Simulation of Extended Sound Sources in Virtual Auditory Environments

Pedro Novo, Noha Korany
Institute of Communication Acoustics, Ruhr-University of Bochum; Email: korany@. novo@ika.rub.de

Introduction

Sound sources found in natural environments exhibit a finite spatial extension. Yet, their actual dimensions and/or their distance to the listener can make them suitable to be modelled in Virtual Auditory Environments (VAE) as point sources [1]. However, when the sound sources are too near and/or are too large, a model capable of accounting for the extension of the sound source has to be employed.

Examples of sound sources likely to require an extended sound source approach are crowd sounds (e.g. applause, speech), environmental sounds (e.g. rain) and man made sounds (e.g. traffic). One possible approach is to model each and every sound source individually. But, for real time interactive performance this is not possible. In fact, if we limit the target hardware platform to PCs, the number of sound sources can not exceed a few dozens (in a 3 GHz PC).

The initial study presented here is limited to the binaural presentation over headphones of the sound sources in free field and further explores the idea of decorrelation between sound sources in order to produce an enlarged auditory event [2]. It focuses in two aspects: the preservation of the perceived quantity of distinctive sound sources when employing the Reduction Method, and the preservation of the Spatial Continuity of the Auditory Event.

The Reduction Method

The Reduction method is employed to reduce the number of talkers around the listening point. The reduction takes place as follows:

(1) The horizontal plane around the listener is divided into *sectors*, each covering an angle $\Delta\phi$ (figure 1a).

(2) Within each *sector* the sound sources, s, are clustered in groups. Within each group the sound sources are replaced by the sound source causing the highest sound pressure level at the listener L . This sound source is renamed to S (figure 1b).

(3) The sound pressure level of each sound source S is raised to equal the total sound pressure level generated at the listener by the totality of the sound sources, s, of its own cluster.

There is however one exception to sound source clustering. Preliminary tests have shown that sound sources in the vicinity of the listener should not be subject to clustering as this results in clearly perceptible differences. Therefore, the sound sources lying inside a restricted area (the dark grey circle in figure 1) are not subjected to clustering. The radius of this circle is defined as the radius of non-clustering. To the purpose of investigating the influence of a radius of non-clustering on the Reduction Method the two following tests were performed.

Tests and Results

For comparison purposes the binaural impulse response at the listening position is calculated before and after employing the Reduction Method. The ratio between the original number of sound sources and the number of sound sources after reduction is defined as the Reduction Ratio.

In the first test, there were employed 94 speech sentences as sound sources, a sector angle of 30 degrees, 6 talkers per cluster for a radius of non-clustering of 0.3m, 1.5m and 2.0m, and 2 talkers per cluster for a radius of non-clustering of 1.0m. The sources were randomly distributed around the listener. Eight subjects have taken part. They were asked to state if the number of talkers perceived before and after using the Reduction Method was { lower, same/nearly the same, higher}. The results of the survey show that the percentage of the number of subjects answering same/nearly the same were 0%, 0%, 44% and 78% respectively for a radius of non-clustering of 0.3m, 1.0m, 1.5m and 2.0m. The Reduction Ratio achieved for these radius of non-clustering was respectively 22%, 57%, 33% and 39%. These results show a stronger dependence of the perceived difference of the number of sources on the radius of non-clustering rather than on the Reduction Ratio.

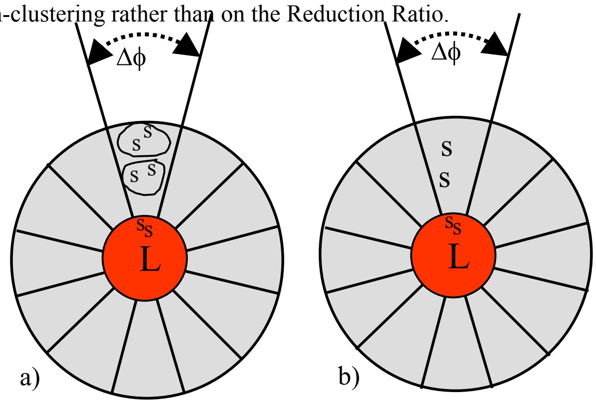

Figure 1: Subdivision of the space around the listener: 's' original sources, 'S' sources after the Reduction Method (sound sources only indicated in one sector). a) step (1) of the Reduction Method; b) steps (2) and (3) of the Reduction Method.

