Application of a transaural focused sound reproduction

Markus Guldenschuh and Alois Sontacchi
Institute of Electronic Music and Acoustics
University of Music and Performing Arts Graz
8010 Graz, Austria
www.iem.at

***Abstract*—In a reference paper, the authors suggested a new sophisticated communication tool for air traffic controllers. The communication tool should consist of a loudspeaker array that focuses spatialized sound. Spatialization of the incoming signals helps to better distinguish between the different communication partners (i.e. the pilot and other co-controllers), and it is achieved with the technique of Transaural Stereo. Sound focusing avoids strong room excitation and is accomplished with a weighted Delay & Sum Beamformer. This paper provides measurement results that proof the concept of the outlined method. In particular, the focusing properties of the array are proofed and the preconditions for a correct rendering of the spatialized signals are investigated.**

***Index Terms*—Transaural stereo, loudspeaker beamforming, weighted delay & sum beamformer, cross talk cancellation.**

I. Introduction

In air traffic control (ATC), the airspace is divided into several sectors. Each sector is controlled by two controllers. Both controllers use headsets as communication interfaces. This might be unpleasant and therefore lower the concentration, especially in long time use. Desktop integrated loudspeaker and microphone arrays as alternatives to head mounted communication interfaces are presented in [1] and [2]. The loudspeaker array is used to produce a controllable focused sound. The position of the controller is steadily tracked such that the sound source can be focused onto his ears. The presented beamforming approach causes two distinct sound spots which also bear the possibility to share communication information like schematized in Fig.1. An underlying binaural encoding of the communication signal further allows spatial augmentation for various distributed communication partners, even placed in a party line.

In this paper, we will present an application of the concepts given in the reference papers [1] and [2]. The paper will include measurement results and discussions on the limitations and capabilities of the application.

This work was supported in part by the Eurocontrol under Research Grant Scheme - Graz, (08-120918-C).

Contact to the authors via e-mail: guldenschuh@iem.at, sontacchi@iem.at.

Figure 1. Two distinct sound bubbles that bear the possibility to share communication information.

II. Concept of transaural stereo with a loudspeaker array

ATC communication uses a bandwidth of only 300 to 2500 Hz. The stereo channels of the ground controllers are split into a *pilot channel* for communication with the pilots and a *controller channel* for communication with controllers from neighboring sectors. This brings an auditive separation between pilots and other controller, which however is not very natural and probably annoying. A natural sound spatialization and separation can be achieved if sound sources are binaurally encoded [3], [4]. For binaural stereo, each transmission channel is directly transmitted to the left and right ear; and any cross talk has to be omitted. Therefore binaural stereo is generally presented via headphones. If loudspeakers are to be used, as suggested in the reference papers, cross talk cancellation has to be applied as in [5]. Fig.2 sketches the concept of the transaural beamformer. The binaural Ambisonics system, introduced in [6], renders arbitrary many sound sources s at position $\mathbf{r}$ as a 2-channel binaural signal. The binaural signal is applied to a 2×2 cross talk cancellation matrix before it reaches the beamforming stage.

In the following chapters the beamforming stage and the cross talk canceler will be investigated in detail. The binaural Ambisonics system is not in the scope of this paper, as it is well

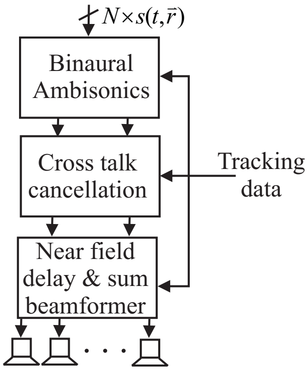


Figure 2. Arbitrary many sound sources can be rendered to a 2-channel binaural signal. The binaural signal runs through a cross talk cancellation filter before it is led to the beamformer.

described in [6]. More on different sound focusing methods and transaural stereo can be read in [7].

