# Aspects of Hearing in Diffuse Sound Fields

P. Novo

*Institute of Communication Acoustics, Ruhr-University of Bochum, Germany, pedro.novo@ruhr-uni-bochum.de*

**ABSTRACT:** The simulation of speech sounds created by a large crowd has produced an unexpected result. The listeners have reported that the auditory events were mainly concentrated to the left and to right, although the sound sources were equally distributed around the listener. To investigate whether this effect was due to the use of non-individual head-related transfer-functions (HRTFs), a reverberant chamber was used to create a diffuse sound field and a test was performed with 6 listeners. The results of this test have confirmed the previously observed lateral concentration of the auditory events. Being assured that the cause was not on the use of non-individual HRTFs an analysis of the diffuse sound field binaural signals (recorded with a 'dummy head') has followed. The binaural signals were filtered in critical bands and half-wave rectified (modelling processes occurring in the cochlea and in the transduction to neuronal signals) and a normalized cross-correlation function was employed for the lateralization prediction within each critical band. The binaural signal analysis suggests that the key aspect regarding the lateralization dominance of the auditory events is connected to the sections where the cross-correlation coefficient assumes negative values.

## 1. INTRODUTION

Distributed sound sources (e.g. crowd sounds) and diffuse sound fields (e.g. occurring in reflective closed spaces) are present in most daily environments. They actually have something in common as a diffuse sound field can be viewed as a dense collection of distributed sound sources (i.e. reflections) and vice-versa. However, relatively little is known about hearing in these conditions when compared to hearing in single source anechoic conditions.

In fact, the background for this study lies in unexpected results obtained from a simulation of crowd sounds. In this simulation, which used a head-tracked headphones-based virtual environment, the sound sources (i.e. speakers) were densely and homogeneously distributed around the listener. When asked to report the location of the auditory events the listeners have unexpectedly indicated that they were mainly concentrated to their left and right [1]. Furthermore, a similar result was obtained with 'dummy head' recordings of applause [2]. These results have prompted an investigation on whether this effect was due to the use of non-individual Head-Related Transfer-Functions (HRTFs). To this aim a diffuse sound field was created in a reverberant chamber and a test was performed with 6 listeners. In addition, a binaural signal analysis was performed with the aim of finding relations between psychoacoustic results and binaural processing model predictions [e.g. 3].

## 2. 3-D DIFFUSE SOUND FIELD TESTS

In a perfectly diffuse sound field the cross-correlation coefficient (CCC) between the pressure at two points, $p_1$ and $p_2$ (eq. 1) follows a *sinc* function (eq. 2),

$$\psi = \frac{\overline{p_1 p_2}}{\sqrt{\left(\overline{p_1^2}\,\overline{p_2^2}\right)}} \qquad (1)$$

$$\psi(kd) = \frac{\sin(kd)}{kd} \qquad (2)$$

where *k* is the wave number and *d* is the distance between the measuring points [4]. To create a diffuse sound field, the set-up described in Figure 1a was installed in a reverberant chamber. The CCC was measured with two microphones located at the listener position (distanced 30 cm apart), for three different directions, and employing decorrelated 1/3 octave band noise signals at each loudspeaker [4]. The fitting between the measurements and the theoretical curve (eq. 2) was very good.

For the listening test, instead of crowd sounds, recorded rain sounds were employed. The reason for this lies in that highly reverberated crowd sounds lose almost all their impulsive character when played in the reverberant chamber. (The influence of the impulsivity of the sound signals arriving at the listener's ears in the lateral concentration of the auditory events is significant and subject of current research). The 8 loudspeakers were supplied with 8 decorrelated sound signals. The listeners were provided a sheet of paper where a circle of 5 cm radius had been printed. They were asked to indicate the location of the horizontal projection of the auditory

