TRANSAURAL STEREO IN A BEAMFORMING APPROACH

Markus Guldenschuh, Alois Sontacchi

Institute of Electronic Music and Acoustics,
University of Music and Dramatic Arts
Graz, Austria
guldenschuh@iem.at, sontacchi@iem.at

ABSTRACT

This paper presents a study on algorithms for headphone-free binaural synthesis using a dedicated loudspeaker configuration. Both algorithms under investigation improve the properties of the binaural synthesis performance of the array. Firstly, beam-forming provides sound radiation localized at two freely adjustable, narrow target spots. Adjusting both spots to the locations of the listener's ears achieves a good basis. Secondly, an additional interaural crosstalk canceler improves the overall result.

1. INTRODUCTION

This paper investigates the capabilities of a loudspeaker array to focus spatialized sound. Humans with normal hearing abilities can identify the position of sound sources due to differences between the right ear signal, and the left ear signal [1]. Such binaural signals can be used to synthesize spatialized sound [2]. However, they need to be fed to the ears directly. In the case of loudspeakers, the influence of the existing cross talk paths has to be removed. This approach is called transaural stereo [3, 4]. A first binaural sound system for loudspeakers and tracked users has been developed in [5]. Improvements were achieved in [6] by using a set of 4 loudspeakers placed around the user which guarantees stable cross talk filters for a full rotation of the user. [7] tried to gain robustness against lateral mismatches by applying a crosstalk network from 6 loudspeakers to 6 control points (instead of to 2 ear positions only). [8] introduced a circular array of 22 loudspeakers that produces a focus point above the listener for which the cross talk filters are applied. Focusing the sound bears two advantages. First, the room is not excited as much as with common open sound systems, and second, the focused beams already cause a reduction of the crosstalk.

We designed a transaural beamformer for the usage in air traffic control. (See also [9].) In this work, as a first step, different beamforming methods (i.e. a near field beamformer, a minimum variance beamformer and a least squares beamformer) are simulated and compared. The second step concludes the examinations with measurements and in a third step a processing efficient cross talk canceler is introduced and evaluated with measurements, too.

2. CONCEPT AND METHODOLOGY

The preconditions in air traffic control deliver two relevant design criteria:

1. The bandwidth in air traffic control reaches from 300 to 2500 Hz.
2. The system should be desktop integrable and processing efficient.

Out of these criteria, the array properties (like shape, size and number of loudspeakers) and the beamforming method have to be deduced.

We simulated different array types and different beamforming methods. The simulations are derived with the help of the Green's function for omnidirectional point sources.

$$G(\mathbf{r}'|\mathbf{r}) = \frac{1}{4\pi|\mathbf{r}'-\mathbf{r}|} e^{-jk|\mathbf{r}'-\mathbf{r}|}, \tag{1}$$

with wavenumber $k = \frac{\omega}{c}$, where ω is the radial frequency and c the speed of sound. The sound pressure in an arbitrary focus point can be calculated over a superposition of Green's functions form every loudspeaker position $\mathbf{r}'_l$ (with $l = 1...L$) to that specific focus point $\mathbf{r}_f$ [10]. Combining the Green's functions to a vector

$$\mathbf{h}(\omega) = \begin{bmatrix} G(\mathbf{r'_1}|\mathbf{r_f}) & G(\mathbf{r'_2}|\mathbf{r_f}) & \cdots & G(\mathbf{r'_L}|\mathbf{r_f}) \end{bmatrix}^T, \tag{2}$$

allows for a compact vector equation

$$p_f(\omega) = \mathbf{h}^T(\omega)\, \mathbf{q}(\omega), \tag{3}$$

where $p_f(\omega)$ is the sound pressure in the focus point and the entries of vector $\mathbf{q}(\omega)$ are the complex weights of the loudspeakers. In addition to the sound pressure in the focus point, N other field points can be considered as control or evaluation points. Using the matrix

