABSTRACT

The aim of this paper is to investigate the possibility of using the spatial extent of sound sources as a mean of carrying information in sonification designs. To do so, we studied the accuracy of the perception of artificially produced sound source extent in a 3D audio environment. We found that the source extent perceived by subjects matched relatively well the intended source extent. Thus source extent could be used as a tool to represent areas, sizes and regions in virtual auditory displays. This paper also reviews the technologies involved in the reproduction and measurement of spatially extended sound sources. Finally, it is shown that the perception of sound source extent can be sensitive to temporal and spectral variations thereby adding extra sonification parameters.

1. INTRODUCTION

The spatial extent of sound sources is a natural and important phenomenon in the perception of sound sources [1]. The spatial extent of a sound source can be defined as the spatial dimension or 'size' of the sound source. For instance, a beach front or an highway are typically perceived as wide horizontal sound sources, while a flying insect is perceived as a tiny point source.

In a sonification perspective, controlled sound source extent could be used to represent areas and regions of activity provided that it is controllable, predictable and that it does not affect sound localisation too much.

In an air-traffic controller sonification application, source extent could be used, for instance, to represent the size and distance of surrounding planes.

Source extent has been studied in a large amount of literature (see [1], [2] and [3] 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) [4], sound loudness [5], pitch and signal duration [6]. The IACC coefficient is a widely used parameter in acoustics [1], [7] 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.

Surprisingly, the binaural system is able to compute IACC coefficients for different frequency bands [8]: this leads to interesting sonification applications where the sound source extend is varying with the signal frequency.

The binaural system is also sensitive to temporal fluctuations [9], [1] of the IACC coefficient. This also can be used as a sonification parameter.

We first review the definition of the inter-aural cross-correlation function and its derivatives used to visualise time and frequency interaural correlation. We then present several techniques that can are employed to render and control the extent of sound sources. Follows a review of decorrelation techniques commonly employed in the reproduction of sound source extent. Finally we present experimental results showing the accuracy of perceived source extent using the source extent rendering methods described below.

2. MEASUREMENTS OF THE INTER AURAL CROSS-CORRELATION FUNCTION

Firstly, the IACC coefficient is reviewed. We then overview other IACC measurements that can be used to study time varying and frequency varying correlation. We finally introduce a technique called the coherence spectrogram which can be used to visualise simultaneously the frequency dependence and time variations of the cross-correlation function.

2.1. Fix and temporal measurements of correlation

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

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

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.

Although useful for finite length signal sequences, the IACC coefficient gives only a coarse, averaged, representation of the cross-correlation function. A problem with the IACC coefficient is that it does not account for temporal variations in the level of inter-aural correlation. Some research [9], [1] has shown that temporal fluctuations of the IACC coefficient can also responsible for the sensation of spaciousness.

To measure the temporal variations of the IACC coefficient, it is possible to compute IACC coefficients per every time frame. A somewhat more accurate technique relies on a model of the binaural system which computes a running correlation function [10], [1].

The running inter-aural correlation function performed by the binaural system can be modeled as follows:

$$IACC(t,\alpha) = \int_{-\infty}^{+\infty} s_L(\alpha) s_R(\alpha-\tau) G(t-\alpha) d\alpha \quad (2)$$

where $G(t-\alpha)$ defines a decaying exponential function defined as:

$$G(x) = e^{-\frac{x}{\tau}} \text{ for } x \geqslant 0 \text{ and } G(x) = 0 \text{ for } x < 0$$

so that, after some time constant τ, past samples are not taken into the calculation of the correlation. It is advised [1] to use few milliseconds for this time constant. The running cross-correlation function allows us to plot a cross-correlation function at any time t, and thus by taking the maximum value we can obtain an IACC value for this particular time.

Methods for measuring changes of correlation over time are useful to test sonification designs that rely on time varying decorrelation.

We now consider frequency varying correlation.

2.2. Sub-band and coherence spectrum measurements

A common way to study the frequency dependence of the inter-aural correlation function is to derive IACC coefficients in 1/3 octave bands [4].

Another technique is to calculate a continuous coherence spectrum [11], [12] defined as follow:

$$IACC(f) = \frac{S_{S_r S_l}(f)}{\sqrt{S_{S_l S_l}(f) S_{S_r S_r}(f)}} \quad (3)$$

where $S_{xy}(f)$ is the cross-power spectral density function defined as:

$$S_{xy}(f) = \lim_{T \to 0} E[\frac{X(f)Y(f)}{T}] \quad (4)$$

where E is the ensemble average function.