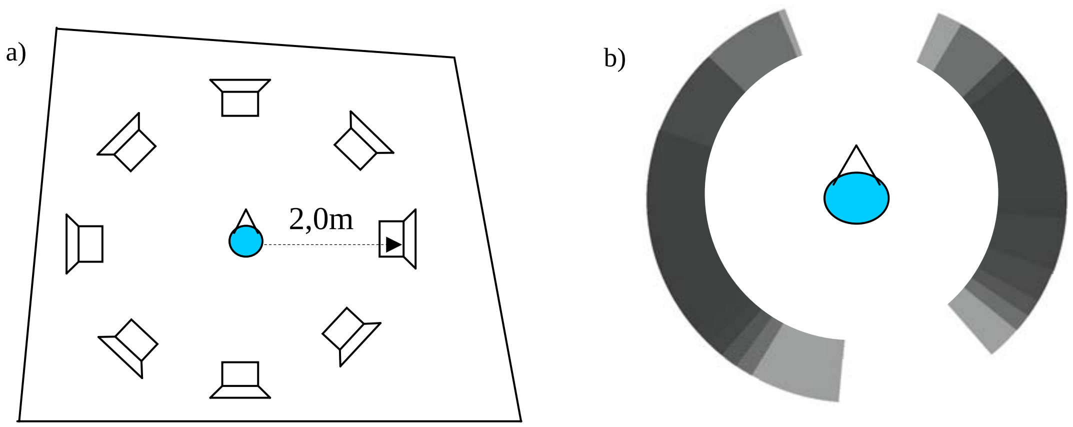


Figure 1- a) Reverberant chamber set-up. Eight decorrelated signals are supplied to the eight loudspeakers. The listener seats in the centre; b) Location of the auditory events. The darker regions indicate a higher incidence of answers.

events, assuming that the centre of the circle was occupied by the listener. The results confirmed the previous observations: a concentration of the auditory events at the lateral positions (Figure 1b). Furthermore, the sound field created in the reverberant chamber was binaurally recorded and subsequently presented to the listeners over headphones. The results of these tests further confirmed the lateral concentration of the auditory events.

## 3. BINAURAL SIGNAL ANALYSIS

Being assured that the cause of this effect was not due to the use of non-individual HRTFs, we proceeded to the analysis of the recorded binaural signals. These were filtered in critical bands and half-wave rectified, modelling processes occurring in the cochlea and in the transduction to neuronal signals [5]. A Normalized Cross-Correlation Function, NCCF, eq. 3,

$$\Phi_{p_l p_r}(\tau) = \frac{\overline{p_l(t) p_r(t+\tau)}}{\sqrt{\left(\overline{p_l^2}\,\overline{p_r^2}\right)}} \qquad (3)$$

was employed for the lateralization prediction within each critical band [5], where $p_l$ and $p_r$ represent the signals recorded at each ear and $\tau$ the interaural delay.

The results of the binaural signal analysis are presented in Figure 2 in the form of a 3D (Figure 2a) and a 2D (Figure 2b) correlogram. The $\tau$ axis represents the interaural delay (in ms) and the frequency axis represents the frequency range (in Hz) being analysed. The vertical axis represents the binaural signals' NCCF values. These results represent an unweighted average of a binaural signal of 1 second duration.

For frequencies below approximately 500 Hz there is a peak in NCCF at $\tau$ = 0.0 ms. From approximately 500 Hz to 1000 Hz there are two peaks of the NCCF at approximately $\tau$ = +-0.725 ms and from approximately 1000 Hz to 1500 Hz there are three peaks: one at $\tau$ = 0.0 ms and two symmetric at approximately $\tau$ = +-0.800 ms. According to this model, a peak at $\tau$ = 0.0 ms indicates an auditory event centred at 0 degrees, in front (or in the back) of the listener, and peaks at +-0.725 ms and at +-0.800 ms indicate auditory events centred near +-90 degrees (i.e. to the left and right sides of the listener).

Another interesting aspect is related to the time of averaging employed to produce the correlogram. In fact if the averaging is performed over a period of 200ms or 600ms as is the case of respectively Figures 3a and 3b the correlogram peaks are not yet completely formed. Only after around 1 sec of averaging (as is the case in Figure 2) does the correlogram starts to arrive to its final form. A 'settle-in time' is in line with the experience reported by the listeners.

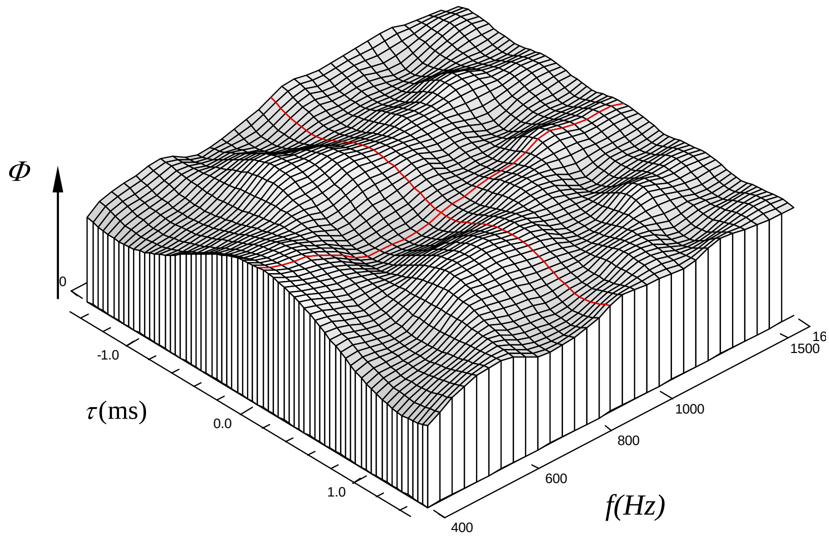


Figure 2- a) 3D Correlogram of the diffuse sound field. $\tau$ axis - interaural delay (associated with lateralization). $f$ axis - frequency range being analysed. Vertical axis - Normalized Cross-Correlation Function, $\Phi$, averaged over 1 second (calculated with [6]).