$$\mathbf{G}(\omega) = \begin{pmatrix} G(\mathbf{r'_1}|\mathbf{r_1}) & G(\mathbf{r'_2}|\mathbf{r_1}) & \cdots & G(\mathbf{r'_L}|\mathbf{r_1}) \\ G(\mathbf{r'_1}|\mathbf{r_2}) & G(\mathbf{r'_2}|\mathbf{r_2}) & \cdots & G(\mathbf{r'_L}|\mathbf{r_2}) \\ \vdots & \vdots & \ddots & \vdots \\ G(\mathbf{r'_1}|\mathbf{r}_N) & G(\mathbf{r'_2}|\mathbf{r}_N) & \cdots & G(\mathbf{r'_L}|\mathbf{r}_N) \end{pmatrix}, \tag{4}$$

yields the N entries long sound pressure vector $\mathbf{p}(\omega)$

$$\mathbf{p}(\omega) = \mathbf{G}(\omega)\mathbf{q}(\omega), \tag{5}$$

due to the source strength vector $\mathbf{q}(\omega)$. We evaluated the sound field in an area of 112×168 cm^2. This area was sampled with a distance of $\Delta x = 7$ cm according to the spatial aliasing constraint

$$\Delta x < \frac{\lambda}{2}, \tag{6}$$

where λ is the wavelength of the upper cut off frequency of the bandwidth.

Based on the simulations, three important characteristics are evaluated. First, the width of the beams, second, the gain (which in microphone array literature is known as white noise gain [11].)

$$\mathrm{WNG}(\omega) = 10 \log \left(\frac{|\mathbf{h^T q}|^\mathbf{2}}{\mathbf{q^H q}} \right), \tag{7}$$

where ^{H} denotes Hermitian transposition, and third, the signal to noise ratio SNR. We define the SNR as the difference between the sound pressure level (SPL) in the focus point (i.e. $20\log(p_f)$) and the SPL L_r of the excited reverberant room. L_r is derived over the acoustical power

$$P_{ak} = \oint_S J\, dS, \tag{8}$$

where $J(\phi)$ is the sound intensity which is evaluated on a half circle around the array. In air traffic control centers, we can at least assume A=100 m^2 reflecting walls. The sound pressure in the reverberant room L_r can then be estimated as [12]

$$L_r = 10\log\left(\frac{P_{ak}}{P_0}\right) - 10\log(A) + 6\text{dB}, \tag{9}$$

where $P_0 = 10^{-12}$ W. Obviously, the SNR increases with the number of used loudspeakers. Our simulations were done with an array of 16 loudspeakers as this proved to be sufficient to cover the given area with focus points of high enough SNRs. All broadband simulations are derived by a uniform superposition of $\mathbf{p}(\omega)$ (and $p_f(\omega)$, respectively) for 65 frequency bins at 6000 Hz sampling frequency. In all sound field figures, the energy was normed to the energy at the focus point.

3. ARRAY PROPERTIES AND FOCUSING METHODS

The first question that arises is the one for the shape of the array. It is evident that a circular array is very suitable to create focus points since the loudspeakers contribute to the focus from every direction. However, it does not fulfill the constraint of an easy to mount and desktop integrable hardware solution. A section of an elliptically bent array is an optimal trade off between a circular and a straight array. It is easier to mount than a circular array and has better focusing properties than a straight array as it can be seen in Fig. 1.

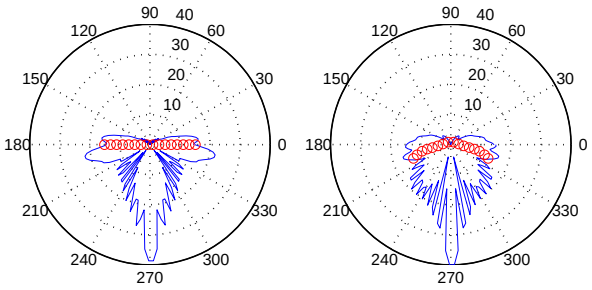

Figure 1: *Focusing properties for a straight and a bent array at 2500 Hz. Circles represent the positions of the loudspeakers. The bent array produces a sharper beam and has lower side lobes than the straight array.*

In the following, all simulations and measurements were done with a bent array of 16 loudspeakers.

