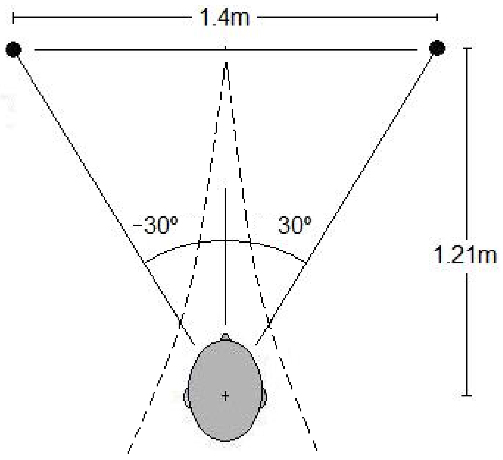
Stereophonic Sound Reproduction for Multiple Listeners

J. Rodriguez Institute of Sound and Vibration Research, University of Southampton, UK
Consejo Nacional de Ciencia y Tecnología, México D.F, México
K.R. Holland Institute of Sound and Vibration Research, University of Southampton, UK

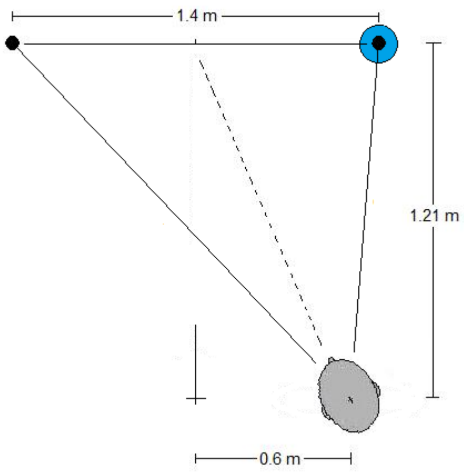
1 INTRODUCTION

Reproduced sound through a stereophonic system can produce the illusion of a sound stage in front of the listener. The necessary localization cues for the spatial illusion are created by using a pair of spaced loudspeakers in accordance with either recorded or adjusted audio signals over a certain frequency range. A major problem of this system is that the mentioned spatial illusion can only be created over a small space symmetrically located in the axis between the loudspeakers called the 'sweet spot' as seen in figure 1a. If the listener moves aside from the sweet spot (asymmetric position), the interaural time difference and interaural level difference (ITD, ILD localization cues), change dramatically and along with the so-called precedence effect, the spatial perception collapses to the nearest of the two loudspeakers, see figure 1b. Therefore it can be said that purist stereophonic system works only for one listener at one specific location. But would it be possible to produce the stereo spatial perception for two listeners simultaneously by using only the original arrangement (two loudspeakers)? In other words, the possibility of enlarging the sweet spot over a broader area including two listeners.

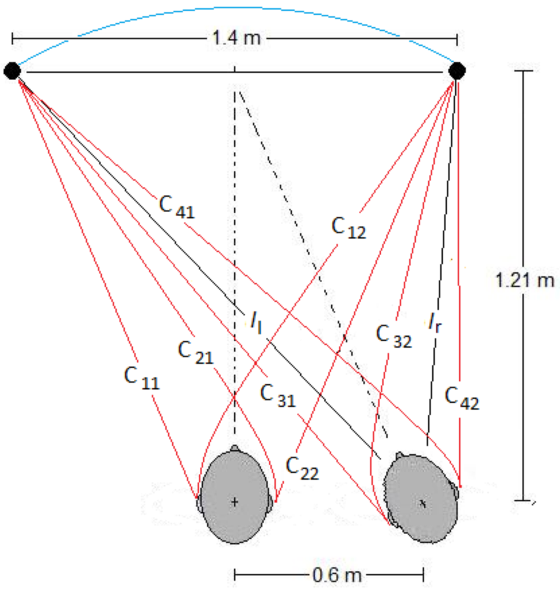
Several works [1, 2] have attempted to overcome this task by means of adjusting the loudspeaker directivity yielding to not really successful results. Furthermore, other systems such as Ambisonics and Wave Field Synthesis try to reproduce a sound field over a larger area. The disadvantage of these is the substantial use of a larger number of loudspeakers which is not practical in a common living room along with complicated recording techniques. This paper describes the viability of finding a signal processing stage by means of an optimization technique (least squares approach) which can be convolved with any stereo signal producing communal spatial perception for two listeners simultaneously. An auditory localization model will be used to assess the resultant perceptions.