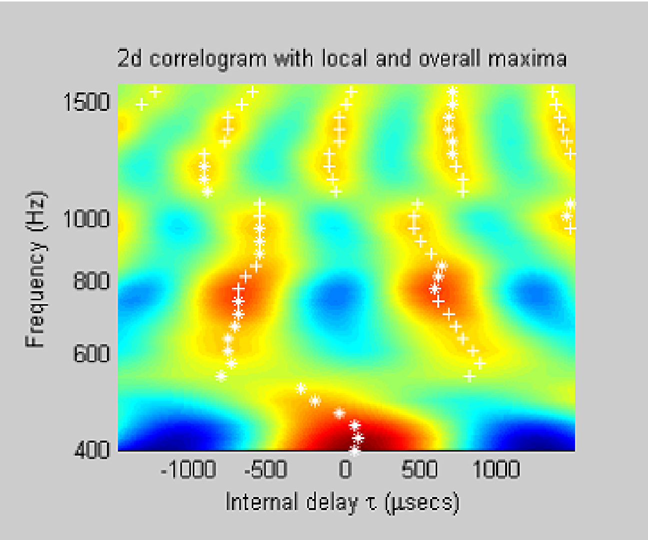


Figure 2- b) 2D correlogram of the diffuse sound field averaged over 1 second. Same data as in figure 2 a) but in a 2D projection: the NCCF values are mapped in colors: red represent the highest and blue the lowest values. The '+'signs represent local maximums and the '*'signs represent global maximums for each frequency band.

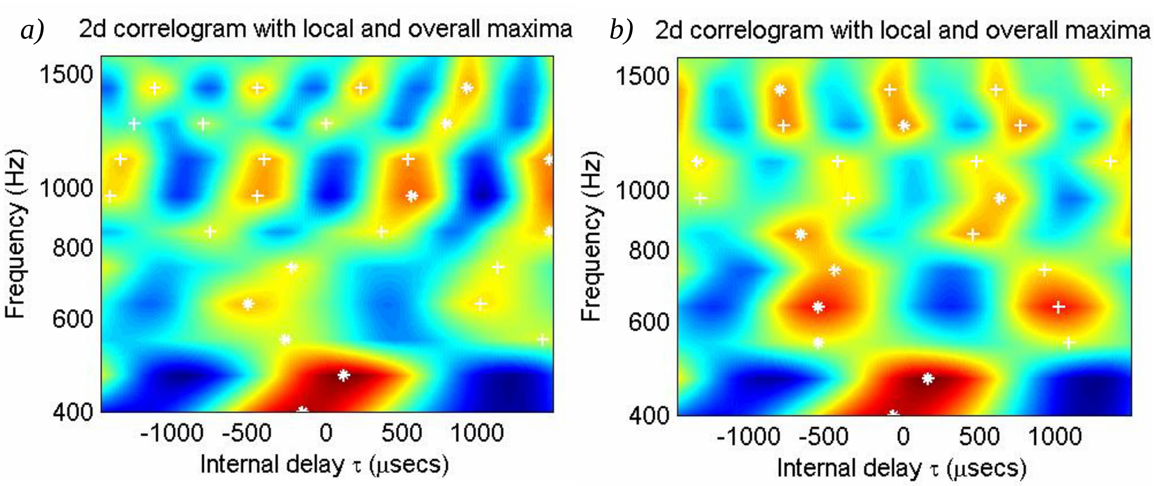


Figure 3- 2D Correlogram of the diffuse sound field averaged over a) 200ms and b) 600ms

Let us now turn the attention to the CCC. When the CCC is calculated using measurements performed at the entrance of the ear canals the first CCC zero occurs at approximately 500Hz, compared to approximately 1200Hz when the CCC is measured at the interaural distance but without the head in-between (Figure 4). The second zero crossing occurs at approximately 1000Hz. In between the CCC takes negative values. Therefore, in the range of frequencies where the CCC (measured with head) takes positive values there is a peak of the NCCF at $\tau$ =0.0 ms and in the range where the CCC takes negative values there are two NCCF peaks symmetric to $\tau$ =0.0 ms (Figure 2). The latter is a very interesting result as it may contribute to the explanation of the lateral concentration of the auditory events previously reported. It is important to note that the absence of a (negative) peak of NCCF at $\tau$ = 0.0 ms for the 500 Hz - 1000 Hz range is related to the halfwave rectification neural transduction model applied in the present study.