III. Near field delay & sum beamformer

The near field delay & sum beamformer (NFDSB) weights the binaural signals with complex weights $g(\omega)$ as shown in Fig.3. The absolute values of the weights are the reciprocal of the distance from the speakers to the focus point. The phase of $g(\omega)$ compensates for the delay of the loudspeaker signal to the focus point. The delay is the same for all frequencies ω, therefore, the NFDSB can easily be realized with a variable delay line and a simple multiplication per channel. This makes the NFDSB very feasible and processing efficient, especially compared to super directive beamformers like it is shown in [8]. We developed a beamformer with 16 loudspeakers, arranged in

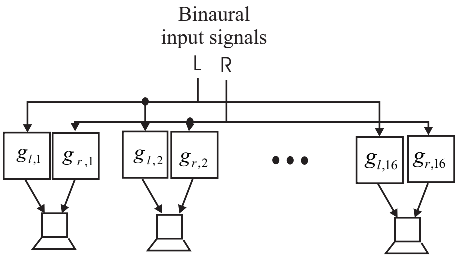


Figure 3. The complex weights $g(\omega)$ can be realized by a delay line and a simple multipicator.

a segment of an ellipse. The loudspeakers can be mounted over (or under) the screens of the air traffic controller, like in Fig.4 for example.

We also developed a simulation tool that allows us to predict the sound field of the beamformer for arbitrary loudspeaker weights $g(\omega)$. Additionally, measurements were accomplished to compare the simulated beams with beams of a real loudspeaker array. The comparison was done for 4 focus positions (marked in Fig.15a) in 12 bark bands in the 384 evaluation points on an area of 112×168 cm. The sound pressure in the evaluation points was measured with a microphone array with a raster of 7×7 cm. This raster prevents spatial aliasing

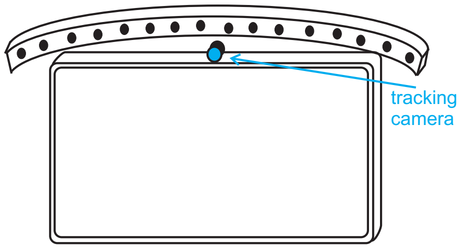
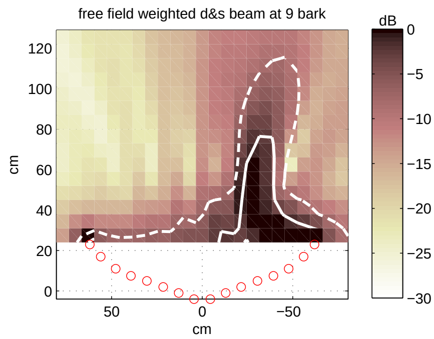


Figure 4. Loudspeaker array over the air traffic control screen with a USB camera for the user tracking.

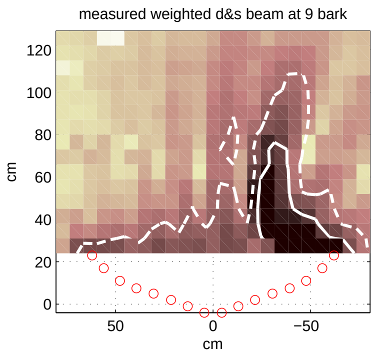
up to 2500 Hz. Bark bands simulate the frequency dependent sensitivity of the ear and are therefore also called critical bands. Their bandwidths were empirically determined by [9]. A reference measurement of all loudspeakers in one meter distance was used to equalize the sound field measurement. In the frequency domain, the equalization can be done by inverse filtering

$$H_{\text{measure.eq}} = H_{\text{measure}} H_{\text{ref}}^{-1}. \tag{1}$$

The comparison will exemplarily be shown at 9 bark for the close side position. The sound pressure level (SPL) distributions are shown in Fig.5, and the absolute value of their differences in Fig.6a. Errors outside of the focused area are less relevant,