(a)



(b)



(c)

Figure 1. (a) Symmetric listener (on-axis), sweet spot area (----). (b) Asymmetric listener (off-axis) showing localization of phantom sources. (c) Geometry used through this work, showing paths for both listeners (plant matrix HRTFs), also showing the distances from left and right sources to the off-axis listener's center head and the main stereo perception of phantom sources (between sources).

2 AUDITORY LOCALIZATION MODEL

The auditory localization model is based on the comparison by the nearest neighbor technique between the so-called characteristic curve and a given target point. By extracting the ITDs and ILDs from the given head related transfer functions (HRTFs) of a set of real sources surrounding a listener [3], a curve (characteristic curve) is built by taking all the ITD-ILD pairs (τ-α space) for each corresponding source angle. As ITD and ILD have different units, they must be remapped into a non-dimensional space (τ'-α') by applying the conversion factors κ_τ and κ_α respectively. The final estimate is then the corresponding angle of the characteristic pair where its distance to a target point (τ'_{tg}-α'_{tg}) is the lowest, in other words, the above mentioned nearest neighbor technique described in Equations 1 and 2. This model takes into account the front and back confusion phenomenon and the mentioned collapse of the spatial perception of an asymmetric listener by means of the interaural phase difference IPD.

One major disadvantage of this model is its lack of integration over frequency. Research on the definition of the critic frequency between waveform (low frequencies) and envelope (high frequencies) ITD is not well defined making it difficult to combine both ranges in the model. This work will use the waveform range, up to 1500 Hz. Although there are other models such as the Pattern Matching auditory model which introduce a frequency weighting function in order to give a single estimate over a broadband input signal, the current model was chosen due to its correspondence with the used optimization method further analyzed in section 3. Both, the characteristic curve and the PM auditory model are fully described in [4].

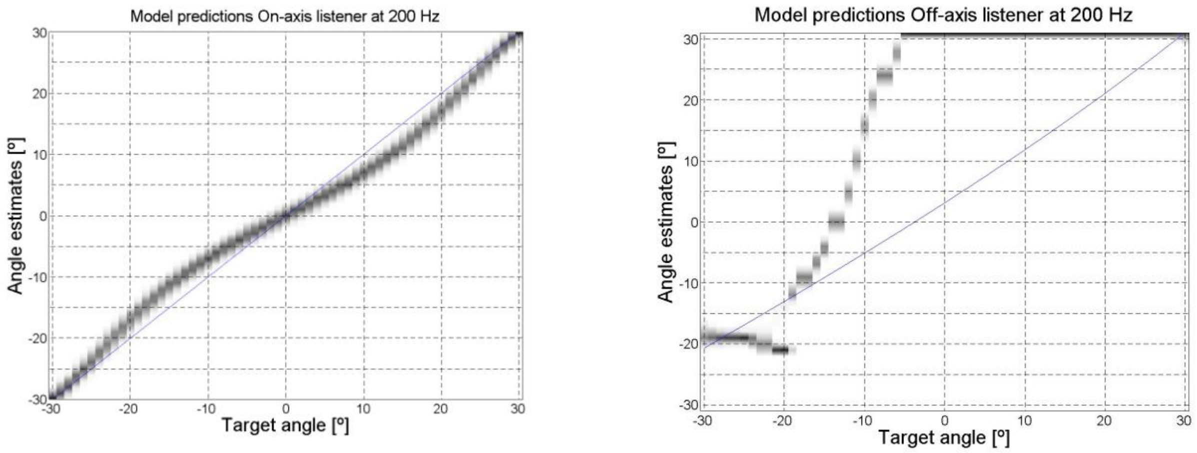
$$e(\theta, f) = \sqrt{\left(\tau'_{tg} - \tau'(\theta, f)\right)^2 + \left(\alpha'_{tg} - \alpha'(\theta, f)\right)^2} \qquad (1)$$