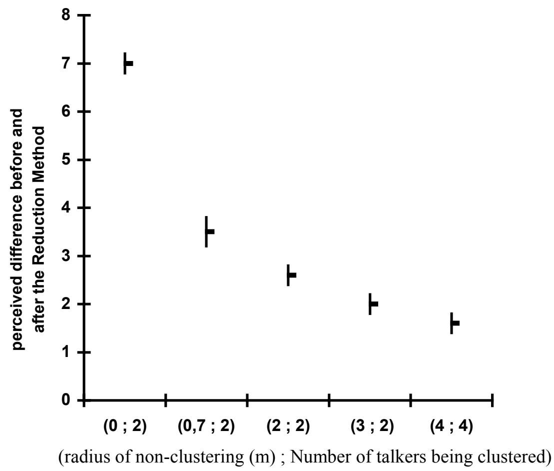

Figure 2 Perceived difference in the distinctive number of talkers as a function of the radius of non-clustering and of the number of talkers per cluster before and after employing the reducing method (8 subjects); '–' average ; '|' standard deviation

In the second test there were employed recorded speech from 31 talkers as sound sources, a sector angle of 30 degrees, a radius of non-clustering of 0.0m, 0.7m, 2.0m, 3.0m and 4.0m and 2 or 4 talkers per cluster. Eight subjects have taken part. They were asked to evaluate the perceived difference in the distinctive number of talkers before and after employing the Reduction Method. The scale {0, 1, 2, …, 10} was employed: 0 means that no difference was perceived and 10 means that the results were perceived as completely different before and after the Reduction Method. The answers of the eight subjects are depicted in graphical form in figure 2.

These results show a strong decrease on the perceived difference in the distinctive number of talkers before and after the Reduction Method with the increase of the radius of non-clustering. The average value for a non-clustering radius of 0.0m is 7 (in the 0 to 10 perceived difference scale where 10 is the maximum difference) which is reduced to 2 for a non-clustering radius of 3.0m.

Spatial Continuity of the Auditory Event

Damaske [2] has shown in an extensive study that the presentation of two decorrelated sound sources at different angles to the listener produces an enlargement of the auditory event. A number of studies that have further investigated the relationship between the decorrelation of the sound signals at the two ears and the imagery of the spatial event (e.g [3], [4] and [5]) have followed.

Damaske also showed that the enlargement of the auditory event does not hold for every angle. Exceeding a certain threshold results in the division of the auditory event into two separate auditory events. To inquire about the directional dependency (as relative to the listener's head) of this threshold some pilot tests were performed.

Test and Results

Two incoherent sound sources of 4 talkers each were simultaneously presented to the listener at different azimuths. The 4 subjects participating in the test were asked to draw the horizontal angle occupied by the perceived auditory event in a circle representing the horizontal plan. A head-tracked binaural system was employed. This system allows up to eight different sound sources being presented simultaneously at different locations.

Figures 3 a) to f) depict the obtained results. On each figure the two arrows pointing inwards represent the directions of incidence of the two sound sources. The arrow in the centre represents the listener head orientation. Although the head-tracking system allows tracking for any movement the listeners were asked to limit their head rotation to small angles (aprox. up to +-20 degrees).

The darker grey sectors in figure 3 represent the more frequently perceived directions of the auditory event. The lighter the grey the less often these directions were perceived. It can be observed from these results that the frontal direction has a lower 'spatial continuity' threshold than the lateral directions.

Figure 5 shows the results obtained when employing 8 decorrelated sound sources. The angles between the sound sources were set based upon results obtained with the two sound sources tests: 30 degrees resolution on the frontal direction, 45 degrees in the rear direction and 60 degrees in the lateral/frontal direction.

It is observed that despite the higher sound sources concentration in the frontal and rear directions the perceived direction of incidence is predominantly from the sides.

