Source Positioning in a Two listener Crosstalk Cancellation System

Bruno Masiero*

Institute of Technical Acoustics, RWTH Aachen University, Germany
Email: bma@akustik.rwth-aachen.de

Introduction

To replicate the spatial impression of a sound scene, the spatial cues of the sound field arriving at the listener's head have to be emulated, which is done by providing the listener with a binaural signal. If the binaural signal [1] is played back to the listener through headphones, each channel will be heard by only one ear. On the other hand, it is long known [2] how to correctly reproduce a binaural signal through a pair of loudspeakers. The binaural signal has to be pre-filtered to compensate for the crosstalk effect that will otherwise ruin the spatial clues contained on the binaural signal.

The filters for the crosstalk cancellation (CTC) system are designed based on the transfer paths between loudspeakers and listener ears (or a pre-recorded *head-related transfer function (HRTF)* database) and usually delivers channel separations of over 20 dB in a small region around the listeners head, the so-called *sweet spot*. If the listener moves away from the sweet spot, channel separation will deteriorate and the spatial cues from the binaural signal will be lost. To allow the users to move and rotate their heads, dynamic CTC systems use a head-tracking device to determine distance and orientation of the listener's head to the loudspeakers, than choosing the corresponding HRTF from the database and constantly updating the cancellation filters [3].

Depending on the relative orientation between head and loudspeakers the matrix formed by the transfer paths between loudspeakers and listener ears may become ill-conditioned for some frequencies, leading to an unstable cancellation filter set. To overcome this problem, Lentz developed a system with four loudspeakers instead of the usual two loudspeakers used for crosstalk cancellation. The system determine which pair of loudspeakers to be used based on the information of head position and orientation, allowing the listener to rotate his head and freely move in between the speakers [4].

Crosstalk cancellation systems were originally proposed for single listener use. But Bauck and Cooper [5] mathematically proved that the CTC method (called by them as *transaural* method) can be expanded for multiple listeners with an increase in the number of required loudspeakers and consequently on the number of cancellation filters. Motivated by the infrastructure of Lentz's *Virtual Head-Phone* setup, a first step towards a multiple listener CTC system was done by testing a practical implementation of a two listener CTC system. The measurements showed an unsatisfactory channel separation.

Kim *et al.* published the first simulations involving a two listeners CTC system [6], with an optimized loudspeaker arrangement to minimize the conditioning of the systems transfer matrix and to construct more stable filter sets. Using an improved model for the two listener CTC system [7], where the heads of the two listeners are modeled as two rigid spheres – allowing phenomena such as wave diffraction and interaction between diffraction from both spheres to be taken into account – an optimization procedure was undertaken to improve the transfer matrix conditioning by repositioning the loudspeakers.

Two Listeners CTC

This section makes a brief recapitulation of the multi-listener transaural reproduction technique proposed by Bauck and Cooper [5].

We assume two binaural signal p_1 and p_2, each containing left and right channels, and we want to reproduce the signals e_1, e_2, e_3 and e_4 at the corresponding ears of two listeners. This is done by feeding these binaural signals to a filter bank that produces four loudspeaker signals s_1, s_2, s_3 and s_4. These signals can now be represented in vector form as follows:

$$\mathbf{p} = [p_1(l) \quad p_1(r) \quad p_2(l) \quad p_2(r)]^T, \tag{1}$$

$$\mathbf{e} = [e_1 \quad e_2 \quad e_3 \quad e_4]^T, \tag{2}$$

$$\mathbf{s} = [s_1 \quad s_2 \quad s_3 \quad s_4]^T, \tag{3}$$

where T represents vector transposition and the signals are assumed to be in frequency domain. We now define two transfer matrices: $\mathbf{X}$ as the four by four acoustic transfer path matrix where each element x_{ij} is the transfer function from the j^{th} speaker to the i^{th} ear and $\mathbf{Y}$ as the four by four crosstalk cancelling matrix where each element y_{ij} is the transfer function from the j^{th} input signal to the i^{th} output signal, as shown in Figure 1.

The acoustic propagation can then written as

$$\mathbf{e} = \mathbf{X}\,\mathbf{s}, \tag{4}$$

and the crosstalk filter action as

$$\mathbf{s} = \mathbf{Y}\,\mathbf{p}. \tag{5}$$

We wish to have the respective binaural signals reproduced at each ear, apart from a time delay intrinsic to the filters and the acoustic system. This leads to the requirement $\mathbf{e} = \mathbf{s}\, e^{-j\omega\Delta}$, where Δ represents the delay. It therefore follows

---

[*] This research is supported by CNPq, National Council for Scientific and Technological Development – Brazil.