The obtained coherence spectrum function is thus the normalised fourier transform of the inter-aural cross-correlation function.

The magnitude spectrum of coherence is therefore bounded between 0 and 1 and represents the correlation coefficient versus frequency function. Fig. 1 depicts an example of the coherence spectrum between two signals. It can be seen that these two signals have a 0.45 correlation coefficient for a certain frequency band but are uncorrelated otherwise. A simple IACC measurement would not be able to reflect this fact.

Finally, since the IACC coefficients can vary in time and frequency, we propose the use of a coherence spectrogram representation (Fig. 2). This consists in computing the coherence spectrum on a time frame.

Fig. 2 represents the coherence spectrogram of two signals which have a common frequency band that is periodically correlated and uncorrelated. These two signals were obtained using a sub-band decorrelation technique described in section 4.3.

2.3. Summary

The correlation measurements described above have been summarised in Fig. 3. The simplest measurement of correlation is given by the IACC coefficient (Fig. 3 top left). One can also derive a running cross-correlation function that gives IACC coefficients in function of time (Fig. 3 top right). To study coherence in the frequency domain, one can derive IACC coefficients in 1/3 octave bands (Fig. 3 bottom left) or derive the inter-aural coherence spectrum (Fig. 3 bottom middle) which gives a continuous representation of the correlation spectrum over frequency. Finally the coherence spectrogram (Fig. 3 bottom right) is used to display the IACC coefficients both versus time and frequency.

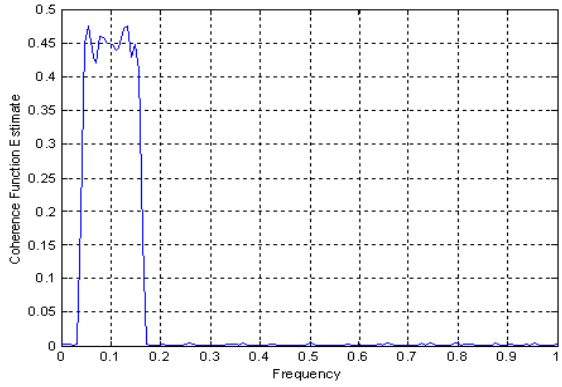

Figure 1: *Coherence spectrum*

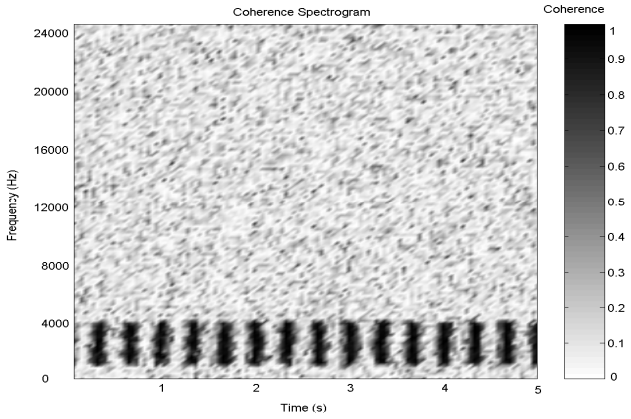

Figure 2: *Coherence spectrogram*

We now review the techniques used to render and control the extent of sound sources.

3. PRINCIPLE OF SOUND SOURCE EXTENT REPRODUCTION

This section describes the techniques employed in auditory displays to create and control the spatial extent of sound sources.

3.1. Source width reproduction principle

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 (Fig. 4a) [13]. 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