For microphone arrays, various focusing methods are known [11, 13, 14]. Due to the tight relations to loudspeaker arrays, these methods can also be applied to calculate the driving functions of the loudspeakers. We examined the classical near field beamforming method, the least squares method and the minimum variance method.

3.1. Near field beamformer

The near field beamformer (NFB) compensates the delays of the Green's functions from the loudspeakers to the focus point. Hence, its source strength vector is the complex conjugate of $\mathbf{h}$. In order to reach a constant WNG, the source strength vector is normalized by the sum of its amplitudes,

$$\mathbf{q}(\omega) = \frac{\mathbf{h}^*(\omega)}{\sum_{l=1}^{L} |h_l(\omega)|}. \tag{10}$$

A simulation of a NFB is shown in Fig. 2.

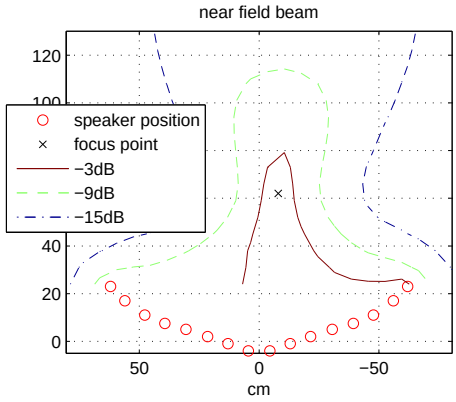

Figure 2: *The simulation shows sound pressure iso-curves for -3, -9 and -15 dB for a near field beamformer.*

3.2. Least squares beamformer

To derive the source strength vector of the least squares beamformer (LSB), we first extend matrix $\mathbf{G}(\omega)$ to

$$\tilde{\mathbf{G}}(\omega) = \begin{bmatrix} \mathbf{h}^T \\ \mathbf{G} \end{bmatrix}, \tag{11}$$

and $\mathbf{p}(\omega)$ to

$$\tilde{\mathbf{p}} = \begin{pmatrix} p_f \\ \mathbf{p} \end{pmatrix}. \tag{12}$$

We can then make a *'wish'* for a sound pressure distribution $\tilde{\mathbf{p}}$ and derive, according to the least squares error solution, the source strength vector

$$\mathbf{q} = (\tilde{\mathbf{G}}^H\tilde{\mathbf{G}})^{-1}\tilde{\mathbf{G}}^H\tilde{\mathbf{p}}. \tag{13}$$

It can be seen that the matrix $(\tilde{\mathbf{G}}^H\tilde{\mathbf{G}})$ has to be inverted. Thus, it has to be taken care which evaluation point are chosen in order to derive a matrix without singularities. For reasons of compactness the dependency on ω is omitted in eq. (13) and in eq. (16), too. The source strength vector $\mathbf{q}$ was calculated for 33 frequency bins at 6000 Hz sampling frequency. The wish for the simulation of

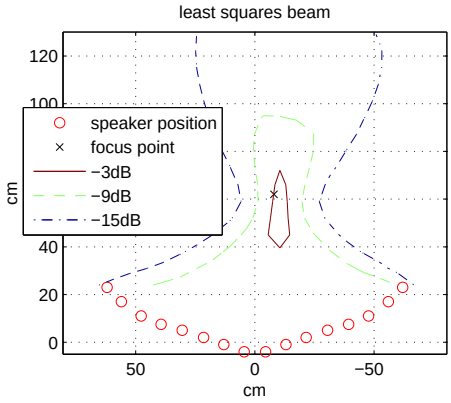

Figure 3: *Sound pressure iso-curves of the least squares beamformer. 64-taps long FIR filters were used to produce the beam.*

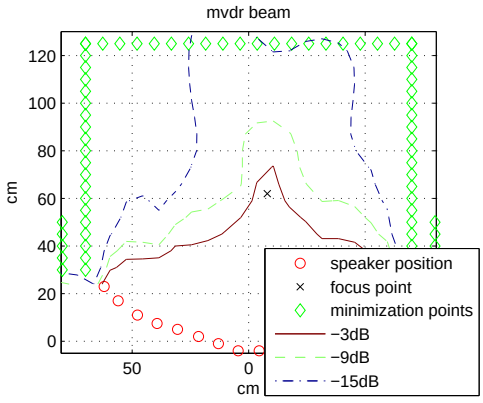

Figure 4: *Sound pressure iso-curves of the minimum variance beam. The beam was produced with 128 taps long FIR filters.*