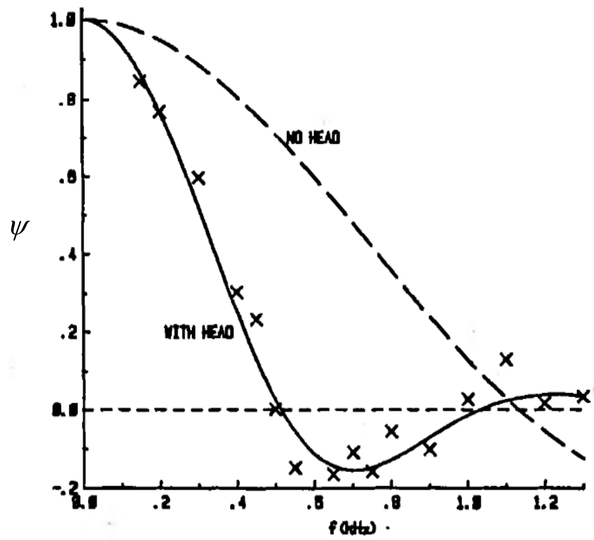


Figure 4 Effect of head on the cross-correlation coefficient of a 3D diffuse sound field [7].

Two further correlograms were calculated. One employing binaural recordings of crowd sounds in a semi-anechoic environment (Figure 5) and another employing binaural recordings of noise impulses emitted by eight loudspeakers in anechoic conditions (Figure 6). In both cases the 'dummy head' employed for the recordings was (homogenously) surrounded by sound sources. The correlograms obtained follow a pattern similar to that of the correlogram of Figure 2 and to the lateralization dominance of the auditory events of Figure 1b.

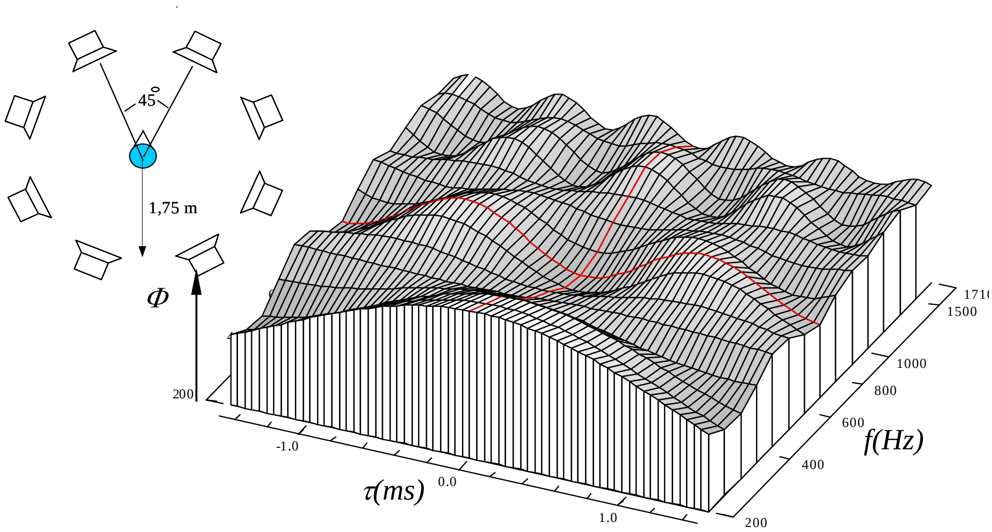


Figure 5- Correlogram, averaged over 1 second, obtained with binaural signals recorded in a sound field created by noise impulses in a anechoic chamber ; a) set-up ; b) 3D correlogram