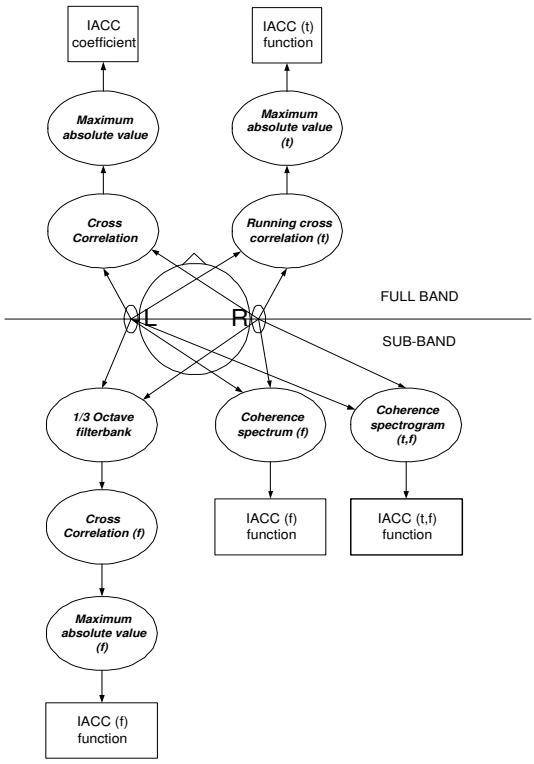

Figure 3: *Summary of correlation measurements.*

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 (Fig. 4b). The position of the center of gravity depends on the positions and intensity gains of the point sources. This can be equated 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 (Fig. 4c). In reality however, if the point sources are densely distributed, it might not be possible to distinguish every single point sources as a different auditory stream because the binaural system produces a final impression of a single, spatially large, sound source.

3.2. Link to the inter-aural cross-correlation function

Another corroborating explanation for the spatial width of weakly point sources is that the Inter Aural Cross Correlation coefficient (IACC) coefficient is decreased. A widely spread literature links a low IACC coefficient with the perception of a large and diffuse source extent [14] and a feeling of spaciousness and envelopment in concert halls [15].

To study the link between correlation of the point sources and the inter-aural correlation, we note that, in anechoic conditions, the signals arriving at the listeners left and right ears are the sums of the source signals convolved with the Head Related Transfer functions (HRTF) for the left and right ears respectively (Fig. 5):

$$L(t) = \sum_{k=1}^{N} H_{L_k} * s_k(t) \tag{5}$$

$$R(t) = \sum_{k=1}^{N} H_{R_k} * s_k(t) \tag{6}$$

where $s_k(t)$ are the signals generated by N point sources and H_{L_k} and H_{R_k} are the HRTF functions for the left and right ears that are in turn dependent on the point source positions relative to the listener.

It can be seen that if the $s_k(t)$ signals are highly correlated or identical, the IACC value will only depend on the decorrelation caused by the HRTF functions; if the signals generated by the point sources have same times of arrival (equidistant sources from the listener), this decorrelation is very weak. The obtained IACC value is high and a narrow sound image is perceived. On the other hand if the $s_k(t)$ signals are totally incoherent, the IACC value decreases, but will not reach zero due to coherence re-introduced by the HRTF functions. It should also be noted that in echoic conditions, room reverberation tend to reduce the IACC [7].

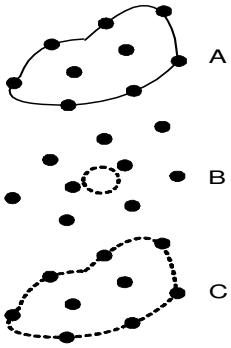

Figure 4: *a) Decomposition of a broad sound source into point sources, b) High correlation between point sources creating a narrow sound image, c) Low correlation creating a wide sound image.*

As we have seen, a large source extent is achieve with a low IACC value which in turn relies on a low level of correlation between point sources.

We now briefly look at other methods to create source extent.

3.3. Other rendering techniques

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 [17], [18] are offspring of Ambisonics theory [19]. 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.

Finally, other approaches used in cinema are more focussed on creating diffuse sound fields rather than a particular source extent; decorrelation and artificial reverberation are commonly used to achieve diffusion.

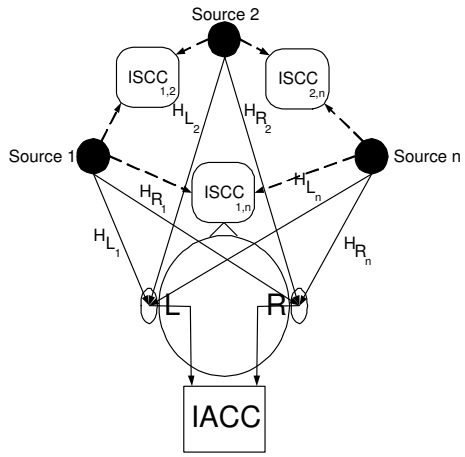

Figure 5: *Relation between the Inter Source Cross Correlation (ISCC) values and the Inter Aural Cross Correlation (IACC) value*

We now review several decorrelation techniques required in the implementation of the uncorrelated point source technique.