Fig. 3 was 0 dB in the focus point and minus infinity in all other evaluation points i.e.

$$\mathbf{q} = (\tilde{\mathbf{G}}^H \tilde{\mathbf{G}})^{-1} \tilde{\mathbf{G}}^H \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \end{bmatrix}^T. \qquad (14)$$

A real focus point is reached (i.e. the island of -3 dB) due to the least squares driving functions and the bent shape of the array. However the frequency response of the LSB in the focus point is not distortionless (see also Fig. 5(a)).

3.3. Minimum variance distortionless response beamformer

In contrast to the LSB, the MVDR beamformer minimizes the sound pressure at some chosen field points under the constraint of producing distortionless 0 dB in the focus point, i.e.

$$1 = \mathbf{h}^T \mathbf{q}. \qquad (15)$$

The solution to the quadratic minimization problem is [13]

$$\mathbf{q}^H = \frac{\mathbf{h}^T (\mathbf{G}^H \mathbf{G})^{-1}}{\mathbf{h}^T (\mathbf{G}^H \mathbf{G})^{-1} \mathbf{h}^*}, \qquad (16)$$

The challenge of the MVDR method is to set the right minimization points. First, the resulting matrix $(\mathbf{G}^H \mathbf{G})$ has to be invertible again, and second, areas which are not in the scope of the minimization might be strongly excited by this method. Fig. 4 shows a constellation of minimization point that causes steep sound pressure decay towards the rear end of the evaluated area. Therefore, the sound pressure close to the array is very high. A filter of 128 taps was needed to derive this result.

3.4. Comparison

The two optimization methods (LSB and MVDR) have source strength vectors that depend on the frequency. Therefore, the WNG depends on the frequency, too. Both optimization methods need high weights at low frequencies to reach their optimum. This causes a poor WNG at this frequencies. The relation between the sound pressure in the focus point and the WNG is depicted in Fig. 5. Table 1 compares the different measures of the 3 examined beamforming methods. It can be concluded that both optimization methods have a severe drawback compared to the NFB.

Table 1: Comparison

	NFB	LSB	MVDR
average WNG	15 dB	-2 dB	-19 dB
filter length	1	64	128
SNR	31 dB	34 dB	32 dB

1. They are driven with filter, which in our case were of length 64 and 128.
2. Both cause a big bass boost in order to reach their minimization optimum. This leads to undesired responses in the case of small phase and amplitude variations of amplifiers and loudspeakers [15].

It is therefore more adequate to use the NFB as it can be driven with one complex weight per loudspeaker. The norm of the weights are constant over frequency and so is the response in the focus point. Measurements of the near field beamformer follow to allow an evaluation of the simulations.

3.5. Measurement

The measurements were done with the same spatial resolution, on the same area and with the same arrangement of loudspeakers like the simulations. This allows for a direct comparison with the simulation results. The impulse responses at the microphone points were derived with exponential sweeps [16] of 2 seconds in the given bandwidth. The length of the impulse responses was reduced to 2.9 ms. The gain of the NFB is the same as in the simulation. The SNR is 34 dB and hence 3 dB lower than in the simulation. Still, this difference in sound pressure is far below the masking level [17]. Thus we can provide a proper loud signal at the position of the user without disturbing coworkers in the rest of the room. The differences of the sound field can be seen in Fig. 6. The measured focus point is smaller and the descent towards the rear end is steeper than in the NFB simulation of Fig. 2. This can be explained by the directivity of the loudspeakers. In the simulations, the loudspeakers were assumed to be omnidirectional point sources. In reality however, they radiate more energy into the direction they are facing. Due to the bent shape of the array, more energy is radiated into the focus point of the elliptical segment.