$$\theta_p(f) = \arg\min_{\theta} e(\theta, f) \tag{2}$$

2.1 AUDITORY MODEL WITH STEREO FOR ON- AND OFF-AXIS LISTENERS

Using the explained auditory localization model along with the amplitude panning stereo mechanism, the on and off-axis listeners' perceptions were estimated using the geometry of figures 1c. Target angle against angle estimates is then shown in figure 2a and 2b for a frequency of 200 Hz. In order to include internal noise in the ear and error measurement, not only one but a set of 500 target points (τ'_{tg}-α'_{tg}) were computed by independent zero-mean Gaussian random processes for ITD and ILD with standard deviation of σ_τ = 10 μs and σ_α = 1 dB respectively. Thus the model estimates are shown as the probability of perceiving a particular angle. The darker the estimate, the higher the probability of that angle perception, this noise inclusion also takes into account the perceived diffuseness of a sound source. It has to be pointed out that the reference for both listeners of the spatial illusion area will be between loudspeakers, where stereo mainly works. It is seen that the angles estimation of the on-axis listener are not exactly over the real sources' location. This reflects the amplitude panning stereo mechanism, more specifically the constant power panning. The perception at the real sources location (-30º and 30º) is stronger than the perceived in between them.

Furthermore, the phantom sources are compressed to the center but the largest error is less than 5º as seen around ±15º as target angles. For the case of the off-axis listener, it is clearly seen that the deviation of the perception dramatically goes toward the right source as expected. Figure 3 shows the perception of the central phantom image across frequency without noise addition for the same geometry (figure 1c). It is seen that for the on-axis listener remains constant at 0º but for the off-axis listener some frequency ranges totally collapses to the nearest source while others appear to be better conditioned. Thus, bad conditioned frequency ranges appear to have great impact in the final estimate and special attention must be bared around them for the signal processing stage.



(a)

(b)

Figure 2. (a) Model predictions of the on-axis listener from amplitude panning stereo mechanism at 200 Hz. The darker the higher probability of the angle estimate. Also shows the real sources location (continuous line). (b) Model predictions of the off-axis listener from amplitude panning stereo mechanism at 200 Hz. The darker the higher probability of the angle estimate. Also shows the real sources location (continuous line).