(a) Simulated beam



(b) Measured beam

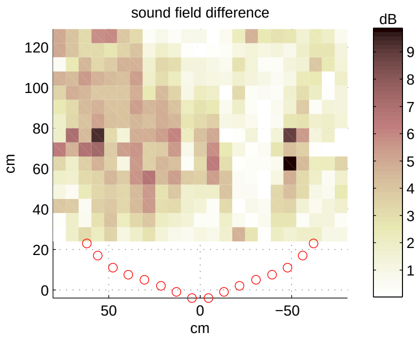
Figure 5. Comparison between the simulated and the measured beam in the close side position in the 9$^{\text{th}}$ bark band. The solid white line represents the -3 dB level line and the dashed line the -9 dB level line.

which is why we introduce a weighted difference e_w, too. The weights are the square root of the simulated sound pressure

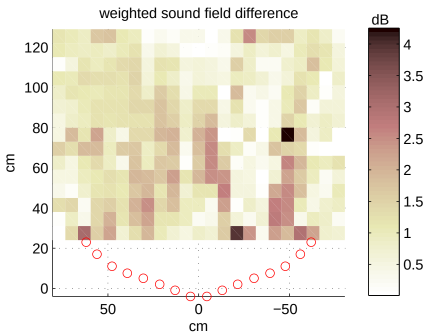
p_{sim} normed by the pressure in the focus point.

$$e_{w,i} = |L_{m,i} - L_{\text{sim,i}}| \sqrt{\frac{p_{\text{sim,i}}}{p_{\text{focus}}}}, \tag{2}$$

where $L_{m,i}$ is the measured SPL in evaluation point i, and L_{sim} the SPL of the simulated sound field, respectively. As a consequence the error in the focus point stays the same, while errors in regions of low SPL are compressed. The weighted difference is shown in Fig.6b



(a) SPL differences in the close side position at 9 bark. The highest difference is 9 dB.



(b) SPL differences, weighted with the square root of p_{sim}. Differences in regions of low SPL (like in the upper left corner) are suppressed. The highest difference is 4 dB.

Figure 6. SPL differences and weighted SPL differences between the measured and the simulated sound field in the 9^{th} bark band.

The median and the mean value over all SPL differences and weighted differences are listed in Fig.15 at the end of this paper. Mean and median of the weighted difference lie under 2 dB, which is a very satisfying result. Standard deviations, higher than 4 dB only occur for the unweighted difference and hence in regions of low SPL and minor interest. The measurement results give reason to the simulation tool and justify its ongoing usage.

With the simulation tool, we also calculated the 3D directivity of the array. The directivity diagram for a central focus point can be seen in Fig.7. With the help of the directivity the SPL L_r of an excited reverberant room can be calculated [10].

$$L_r = 10 \log \frac{P_{ac}}{P_0} - 10 \log A + 6\,\text{dB}, \tag{3}$$

where A is the sum of reflecting surfaces, P_0 is the reference sound power of 10^{-12} W and P_{ac} is the acoustic power of the

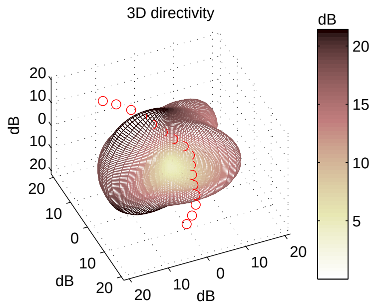


Figure 7. The array has a strong directionality towards the focus point due to its bent shape.