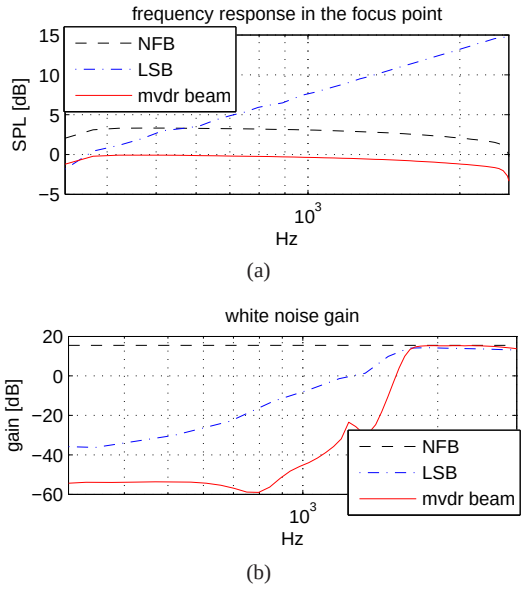

Figure 5: (a) *The LSB does not produce a distortionless response in the focus point.* (b) *Due to their bass boost, the LSB and the MVDR beamformer dramatically loose gain at low frequencies.*

4. TRANSAURAL STEREO

The width of the focus point Δd, is proportional to the signal wavelength and is limited by [18],

$$\Delta d > \frac{\lambda}{2}. \tag{17}$$

In Fig. 7, a virtual head was placed into the measured sound field. It can be seen, that the decay at the position of the contralateral ear is already -9 dB. However, according to the physical limits given in eq. (17), the beam will reach the contralateral ear up to 1700 Hz which is an undesired cross talk. This cross talk was measured with a dummy head and is shown in Fig. 8. Below 500 Hz the channel separation is 5 dB only. Above 1 kHz it increases to 10 dB. The crosstalk can be further reduced by applying a transaural filter matrix as it can be seen in the following.

For the case of 2 loudspeakers the following matrix equation describes the relation between the ear signals $E_l(e^{j\Omega}), E_r(e^{j\Omega})$ and the loudspeaker signals $L_l(e^{j\Omega}), L_r(e^{j\Omega})$

$$\begin{pmatrix} E_l \\ E_r \end{pmatrix} = \begin{pmatrix} T_{ll} & T_{rl} \\ T_{lr} & T_{rr} \end{pmatrix} \begin{pmatrix} L_l \\ L_r \end{pmatrix}, \tag{18}$$

where T_{rl} denotes the transfer function from the right speaker to the left ear a.s.o. The filter matrix that needs to be applied to the speaker signals in order to reach the transaural solution is the inverse of the transfer function matrix

$$\frac{1}{T_{ll}T_{rr} - T_{rl}T_{lr}} \begin{pmatrix} T_{rr} & -T_{rl} \\ -T_{lr} & T_{ll} \end{pmatrix}. \tag{19}$$

We have an array of 16 loudspeakers. However, the equation still holds because we only use 2 outgoing signals. The transfer function matrix is derived over the complex weights $g_{i,j}$ of the beamformer and the HRTFs $H_{i,j}$ from every loudspeaker to both ears

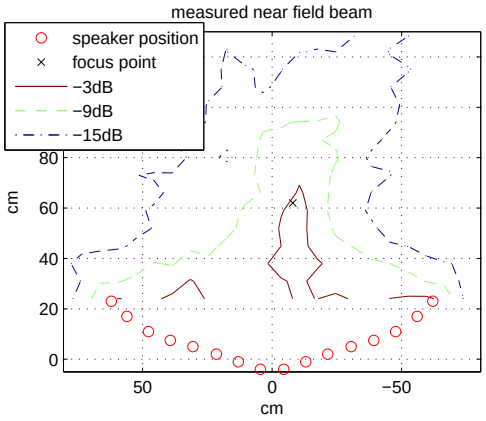

Figure 6: *The focus point of the measured near field beam is smaller and the sound pressure decay towards the rear end is steeper than the simulated near field beam in Fig. 2.*