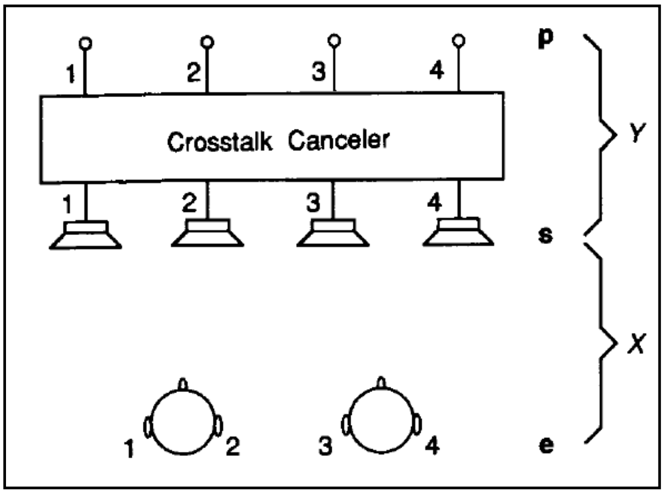

Figure 1: Schematics of a two listener CTC system. Adapted from [5].

$$\mathbf{XY} \triangleq \mathbf{I} \cdot e^{-j\omega\Delta}, \tag{6}$$

where $\mathbf{I}$ denotes the identity matrix; thus at each angular frequency ω the solution for the cross-talk cancellation filter is given by

$$\mathbf{Y}(\omega) = \mathbf{X}(\omega)^{+} \cdot e^{-j\omega\Delta}. \tag{7}$$

Note that $(\cdot)^{+}$ indicate the generalized inverse of a matrix and in the case of a square matrix can be calculated by $\mathbf{X}^{+} = (\mathbf{X}^{H}\mathbf{X})^{-1}\mathbf{X}^{H}$. The matrix $\mathbf{X}$ gets ill-conditioned if, for example, we were to place the loudspeakers in the vertices of a square and the listeners over a line that divides this square in two equal parts. In this situation the acoustic paths from the loudspeakers at each side of the dividing line to the ears of the listeners would be practically the same, meaning that the columns of the resulting transfer matrix would be linearly dependent and are the matrix, therefore, ill-conditioned. This square source distribution could then be called the “worst case scenario” for the two listener CTC system. Nevertheless, other source distributions will also result in ill-conditioned transfer matrices at specific frequencies. A well-established technique for dealing with ill-conditioned inversion problems is the use of regularization [8], which results in

$$\mathbf{Y}(\omega) = [\mathbf{X}(\omega)^{H}\mathbf{X}(\omega) + \beta\mathbf{I}]^{-1}\mathbf{X}(\omega)^{H} \cdot e^{-j\omega\Delta}, \tag{8}$$

where β is the regularization parameter.

A measurement was made in a hemi-anechoic chamber using four monitor loudspeakers and two artificial heads placed as illustrated (not in scale) in Figure 2.

The obtained channel separations are shown in Figure 3. These graphics shows a channel separation in the order of 10 dB, a value insufficient for a realistic virtual reality environment, since typically the *interaural level difference* cue have values up to 20 dB. In a dynamic situation, where the listener does not stay completely still, if the transfer matrix is ill-conditioned then a small movement of the listeners will result in a big error in the received binaural signal, making the reduced channel separation even more critical.

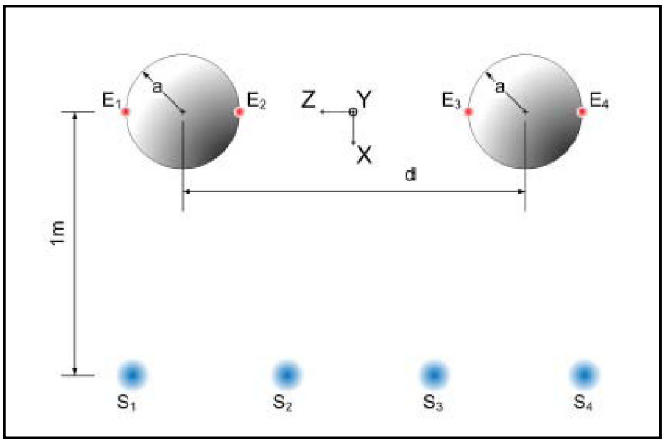

Figure 2: Schematic transducers position. The distance *d* between the artificial heads corresponds to 0,5 m and the distance between the sources is 0,6 m.