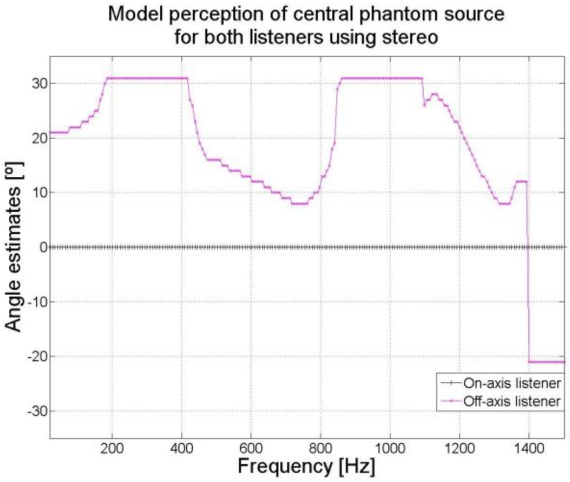
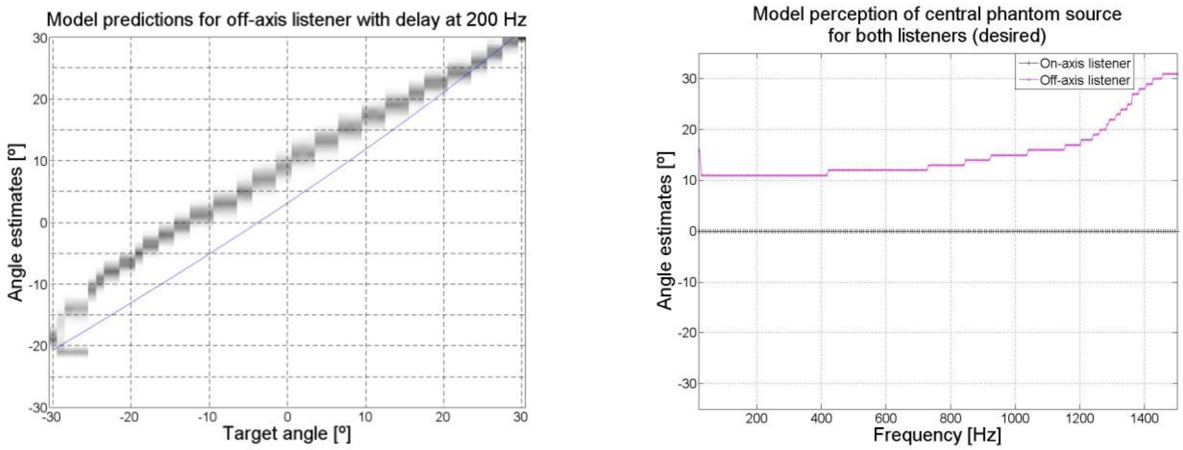


Figure 3. Model predictions of the on- off-axis listeners for a central phantom source using the amplitude panning stereo mechanism across frequency.

2.2 CENTRAL PHANTOM SOURCE DESIRABILITY

The coincidence between listeners' spatial perception at all angles between the loudspeakers using stereo is a very difficult task to achieve. In this paper, the work was done towards the coincidence of only a central phantom source assuming that real sources perceptions are already matched. The work done in [5], adjusted the amplitude and delay of the stereo signals adaptively according to a single listener position in respect to the arrangement. It showed that the perception of a central source is improved by making the signals from the loudspeakers to arrive at the same time and with the same amplitude at the off-axis listener's center head. With this low frequency approach, it effectively released the listener from a fixed symmetric position.

This paper will use the same approach in order to create the off-axis listeners' central perception but with the difference that only a delay will be applied to the nearest source which in this case is the right loudspeaker and no attenuation will be used. For the same frequency, figure 4a shows a considerable difference about the perception of the off-axis listener for central phantom source by applying the corresponding delay. While for the central phantom image is perceived at the right loudspeaker location in figure 2b, the applied delay change it around 9º. Thus it can be said that this approach is a sensible way to build a "desired" perception for the off-axis listener which will be used in the signal processing stage in the next section. Finally figure 4b show the central phantom source perception for both listeners across frequency. As seen the delay approach works mainly at low frequencies.



(a) (b)

Figure 4. (a) Model predictions of the off-axis listener with the corresponding time delay in order to perceive a central phantom source at 200 Hz. The darker the higher probability of the angle estimate. (b) Model predictions of the desired angle estimates for both listeners across frequency (with and without delay for on- and off-axis listeners respectively.

3 DESIGNING THE SIGNAL PROCESSING STAGE

The signal processing stage must be designed so by using only two signals to create same central perception for both listeners. In other words, four control points must be satisfied with only two signals. Thus it is an underdetermined system and does not have an exact solution, it is impossible to fit exactly the required pressures at the listeners' ears. Furthermore as proved above, the off-axis listener cannot have the same perception as on-axis listener even with the introduction of an adequate delay. For this reason, an optimization method must be used dealing with a trade-off between listeners' central phantom source perception simultaneously. This work used the so-called least squares optimization method as it seems to be appropriate to approximate the best possible solution to the problem.