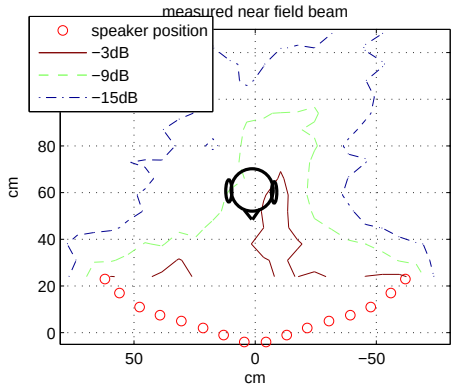

Figure 7: *The sound pressure decay at the position of the contralateral ear is -9 dB.*

(see also Fig. 9)

$$\begin{pmatrix} T_{ll} & T_{rl} \\ T_{lr} & T_{rr} \end{pmatrix} = \begin{pmatrix} H_{1,l} & H_{2,l} & \cdots & H_{16,l} \\ H_{1,r} & H_{2,r} & \cdots & H_{16,r} \end{pmatrix} \begin{pmatrix} g_{l,1} & g_{r,1} \\ g_{l,2} & g_{r,2} \\ \vdots & \vdots \\ g_{l,16} & g_{r,16} \end{pmatrix}. \tag{20}$$

This transfer function matrix has to be adapted and inverted with every movement of the user. Therefore, short transfer functions and the reduction to a 2×2 matrix are essential to save computation power. The FIR filters used for the LSB and the MVDR beam in subsection 3.2 and 3.3 would prolong the transfer functions with 63 taps or 127 taps, respectively. The inversion is done for every frequency bin. Thanks to the limited bandwidth and the reduction to a simple 2×2 matrix, the transfer function matrices are of outstanding smoothness. They can easily be inverted without the need for regularization.

Applying the filter matrix of eq.(19) to the binaural signals leads to a considerable improvement of the crosstalk reduction of constantly 20 dB as it can be seen in Fig. 10.

The filtering is especially important in constellations where it is difficult to produce two distinct beams. Such a constellation is

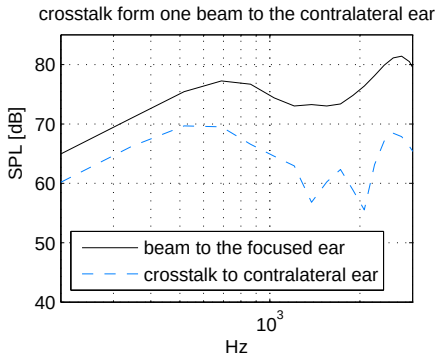

Figure 8: *SPL of the crosstalk over frequency. The beam width decreases with the frequency. Therefore the crosstalk decreases, too.*

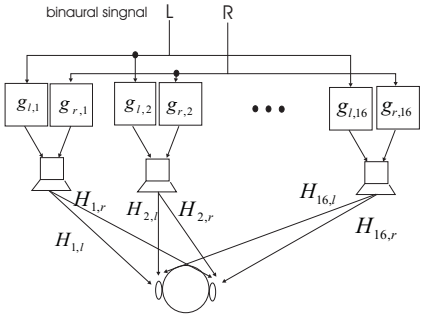

Figure 9: *The binaural signals are delayed and weighted by the complex factors $g_{i,j}$ and reach the ears via the HRTFs $H_{j,i}$. The overall transfer functions are derived by superimposing these weighted HRTFs.*

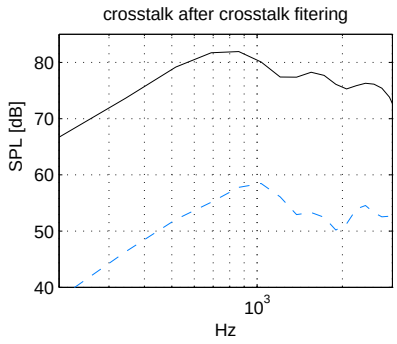

Figure 10: *This figure was derived under the same conditions like Fig. 8 except for applying the filter matrix of eq.* (19)*. A constant crosstalk suppression of 20 dB is reached.*

depicted in Fig. 11. The crosstalk that arises from this constellation is shown in Fig. 12. The crosstalk signal is even stronger than the signal at the focused ear. After filtering however, a channel separation of 15 dB can be gained, as it is shown in Fig. 13. Performance losses only occur at the side ends of the array where the channel separation may decreases to 8 dB for some frequencies. Such a case is depicted in Fig. 14.