4. DECORRELATION TECHNIQUES

We now discuss techniques and challenges involved in obtaining a set of uncorrelated signals from a monaural source. We first look at time-invariant decorrelation, then at dynamic decorrelation. Finally we present a novel sub-band decorrelation technique that allows to alter the correlation level for each frequency band independently.

4.1. Time invariant decorrelation

4.1.1. Time delay

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 effect caused by inter-signal delays.

4.1.2. All pass filtering

Decorrelation is most commonly achieved by filtering the input signal with all-pass filters having random, noise-like, phase responses [13]. 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 [13] 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 100 zeros.

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

4.2. Dynamic decorrelation

Time-varying or dynamic decorrelation can be defined by the use of time-varying all-pass filters [13]. 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 [13] that dynamic decorrelation creates micro-variations simulating the time-varying fluctuations caused by moving air.

However we found that dynamic decorrelation can 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 Interaural Time Differences (ITD). Therefore, it is left to the discretion of the sonification designer whether fix or dynamic decorrelation should be used.

4.3. 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 Fig. 6. 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 decorrelation technique described above. Cross-fader 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 decor-

related signals are added together to form the final set of partially decorrelated signals.

We have implemented such a decorrelator on the MAX/MSP platform [20] 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.

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.

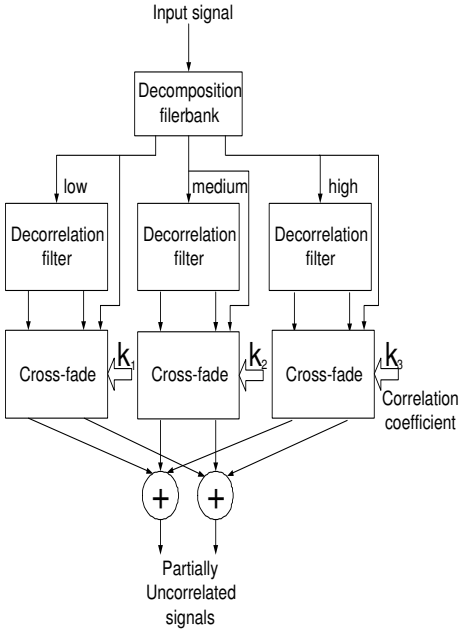

Figure 6: *A three sub-band decorrelator.*

4.4. Other decorrelation techniques

Other decorrelation techniques besides delay and all-pass filtering exist (see [11] for an overview), these are often used in echo-cancelation 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.

We now present experiments we have carried out regarding the use of sound source extent as a sonification parameter.

5. EXPERIMENT I: SOURCE EXTENT AS SONIFICATION DIMENSION

The aim of this experiment was to assess whether the spatial extent of sound sources can reliably be used as a sonification parameter in a 3D audio environment. That is, we were interested in the shift between the intended source extent and the actual perceived source extent by subjects.

The experiment was performed on the Configurable Hemispheric Environment for Spatialised sound (CHESS) [21] which uses fourth order Ambisonics spatialisation on a 16 speaker dome array. The space is not anechoic but has some acoustic proofing.

5.1. 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 (partially shown in Fig. 7). 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).

5.2. 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 the most 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.

5.3. 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 (Fig. 7).

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 [22].

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 sequence 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.

As far as sonification is concerned, it seems that spatial extent of sound sources could be used to carry information, however a sharp mental imaging of the source extent does not seem possible. 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. Training of subjects could be improved so that they improve their ability to judge source extent. Elicitation with point devices and head-tracker would seem also to be more accurate.

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

6. CONCLUSION

Subjective experiments showed that sound source extent can be used to display certain 'sound areas' on a speaker based 3D sound system. With white noise signals, subjects were able to locate the centers of spatially extended sound sources, assess their spatial extents and discriminate sources with an horizontal extent from sources with a vertical extent. These abilities were lessened for sounds coming from behind.

We have also highlighted the link between the inter-correlation of point sources and the inter-aural cross-correlation. We then presented techniques to measure the temporal and spectral variations of the IACC coefficient. These methods are essential tools in the design of auditory displays that use sound source extent and diffuse sound fields.

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.

By finding the correct rendering and user training methodologies, it would be possible to use source extent as a powerful mean to carry size and area information in auditory displays.