3.1 LEAST SQUARES METHOD

The block diagram for obtaining the required signal processing stage can be seen in figure 5. The least squares approach constructs a cost function curve based on the sum of the squares of the error vector **e** which can be expressed as J. The error vector is formed by the subtraction of the desired signals vector **d** with the reproduced signals vector **w**. In turn, vector **d** is formed by the stereo signals **s** pass through the target matrix **T**. Meanwhile **w** is formed by the stereo signals passed through the required signal processing stage (matrix of filters) **H** and the plant matrix **C**. After some algebra which is fully described in [6] the least squares solution is given by.

$$\mathbf{H} = \left[\mathbf{C}^{H}\mathbf{C}\right]^{-1}\mathbf{C}^{H}\mathbf{T} \qquad (3)$$

The plant matrix **C** is a 4x2 matrix composed by the HRTFs of the listeners due to the geometry presented in figure 1c. The target matrix **T** which is also a 4x2 matrix was formed by adjusting HRTFs of the listeners. More specifically these HRTFs will contain the corresponding delay applied to the off-axis listener as explained in the last section, while for the on-axis listener the same amount of delay is going to be applied to both sources in order to keep causality in the filters, see Equations 4-6. This optimization method is a frequency based method in which the required filters will be computed frequency by frequency called Fast deconvolution method and corresponds with the used auditory model to assess the perceptions.

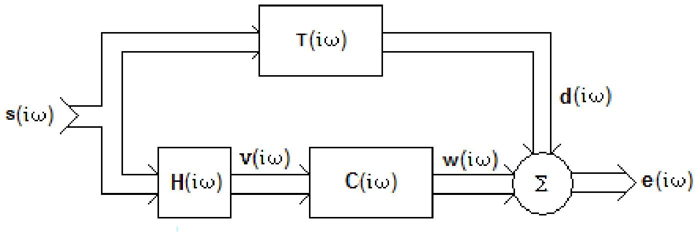


Figure 5. Least squares block diagram.

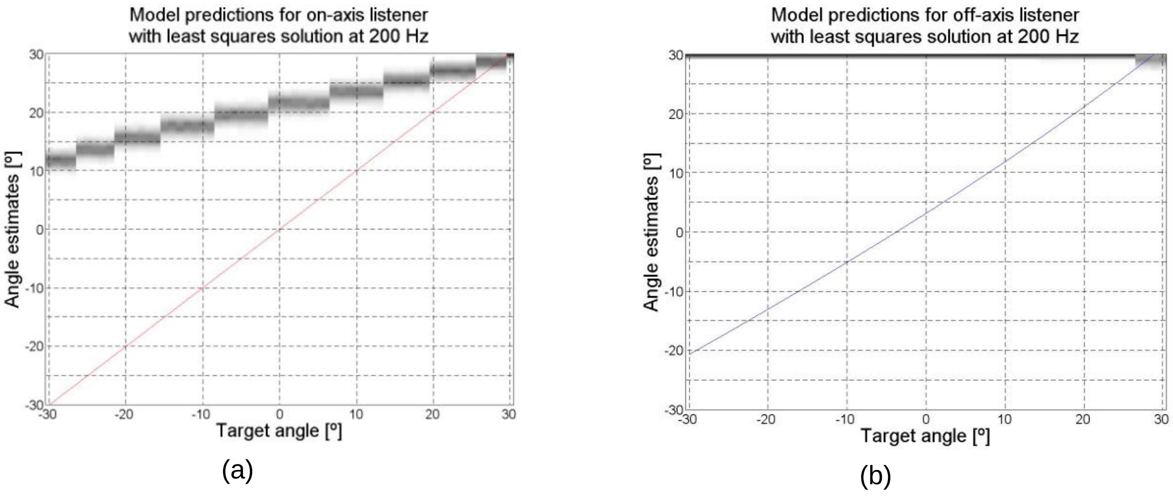
$$\mathbf{C} = \begin{bmatrix} C_{11}\, C_{12} \\ C_{21}\, C_{22} \\ C_{31}\, C_{32} \\ C_{41}\, C_{42} \end{bmatrix} \tag{4}$$