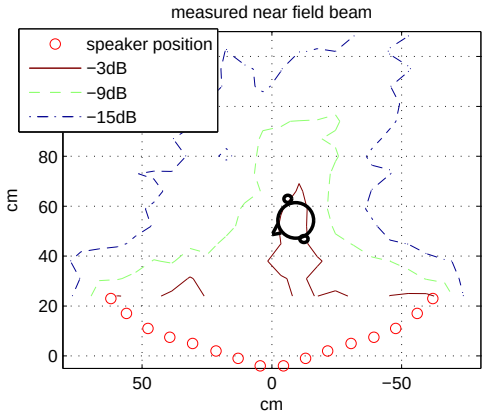

Figure 11: *In a constellation like this, the crosstalk from the focused right ear to the contralateral left ear will be very high, for the contralateral ear is facing the loudspeaker array.*

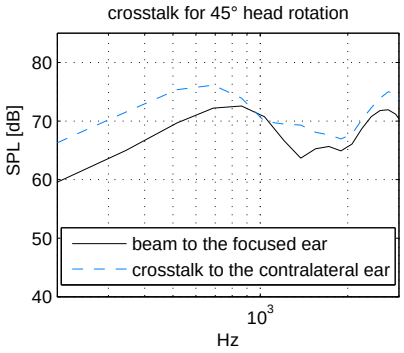

Figure 12: *According to the constellation depicted in Fig. 11: The crosstalk is even higher than the beam at the focused ear itself.*

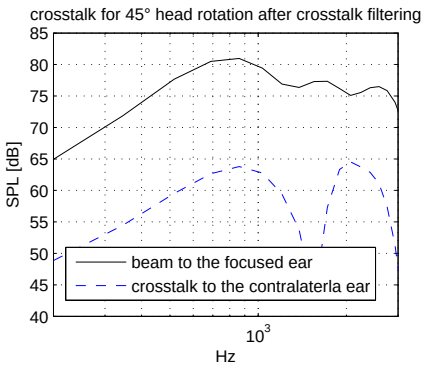

Figure 13: *The transaural filter reduces the crosstalk to 15 dB, even in disadvantageous constellations like depicted in Fig. 11.*

In contrast to [19] where only 2 loudspeakers are used for an adaptive transaural system, our cross talk cancellation works for any head rotations. In its limited bandwidth, it shows better results than [8] for central head positions and allows for a larger radius of movements.

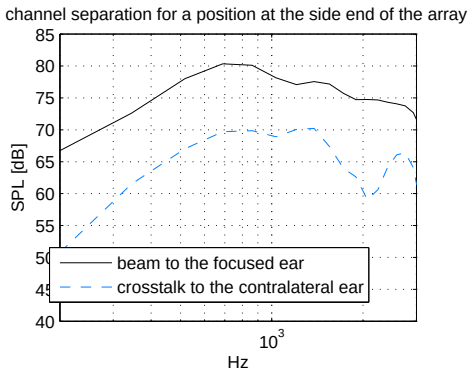

Figure 14: *At the side end of the array, the channel separation after cross talk filtering lies between 17 and 8 dB.*

5. CONCLUSION AND OUTLOOK

We designed a loudspeaker array that steers binaural signals to the ears of a user. Therefore, we simulated and compared a near field beamformer, a least squares beamformer and a MVDR beamformer. The latter two cause big enhancements of low frequencies which makes them unfeasible for standard amplifiers and loudspeakers. A measurement of a near field beam concluded the examination. It proved the focusing qualities and gives reason to our concept of an alternative to headphones that does not exceed the noise floor in the rest of the room. Further simulations could include the directivity of the loudspeakers and aim to find optimization conditions which lead to smaller filter lengths and less bass boost. E.g. recent work has derived robust superdirective microphone beamformers by including a white noise constraint [20].

Future investigations could also examine the abilities of modeling the transaural filters with parametric equalizers. This would further reduce the processing cost dramatically.