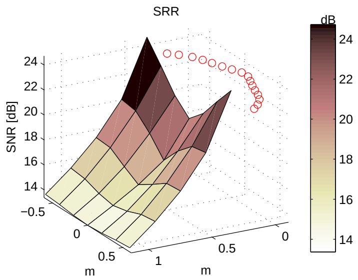


Figure 8. SRR in 35 focus points. The SRR decreases with the distance of the focus point from the array. This means that the room excitation in relation to the SPL in the focus is higher if the beam has to be steered further away.

array. P_{ac} is the average sound intensity on a surface of a sphere with radius r:

$$P_{ac} = \frac{1}{N} \sum_{n=1}^{N} p_n^2 \frac{4\pi r^2}{\rho c}, \tag{4}$$

where ρc is the acoustic impedance which has 408 $\frac{\text{kg}}{\text{m}^2\text{s}}$ at 20 °C. We define the difference between the SPL L_{focus} in the focus point and L_r as spatial rejection ration (SRR):

$$\text{SRR} = L_{focus} - L_r \tag{5}$$

The SRR for 35 focus positions is depicted in Fig.8.

Fig.9 shows the measured broad-band sound field of our beamformer. Two effects can be observed. First, the beam decays very steep toward the sides and the rear end of the sound field. As a consequence, the room will only be excited marginally with sound energy. Second, the beam already produces a desired cross talk reduction. This cross talk reduction can be seen in Fig.10.

A further cross talk reduction can be achieved by the usage of HRTF filters, as it can be read in the following section.

IV. Cross Talk Cancellation

The binaural signals L and R are split into 2×16 loudspeaker signals in the beamforming stage. The loudspeaker weights are indicated with g_{ji}. These 2×16 loudspeakers signals reach the ears over 16×2 head related transfer functions H_{ij}

$$\begin{pmatrix} E_l \\ E_r \end{pmatrix} = \begin{pmatrix} H_{1,l} & H_{2,l} & \cdots & H_{16,l} \\ H_{1,r} & H_{2,r} & \cdots & H_{16,r} \end{pmatrix} \begin{pmatrix} g_{l,1} & g_{r,1} \\ g_{l,2} & g_{r,2} \\ \vdots & \vdots \\ g_{l,16} & g_{r,16} \end{pmatrix} \begin{pmatrix} L \\ R \end{pmatrix}. \tag{6}$$

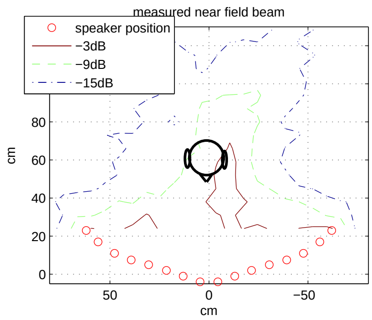


Figure 9. The broad-band (300-2500 Hz) sound pressure attenuation of a measured beam is indicated through three level lines. A head was drawn into the sound pressured distribution to indicate that the beam already causes a notable cross talk reduction at the contralateral. The cross talk over frequency can be seen in Fig.10.

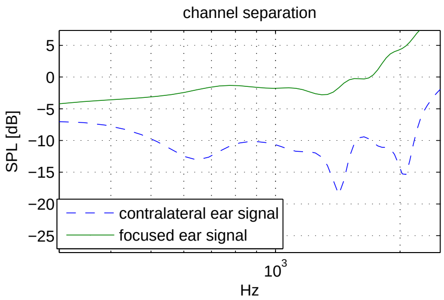


Figure 10. Cross talk reduction due to beamforming, measured with a dummy head. The beam width decreases with the frequency [2]. Therefore, the cross talk decreases, too.

Let us define the product of the beamforming matrix of g_{ji} and the HRTF matrix as transfer function matrix $\mathbf{T}$. Then

$$\begin{pmatrix} E_l \\ E_r \end{pmatrix} = \begin{pmatrix} T_{ll} & T_{rl} \\ T_{lr} & T_{rr} \end{pmatrix} \begin{pmatrix} L \\ R \end{pmatrix}. \qquad (7)$$