7. REFERENCES

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

[2] H. Lehnert "Auditory Spatial Impression" in *Proc. of the 12th Audio Engineering Society Conf.*, Copenhagen, Denmark, June 1993.

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

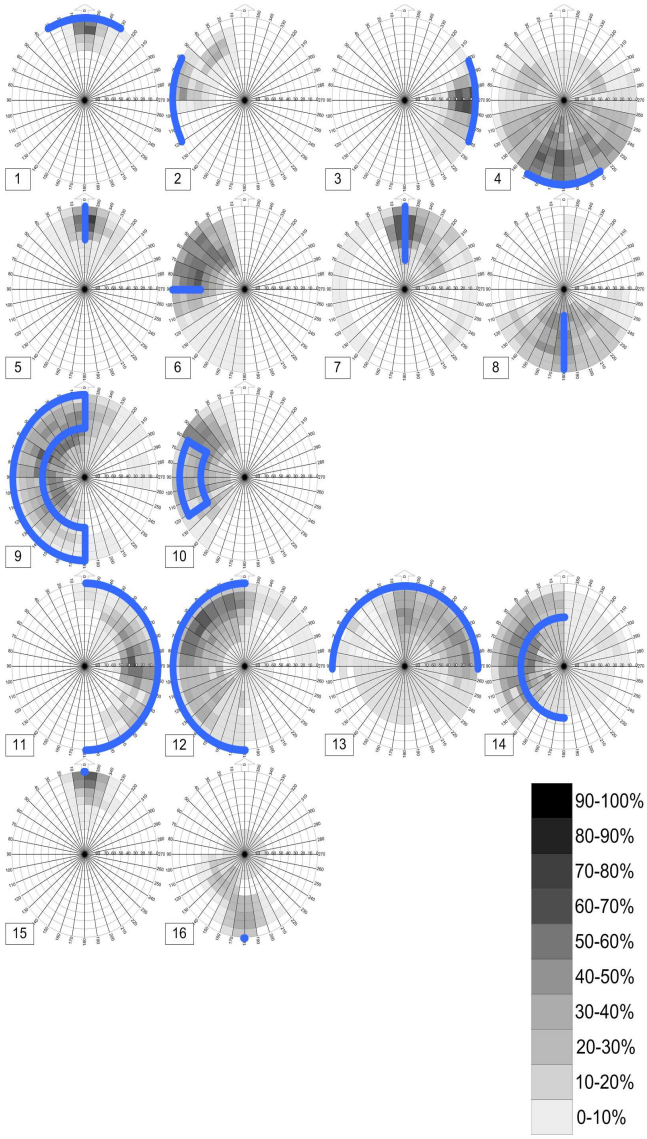

Figure 7: *Density plots of sound source extent perception*

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

[5] 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.

[6] 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.

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

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

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

[10] B.M. Sayers and E.C Cherry, "Mechanism of binaural fusion in the hearing of speech" *J. Acoust. Soc. Am.*, vol. 29, no. 9, pp. 973-987, Sep. 1957.

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

[12] C. Fancourt, L. Parra "A comparison of decorrelation criteria for the blind source separation of non-stationary signals" in *Proc. IEEE Sensor Array and Multichannel Signal Processing Workshop*, Rosslyn, VA, 2002, Aug. 2002, pp 165-168

[13] G.S. Kendal, "The decorrelation of audio signals and its impact on spatial imagery" *Computer Music J.*, vol. 19, no. 4, pp. 71-87, Winter 1995.

[14] D. Griesinger, "Spaciousness and envelopment in musical acoustics"in *Proc. of the 101st Audio Engineering Society Convention*, preprint 4403, Nov. 1996

[15] H. Kuttruff, *Room Acoustics, 3rd edition*, Elseiver applied science, 1991.

[16] F. Rumsey, *Spatial audio*, Focal Press, 2001.

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

[18] D. Menzies, "W-panning and o-format, tools for spatialisation" in *Proc. of the ICAD 2002 Conference*, kyoto, Japan, July 2002.

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

[20] www.cycling74.com

[21] M. O'Dwyer, G. Potard, I. Burnett, "A 16-speaker 3D audio-visual interface and control system", in *Proc. of the ICAD 2004 Conference*, Sydney, Australia, July 2004

[22] G. Potard and I. Burnett, "A study of sound source apparent shape and wideness" in *Proc. of the ICAD 2003 Conference*, Boston, USA, July 2003, pp. 25-28.