6. ACKNOWLEDGMENT

This work was supported in part by Eurocontrol under Research Grant Scheme - Graz, (08-120918-C). We thank Horst Hering and Robert Höldrich for fruitful discussion and Franz Zotter for valuable advice.

7. REFERENCES

[1] J. Blauert, *Spatial Hearing*, MIT Press, Cambridge, MA, USA, revised edition, 1997.

[2] M. Noisternig, A. Sontacchi, T. Musil, and R. Höldrich, "A 3d ambisonics based binaural sound reproduction system," in *24th international AES Conference: Multichannel Audio*, 2003.

[3] J. Bauck and D. H. Cooper, "Generalized transaural stereo and applications," *J. Audio Eng. Soc.*, vol. 44, no. 9, pp. 683–705, 1996.

[4] B. S. Atal and M. R. Schroeder, "Computer simulation of sound transmission in rooms," in *IEEE Conv. Record*, 1963.

[5] W. G. Gardner, "Head tracked 3-d audio using loudspeakers," in *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, New York, USA, Oct 19 - 22 1997.

[6] T. Lentz, "Dynamic crosstalk cancellation for binaural synthesis in virtual reality environments," *J. Audio Eng. Soc.*, vol. 54, no. 4, pp. 283–295, 2006.

[7] M. R. Bai, C.-W. Tung, and C.-C. Lee, "Optimal design of loudspeaker arrays for robust cross-talk canellation using the taguchi method and the genetic algorithm," *J. Audio Eng. Soc.*, vol. 117, no. 5, pp. 2802–2813, May 2005.

[8] K. Laumann, G. Theile, and H. Fastl, "A virtual headphone based on wave field synthesis," in *Acoustics 08 Paris*, Paris, France, June 29 - July 4, 2008.

[9] A. Sontacchi, M. Guldenschuh, Th. Musil, and F. Zotter, "Demonstrator for controllable focused sound source reproduction," in *Eurocontrol INO Workshop*, Paris, France, November 2008.

[10] E. Williams, *Fourier Acoustics*, Academic Press, London, 1998.

[11] D. B. Ward, R. A. Kennedy, and R. C. Williamson, *Microphone Arrays*, chapter Constant Directivity Beamforming, pp. 3–17, M. Brandstein and D. Ward, Springer-Verlag, Berlin Heidlberg New York, 2001.

[12] W. Ahnert and F. Steffen, *Beschallungstechnik: Grundlagen und Praxis*, Hirzel, Stuttgart, D.

[13] J. Bitzer and U. Simmer, *Micorphone Arrays*, chapter Superdirective Microphone Arrays, pp. 19–38, M. Brandstein and D. Ward, Springer-Verlag, Berlin Heidlberg New York, 2001.

[14] T.-C. Liu and B. Van Veen, "A modular structure for implementation of linearly constrained minimum variance beamformers," *IEEE Transactions on Signal Processing*, vol. 39, no. 10, pp. 2343–2346, 1991.

[15] E. Mabande and W. Kellermann, "Towards superdirective beamforming with loudspeaker arrays," in *Conf. Rec. International Congress on Acoustics*, Madrid, Spain, September 2007.

[16] A. Farina, "Simultaneous measurements of impulse response and distortion with a swept-sine technique," in *AES 108th convention*, Paris, France, February 2000.

[17] B. C. Moore, *Hearing*, chapter Frequency Anlaysis and Masking, pp. 161–206, Academic Press, San Diego, USA / London, UK, 1995.

[18] S. Yon, M. Tanter, and M. Fink, "Sound focusing in rooms: The time-reversal approach," *J. Audio Eng. Soc.*, vol. 113, no. 3, pp. 1533–1543, 2003.

[19] T. Lentz and O. Schmitz, "Realization of an adaptive cross-talk cancellation system for a moving listener," in *21st AES Conference*, St. Peterburg, Russia, June 1-3, 2002.

[20] E. Mabande, A. Schad, and W. Kellerman, "Robust superdirectional beamforming for hands-free speech capture in cars," in *NAG/DAGA 2009*, Rotterdam, NL, March 23 - 26, 2009.