As already noted by Nelson and Rose [8], the reduced channel separation is probably a by-product of the use of regularization. But abandoning the use of regularization will result in a filter matrix $\mathbf{Y}$ whose elements might have a higher order of magnitude or, in other words, in a filter with greater amplification in unstable frequencies. At these frequencies the filters will send a great amount of energy to the transducers, energy which will be cancelled only to leave a small level of synthesized binaural signal at the listeners’ ears. The higher amplification directly results in the loss of dynamic range, since it forces the global gain of the system to be reduced, thus, preventing digital clipping of the signal and non-linear distortion of the loudspeakers in these frequencies [9].

Optimizing Source Position

Bearing in mind the situation expose above, in a practical implementation of a two listener CTC system the position of the loudspeakers must be optimized to reduce the condition number of the transfer matrix, in this way allowing for more stable cancellation filters and consequently bigger dynamic range and larger sweet spot.

The single CTC system gave origin to a long discussion about the ideal placement of its transducers, with some researchers proposing an angle of 10° between the speakers (known as “Stereo Dipole”), others an angle of 60° or even 120° between the speakers. It seems that the discussion is now converging to the solution named “Optimal Source Distribution” [9], which claims that, for each frequency, an ideal angle between speakers exists, being this angle near 180° for low frequencies and near 0° for the higher frequencies.

Following the same idea, Kim *et al.* published the first simulations involving a two listeners CTC system [6]. They used a free-field model with four spherical point sources and four point receivers to optimize the sound source arrangement by minimizing the condition number of the systems transfer matrix.

To take into account the head reflections, a new model was used with two rigid spheres to represent the listeners’ heads, as explained in [7].