$$\mathbf{T} = \begin{bmatrix} C_{11}z^{-m} & C_{12}z^{-m} \\ C_{21}z^{-m} & C_{22}z^{-m} \\ C_{11} & C_{12}z^{-m} \\ C_{21} & C_{22}z^{-m} \end{bmatrix} \tag{5}$$

$$m = \frac{(l_l - l_r)fs}{c} \tag{6}$$

Where m is the time delay in samples, l_l and l_r are the distances from the left and right loudspeaker to the off-axis listener center head respectively (see figure 1c), *c* is the speed of sound and *fs* is the sampling frequency of the measured HRTFs.

The listeners’ spatial perceptions were computed after passing the stereo signals through the matrix of filters **H** obtained with the least squares solution (Equation 3). Target angles against angle estimates can be seen in figure 6 for a frequency of 200 Hz. As seen, the on-axis listener’s spatial perception is located to its right side (figure 6a); the desired central phantom source is located around 22º from its position. So it can be said that the filters lower the original stereo spatial perception as seen in figure 2a. For the case of the off-axis listener’s perception (figure 6b), it is seen that all the phantom sources are located at the right loudspeaker. In contrast with the perceptions shown in figure 2b, the filters totally pulled the phantom sources at the right side. Finally figure 7 show the comparison of central phantom source perceptions with and without applying the computed filters to the appropriate stereo signals.



(a)

(b)

Figure 6. (a) Model predictions of on-axis listener for the filtered stereo signals using the least squares method at 200 Hz. The darker the higher probability of the angle estimate. (b) Model predictions of off-axis listener for the filtered stereo signals using the least squares method at 200 Hz. The darker the higher probability of the angle estimate.

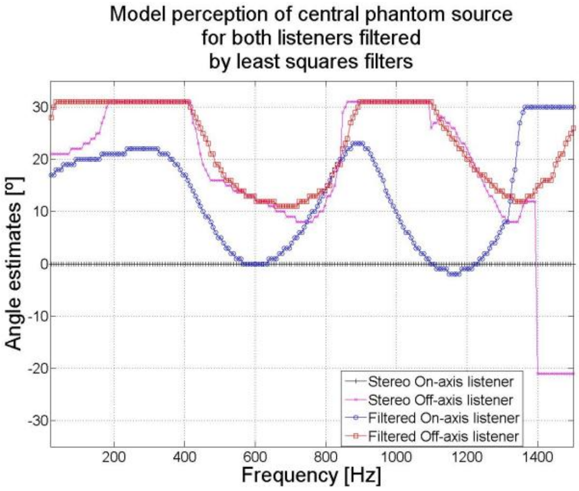


Figure 7. Model predictions of the central phantom source for both listeners across frequency for filtered stereo signals using the least squares method.

4 DISCUSSION

It is then clear that for a stereo signal at 200 Hz, the least squares filters just worsens the spatial perceptions. The ideal perceptions across frequency would be the shown in figure 4b but the obtained were similar to the case where no filters were applied as seen by comparison of figure 7. Furthermore, the filtered perceptions at frequencies lower than 200 Hz are panned to the right source due to the difficulty of inversion at this range. It has to be mentioned that all the presented figures in this work were computed using as stereo signals, two Kronecker delta functions with its corresponding amplitude panning mechanism. Therefore the energy across frequency is the same, so the results show that in principle, the main effort of the filters must correct the perceptions where the central phantom is collapsed to one of the sources. For this particular geometry around 200-400

Hz and 850-1250 Hz but obviously it has to be closely related with the desired signals. Let's speculate that even if we could design an optimal set of filters which works perfectly for all frequency ranges, if the main power of the stereo signals are around 1200-1500 Hz, the optimal perceptions for the off-axis will also be at the nearest source due to the choice of the delay compensation explained in section 2.2 (see figure 4b).