The binaural signals L and R shall reach the ears without being changed. In other words $\left(\begin{smallmatrix} E_l \\ E_r \end{smallmatrix}\right)$ shall equal $\left(\begin{smallmatrix} L \\ R \end{smallmatrix}\right)$. This is achieved by introducing a filter matrix $\mathbf{C}$, which is the inverse of $\mathbf{T}$, such that $\mathbf{TC} = \mathbf{E}$ where $\mathbf{E}$ is the unit matrix.

$$\left(\begin{smallmatrix} E_l \\ E_r \end{smallmatrix}\right) = \mathbf{TC} \left(\begin{smallmatrix} L \\ R \end{smallmatrix}\right). \qquad (8)$$

The cross talk canceler (XTC) not only aims to cancel the cross talk, it also equalizes the spectrum of the focused ear signal, because the binaural signals should reach the ears without being altered. Both effects of the cross talk canceler can be seen in Fig.11. The channel separation raises to at least 15 dB and the ripples of the focused ear spectrum are smaller than 3 dB.

The channel separation due to the beamformer works well for central user positions and 0^{circ} head rotation. For a constellation like depicted in Fig.12, the beam is very strong at the unfocused ear, too. The cross talk canceler however, causes a channel separation of 15 dB for this constellation, too. Fig.13 shows the ear signals of the beamformer without XTC and after XTC.

We measured the cross talk for 4 head positions with 0^{circ} and 30^{circ} head rotation each. The measurement positions are

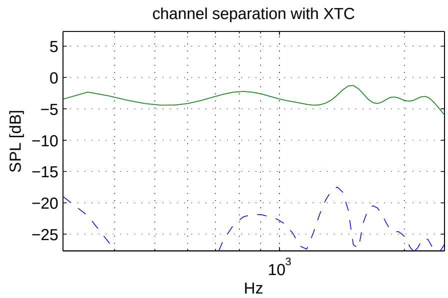


Figure 11. The channel separation with cross talk cancellation is much higher. Especially for low frequencies.

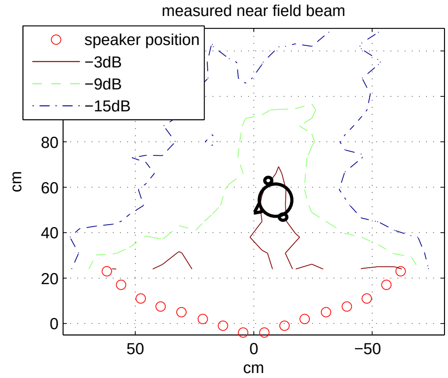


Figure 12. The rear right ear is focused, but the left ear is closer at the loudspeaker array. That is the reason why the beamformer does not cause any cross talk reduction for this constellation.

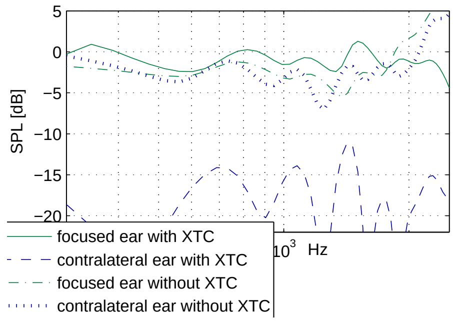


Figure 13. The cross talk canceler suppresses the unfocused ear signal more than 15 dB, even in constellation where the beamformer cannot contribute to the channel separation.

the same as in Fig.15a. The channel separation is at least 10 dB and the ripple in the spectrum of the focused ear is 5 dB in the worst case. These results are very satisfying, however, they vary with the head position and rotation. A quality measure is introduced to judge the dependency on these positions and rotations. The quality measure Q shall consist of the channel separation and the ripple in the frequency response. The channel separation $\mathrm{SPL_{dif}}$ is simply given by the average amplitude difference in dB. The ripple $\mathrm{SPL_{var}}$ will be defined as the variance of the amplitude over frequency. The average channel separation is 14 dB in the worst case and 22 dB in the best case.