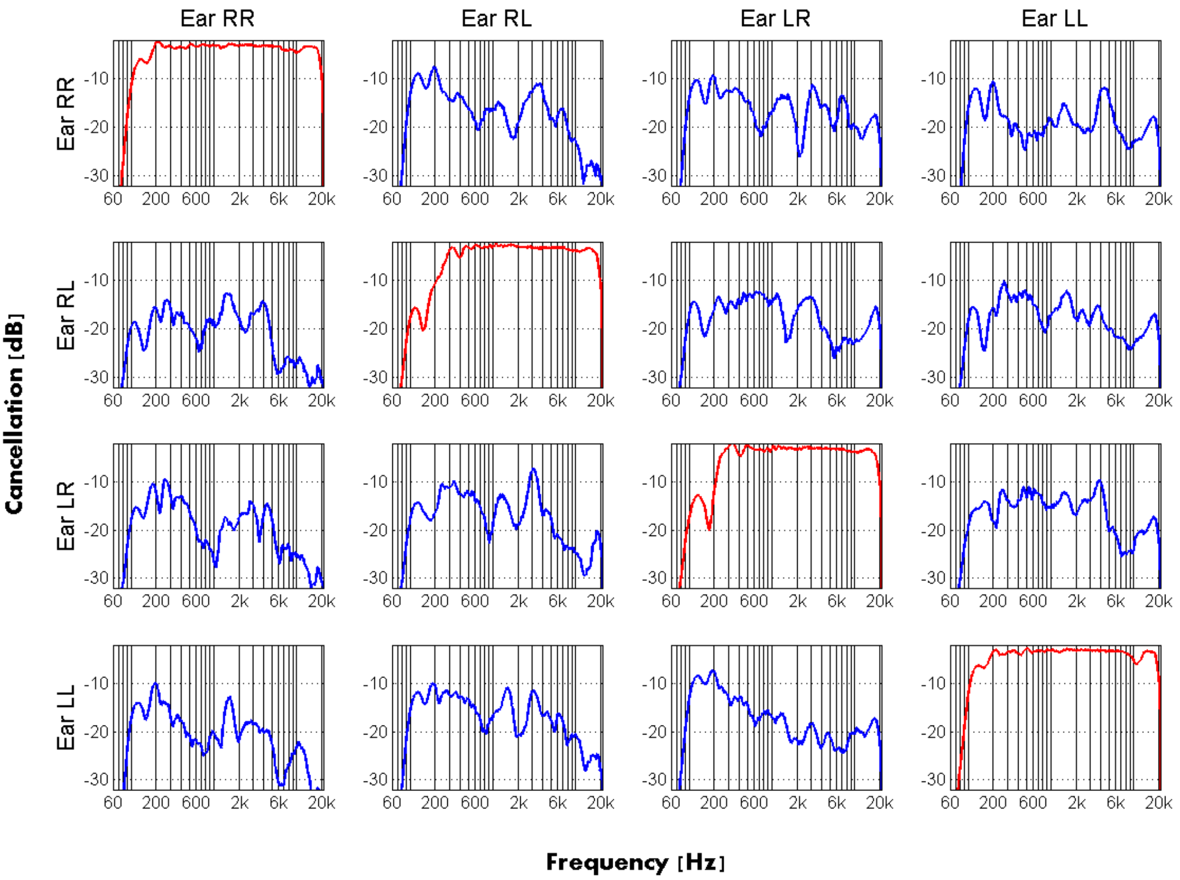

Figure 3: Magnitude spectra of the signals delivered to the microphones of the artificial heads, normalized by the maximum observed value. The used excitation signal was a log-sweep ranging from 100 Hz to 20 kHz. LR stands for left head, right ear, RL for right head, left ear and so on.

The simulated geometry is the same used in the measurements (Figure 2) with the origin of the coordinate system between both spheres. The radii of the spheres were chosen to be $a = 0.09$ m and the distance between their centers was set to $d = 0.52$ m. The possible source locations are assumed to be at 0.05 m intervals within the range from -0.6 to 0.6 m along the line $(x = 1, y = 0)$ m. These distances also coincide with the ones used by Kim *el al.* for their simulations.

When using only one source arrangement for the two listeners CTC system, it is necessary to account for an elevated condition number at certain frequencies, as depicted in Figure 4. This figure shows the value of the condition number as a function of frequency for the source arrangement that minimizes the average condition number over the frequency band from 0 to 5 kHz. The source arrangement that minimizes the average condition number was found by placing the point sources at $z = (-0.6, -0.3, 0.3, 0.6)$. It is important to point out that other source arrangements not tested here might yield better results. The peaks present in this curve occurs due to the fact that the system is badly conditioned in this regions, what is equivalent to saying that the filter gains at these frequencies are considerably high or that the generated sound field varies rapidly in space.

By choosing an appropriate source distribution for each frequency it is possible to keep the condition number below 5 dB for frequencies higher than approximately 1 kHz. The value of the smallest condition number found for each frequency can also be seen in Figure 4. The position that gives the smallest condition number for each frequency for the two rigid sphere model is presented at Figure 5. Note that even though no symmetry was forced for the source positioning, the source distribution is relatively symmetrical and, like verified for the free-field model [6], each neighboring frequency requires a different source arrangement. This leads to an impractical system implementation, since it would require a great number of loudspeakers positioned very closely to each other and filters with very narrow band pass region to distribute the signal to the correspondent loudspeakers.

In search of a compromise between the number of required sources and the value of the condition number Kim [6] defined six frequency bands and found for each band the source arrangement that minimizes the average condition number within each band. This procedure was repeated for the two sphere model and the results are presented in Figure 6. In comparison with the results obtained for the free-field model, an improvement of around 5 dB can be verified at the two higher frequency bands. This improvement can be explained by the fact that the scattering from a sphere becomes very directional for $ka > 5$ (in this case $f > 3$ kHz) and if one sphere is positioned in the shadow region

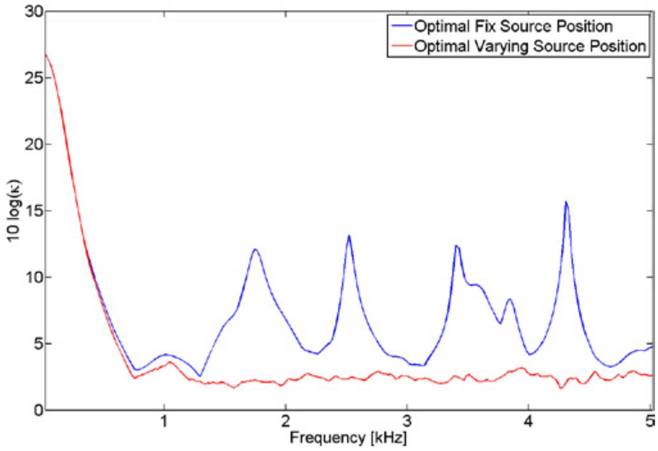

Figure 4: The behavior of the condition number from the transfer matrix with one fix position that gives the smallest mean condition number (with peaks) and with varying source position (smoother curve).