The work done in [5] used an auditory model able to integrate over frequency, three broadband (white noise) signals were used, more specifically 300-1300 Hz, 300-2300 Hz and 300-4500 Hz. Either the first and the third broadband signals gave better perceptions than the second, giving an insight that even if this delay compensation comes from a low frequency approach, it may also work by including higher frequencies (above 2300 Hz). Unfortunately the model used in this work does not give truthful estimates above 1500 Hz due to the exclusion of the ITD information from the signals' envelope. But it would be worth knowing the influence of several frequency ranges on the perceptions of a central phantom source. Into what extent the relative power of the stereo signals in several frequency ranges would make the perception to collapse to one of the sources. This may open the scope on choosing another target matrix rather than delaying signals which has less impact or be less dependent across frequency.

5 CONCLUSION

It is clear that under the geometry and characteristics of the system, the least square optimization method is not "powerful" enough to fit both listeners' perceptions simultaneously. The showed results are not out of logic, it is a hard task to achieve by using only two sources. Even more it has being claimed in [7], that it is impossible to achieve accurate results by using fewer loudspeakers than control points. But as mentioned above, stereophonic technology is still widely used and would be worth if it can be shared outside the sweet spot by using its original configuration.

6 FUTURE WORK

Other optimization methods such as steepest descent, downhill simplex, the Powell's method, conjugate gradient, the Quasi-Newton and finally the Simulated Annealing method are being tried in order to obtain the minimum points of the cost function error and compare between them for the best option. These methods are also computed in the frequency domain thus an extension of the auditory localization model for higher frequencies will be required and a further integration in order to give one perception estimate. It will also be tried with other desired signals as mentioned in the discussion. Furthermore the estimates of the auditory model will be used in order to build the cost function, thus the filters will be designed straight from the auditory model.

As mentioned in the discussion, figure 7 shows some frequencies where are naturally good behaved in terms of matching both perceptions, this gives an insight that there is a correspondence between optimum geometries for particular frequency ranges. This is the optimal source distribution idea fully explained in [8]. But this departs from the original idea of using only two sources. By adding extra sources they must be fitted into two cabinets, for example a pair of dipoles mounted each into a cabinet and see if with this approach can release the control effort made by the filters for a wider frequency range.

7 REFERENCES

[1] B. Bauer, “Broadening the Area of Stereophonic Perception*,” J. Audio Eng. Soc*., vol. 8, pp. 91-94 (1960).

[2] R. M. Aarts, “Enlarging the Sweet Spot for Stereophony by Time/Intensity Trading,” presented at the 94th Convention of the Audio Engineering Society, *J. Audio Eng. Soc*. (*Abstracts*), vol. 41, p.387 (1993 May), preprint 3473.

[3] W. Gardner and K. Martin, “HRTF Measurements of a KEMAR Dummy-Head Microphone,” Tech. Rep. 280, MIT Media Lab Perceptual Computing, Cambridge, MA (1994).

[4] M. Park, “Models of binaural hearing for sound lateralisation and localisation,” Ph.D. dissertation, Southampton Univ., U.K., 2007.

[5] S. Merchel and S. Groth, “Adaptively Adjusting the Stereophonic Sweet Spot tot eh Listener’s Position,” *J. Audio Eng. Soc*., vol. 58, pp. 809-817 (2010 October).

[6] P. A. Nelson and S. J. Ellioitt, *Active Control of Sound*, Academic Press, London, 1992.

[7] O. Kirkeby and P. A. Nelson, “Digital Filter Design for Inversion Problems in Sound Reproduction,” *J. Audio Eng. Soc*,. vol. 47, pp. 583-595 (1999 July).

[8] T. Takeuchi and P. A. Nelson, “Subjective and Objective Evaluation of the Optimal Source Distribution for Virtual Acoustic Imaging,” *J. Audio Eng. Soc*,. Vol. 55, pp 981-997 (2007 November).