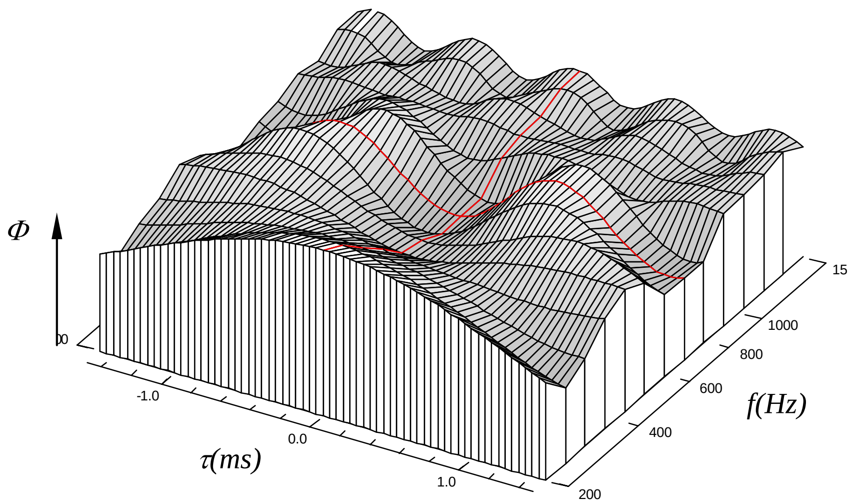


Figure 6- Correlogram, averaged over 1 second, obtained with binaural signals recorded in a sound field created by 40 speakers homogeneously distributed between a radius of 2m and 4m around  the listener.

The CCC of a 2D diffuse sound field follows a Bessel function of order zero and not a sinc function as is the case of a 3D diffuse sound field [8]. In practice this implies that, in an ideal 2D diffuse sound field, the zeros of the CCC occur at a lower frequency than in a 3D ideal diffuse sound field.

It is, then, plausible that the two peaks of the NCCF occurring approximately between 500Hz and 1000Hz in Figure 2 may play a role in the lateral concentration of the auditory events displayed in Figure 1b. In fact preliminary listening tests performed with narrow band signals support this hypothesis. Furthermore, this frequency region fits within the range reported in Potter et. al [10] as optimal for spaciousness and conforms with the findings of Blauert and Lindemann [9], which report that the low frequency region adds front-back extension to the auditory image.

## 4. CONCLUSION

1) We have started with a 'problem' of unexpected lateral concentration of the auditory events in virtual environments. Lateralization is not a problem from the simulations, it occurs in real environments.

2) It is proposed that the key aspect regarding the (right/left) lateralization dominance observed with a variety of signals, all near a 2D or a 3D diffusion situation, is related to the sections where the correlation coefficient takes negative values. In this situation cross-correlation function peaks symmetric to $\tau = 0.0$ms occur. When the correlation coefficient is positive, a peak appears in the cross correlation function at $\tau = 0.0$ms.

3) The correlogram arrives to values that support the explanation proposed in point 2) only after about 1 sec of averaging. In fact, a 'settle-in time' is in line with the experience reported by the listeners.

4) As the correlation coefficient is governed by a sinc function (for a 3D diffuse sound field) there is an alternation between positive and negative values of the correlation with changing frequency. The peaks delays are frequency dependent, therefore contributing to an extended auditory event.

## ACKNOWLEDGEMENTS

I would like to thank Prof. Jens Blauert's many comments in the development of the work presented here as well as to my colleague Juha Merimaa for the various fruitful discussions on binaural modeling. I would like also to express my appreciation to the colleagues who have generously offered their time for the psychoacoustics experiments.

## REFERENCES

[1] Novo, P., Korany, N. (2003) Simulation of Extended Sound Sources in Virtual Environments. In: Fortschr. Der Akustik- DAGA'03, Dtsch. Ges. Akust., D-Oldenburg
[2] Kleiner, M. Private Communication
[3] Novo, P. (2004) Experiments with Spatially Distributed Sound Sources in Real and Virtual Environments, In proceedings of DAGA'04, Strasbourg, (in print)
[4] Cook, R., Waterhouse, R., Berendt, R., Edelman, S., Thompson, M. (1955) Measurements of Correlation Coefficient in Reverberant Sound Fields. J. Acoust. Soc. Am. 27 (6), pp. 1072-1077
[5] Blauert, J. (1996) Spatial Hearing: The Psychophysics of Human Sound Localization. Rev. Edition. The MIT Press, Cambridge Mass.
[6] Akeroyd, M. (2001) A Binaural Cross-Correlogram Toolbox for MATLAB. www.biols.susx.ac.uk/ home/Michael_Akeroyd/
[7] Benade, A., Lindevald, I. (1986) Two-ear correlation in the statistical sound fields of rooms J. Acoust. Soc. Am. 80, pp. 661-664
[8] Kuttruff, H. (2000) Room Acoustics, fourth edition, Spon Press
[9] Blauert, J., Lindemann, W. (1986) Auditory Spaciousness: Some Further psychoachoustic analysis. J. Acoust. Soc. Am., 79, pp.806-813
[10] Potter, J., Bilsen, F., Raatgever, J. (1995) Frequency Dependence of Spaciousness. Acta acustica, vol.3, pp. 417-42