The variance varies between 1 and 3 dB^2. To get a equal range, $\mathrm{SPL_{dif}}$ is divided by 2 and $\mathrm{SPL_{var}}$ is multiplied with 2, such that

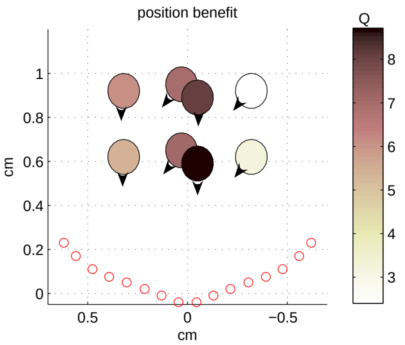
$$Q = \frac{1}{2}\mathrm{SPL_{dif}} - 2\ \mathrm{SPL_{var}}. \tag{9}$$

The results of Q can be seen in Fig.14a. The central positions have the best XTC conditions, while the side positions with 30^{circ} rotation have the worst.

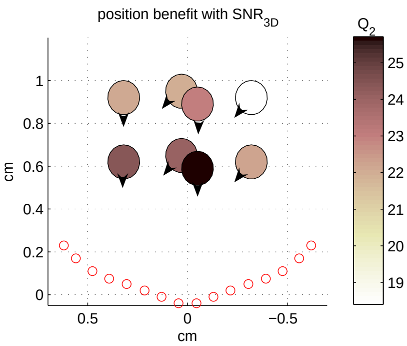
Until now, only the correct binaural perception has been considered for the quality factor Q. The excitation of the room, however, should also be taken into account. The SSR of the NFDSB varies between 15 and 19 dB for the given head positions and has therefore the same rage as the two already considered (scaled) properties. The quality measure Q_2 includes the room excitation and is defined as

$$Q_2 = \frac{1}{2}\mathrm{SPL_{dif}} - 2\ \mathrm{SPL_{var}} + \mathrm{SSR}. \tag{10}$$

The results of Q_2 are shown in Fig.14b. It can be concluded that the quality of the transaural beamformer decreases with the distance from the array, with the degree of head rotation and with the distance from the symmetry axis.



(a) XTC quality. Central positions and 0^{circ} head rotations have the best preconditions for a correct binaural perception.



(b) Transaural beamforming quality. The room excitation is included in the measure. Beams to the close side positions cause little room excitation (see also Fig.8), that is why the close side head positions also perform better in terms of the transaural beamforming quality Q_2.

Figure 14. Evaluated head positions and rotations. Please note that the head positions on the symmetry axis are coincident for both rotations. The are only displaced for the representation. The quality of the transaural beamformer decreases with the distance from the array, with the degree of head rotation and with the distance form the symmetry axis.

V. Conclusion and Outlook

Following the reference papers [1] and [2], we developed a transaural beamformer for the usage in air traffic control. The application is dynamic, since the user position is steadily tracked with a camera. Measurements proofed the concept of the reference papers and show satisfying results in terms of room excitation and cross talk cancellation.

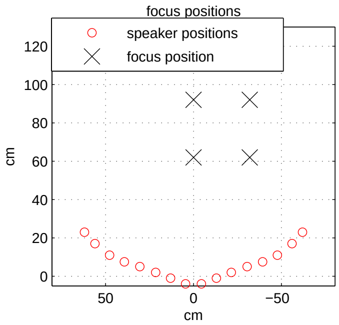
A study on the comfort of headphone free telephony would be very meaningful to evaluate the benefits of the transaural beamformer. Future work could also examine the possibilities to model the cross talk cancellation filter with parametric equalizers. This would avoid matrix inversions and therefore further reduce the processing costs.

Acknowledgment

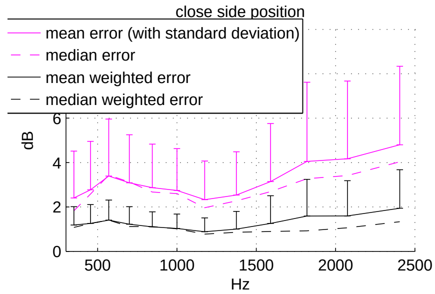
The authors would like to thank Horst Hering for his description of air traffic controller's workstations and work conditions.