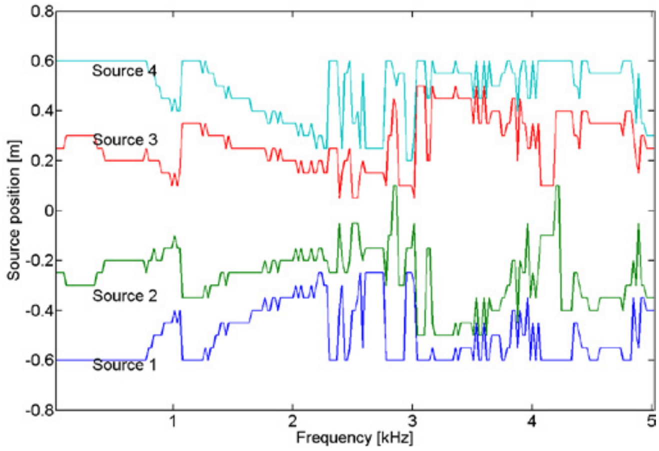

Figure 5: Position of the four point sources that gives the smallest condition number at each frequency. Even though no symmetry was forced within the possible combinations, the distributions tend to be symmetric for the given receiver geometry.

of the other sphere, the systems coupling is reduced and thus the condition number decreases as well.

Conclusions

In this paper we verified that the two listeners crosstalk cancellation system, already proven by Bauck and Cooper to be mathematically feasible, is realizable but under severe practical limitations, since the measured channel separation was approximated only 10 dB and the signal levels at the receivers were highly reduced due to the unstable crosstalk cancellation filter that result from an ill-conditioned transfer matrix.

Simulation with the rigid sphere model gave a source arrangement that minimizes the condition number of the transfer matrix for every frequency. This ideal position varies significantly for neighboring frequencies, frustrating a practical implementation of the system, since very narrow frequency dividing filters would be needed. A band optimized system reduces the overall conditioning of the transfer matrix and reduces the number of sources needed and is possibly the optimal compromise to guarantee stability to this system. Even though, the number of required loudspeakers would still be considerably high.

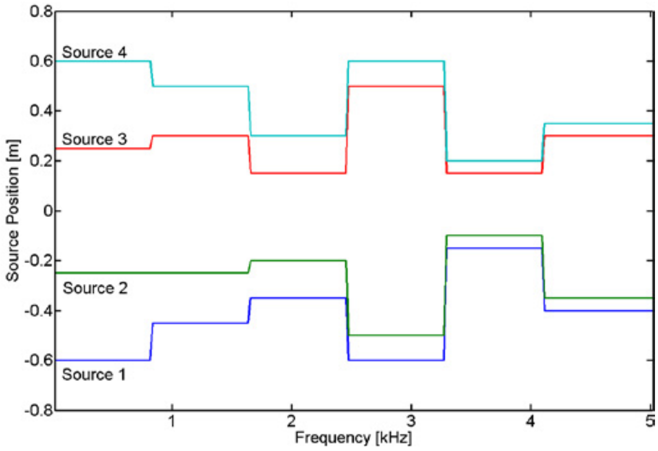

Figure 6 The point source positions that minimize the average condition number over each frequency band.

It was verified that a two listeners CTC is a very challenging problem, considerably more complicated than its single user version and with this in mind, it is hard to foresee the development of a multi-listener CTC in the near future.

References

[1] Vorländer, M. Auralization *Springer-Verlag,* 2008

[2] B. S. Atal, M. Hill, and M. R. Schroeder, "Apparent Sound Source Translator," U.S. Patent 3,236,949 (1966 Feb. 22).

[3] Lentz, T. Dynamic Crosstalk Cancellation for Binaural Synthesis in Virtual Reality Environments, Journal of the Audio Engineering Society **54** (2006), no. 4, 283–294.

[4] Lentz, T. Binaural Technology for Virtual Reality, *Institut für Technische Akustik, RWTH-Aachen,* 2007

[5] Bauck, J. and Cooper, D. H. Generalized transaural stereo and applications, *Journal of the Audio Engineering Society,* 1996, 44, 683-705

[6] Kim, Y.; Deille, O. and Nelson, P. A. Crosstalk cancellation in virtual acoustic imaging systems for multiple listeners, *Journal of Sound Vibration,* 2006, 297, 251-266

[7] Masiero, B. and Qiu, X., Two Listeners Crosstalk Cancellation System Modelled by Four Point Sources and Two Rigid Spheres, *Acta Acustica united with Acustica,* 2009, 95, 379-385.

[8] Nelson, P. A. and Rose, J. F. W. The time domain response of some systems for sound reproduction, *Journal of Sound and Vibration,* 2006, 296, 461-493

[9] Takeuchi, T.; Teschl, M. & Nelson, P. Subjective evaluation of the Optimal Source Distribution system for virtual acoustic imaging. *19th International Conference: Surround Sound - Techniques, Technology, and Perception,* 2001