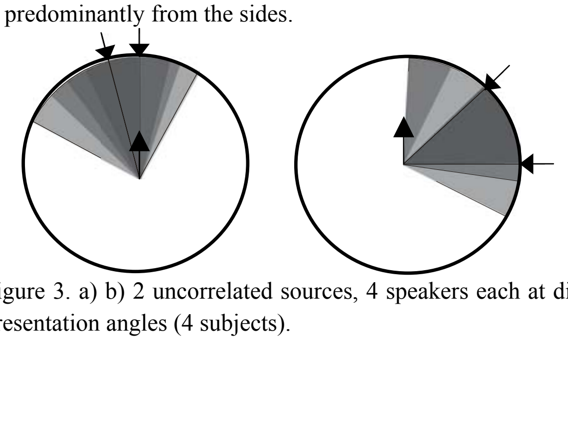

Figure 3. a) b) 2 uncorrelated sources, 4 speakers each at different presentation angles (4 subjects).

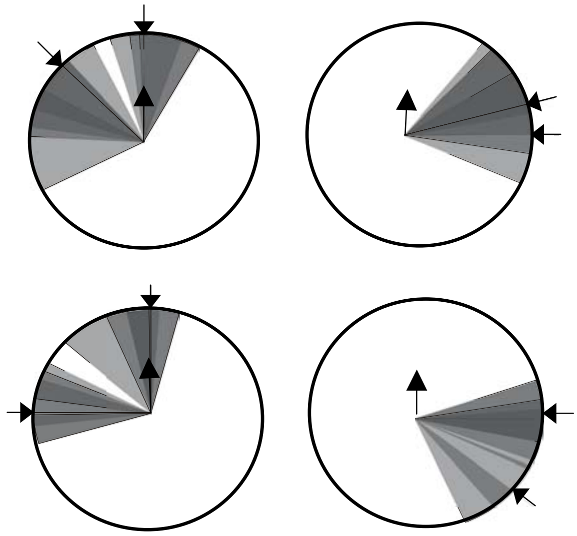

Figure 3. c) to f) 2 uncorrelated sources, 4 speakers each at different presentation angles (4 subjects).

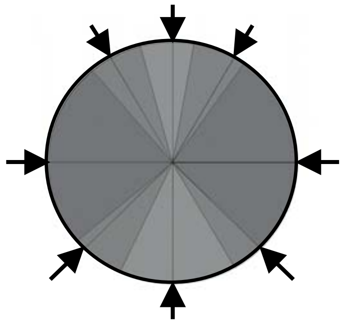

Figure 5 8 uncorrelated sources, 4 speakers each (4 subjects).

Conclusions

An algorithm aiming at reducing the number of sound sources to be modelled in a Virtual Auditory Environment, while maintaining the perceived number, type and horizontal angle of the spatial event was proposed. The sounds were rendered binaurally with only the direct sound being modelled. The results of the tests indicate that the reduction of the number of sound sources should only be performed at a distance from the listener higher than 2m. Also the results show that the generation of a continuous auditory event all around a listener (360 degrees) using eight decorrelated sound sources distributed according to results obtained with two decorrelated sound sources exhibits nevertheless a lateral predominance of the auditory event.

References

[1] Pedro Novo, Auditory Virtual Environments, Proceedings of DAGA 02, Bochum, Germany, 2002.

[2]Damaske, P., "Subjective Untersuchungen von Schalfeldern" Acustica 19, 198-213, 1968

[3]Kurozumi, K. Ohgushi, K., "The relationship between cross-correlation coefficient of two channel acoustic signals and image quality" J.A.S.A. vol. 74(6), pp 1726-1733, Dec. 1983.

[4]Blauert, J. et Lindeman, W., "Auditory Spaciousness: Some further Psychoacoustic Analysis", J.A.S.A., 80(2), August 1986.

[5]Kendall G. S., "The Decorrelation of Audio Signals and its Impact on Spatial Imagery", Computer Music Journal, vol. 19(4), pp 71-87, 1995.

The authors would like to thank Dipl.-Ing. Andreas Silzle for fruitfull discussions.