References

[1] A. Sontacchi, M. Guldenschuh, T. Musil, and F. Zotter, "Demonstrator for controllable focused sound source reproduction," in *7th Eurocontrol INO Workshop*, 2008.
[2] M. Guldenschuh, A. Sontacchi, F. Zotter, and R. Höldrich, "Principles and considerations to controllable focused sound source reproduction," in *7th Eurocontrol INO Workshop*, 2008.
[3] J. Blauert, *Spatial hearing; The psychophysics of human sound localization.* Cambridge, Mass. [u.a.]: MIT Pr., 1999; 1999.
[4] F. Baumgarte and C. Faller, "Binaural cue coding-part i: psychoacoustic fundamentals and design principles," *Speech and Audio Processing, IEEE Transactions on*, vol. 11, no. 6, pp. 509–519, 2003.
[5] J. Bauck and D. H. Cooper, "Generalized transaural stereo and applications," *Journal of the Audio Engineering Society.*, vol. 44, no. 9, p. 683, 1996, iD: 89400647.
[6] M. Noisternig, A. Sontacchi, T. Musil, and R. Höldrich, "A 3d ambisonics based binaural sound reproduction system," in *24th international AES Conference: Multichannel Audio*, Banff, Canada, June 2003.
[7] M. Guldenschuh, "Transaural beamforming," Master's thesis, Graz University of Technology, Graz, Austria, 2009.
[8] M. Guldenschuh and A. Sontacchi, "Transaural stereo in a beamforming approach," in *DAFx-09*, September 2009.
[9] E. Zwicker, *Psychoacoustics; Facts and models*, Berlin [u.a.], 1990.
[10] W. Ahnert, *Beschallungstechnik; Grundlagen und Praxis*, Stuttgart [u.a.], 1993.



(a) The comparison between the simulated and the measured beams was done for the marked 4 positions. As we assume symmetry of the sound field, all positions are chosen to be on the negative side of the x-axis.



(b) Error (i.e. the SPL differences between the measured and the simulated beams) over frequency at the four focus positions.

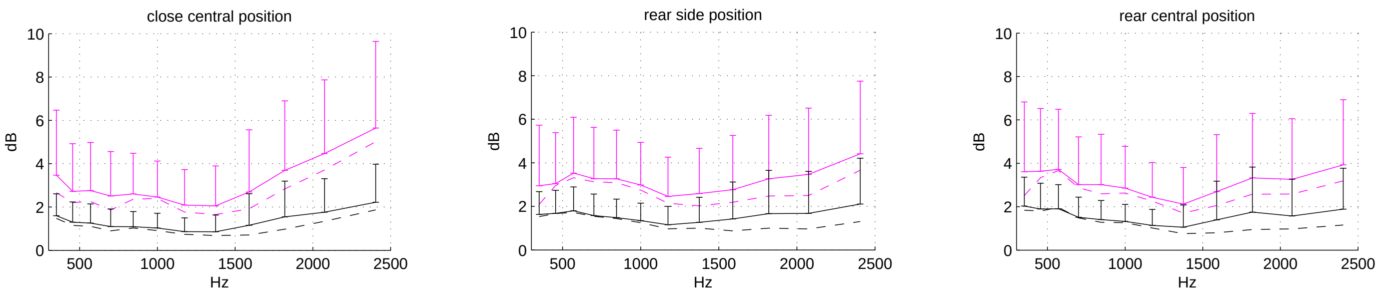


Figure 15. The errors are independent of the focus position. However they increase with the frequency, where the areas of constructive and destructive superposition are smaller. A little phase error can then cause a constructive superposition at a location where the simulation predicts a destructive superposition, or vice versa.