EuroRegio2016
June 13-15, 2016
Porto ▪ Portugal

Developing an In-Car 3D Audio System using the latest Virtual Audio Methods

Daniel Wallace, Delphis Migliori, Filipe Soares, Gergo Orosz, Ergo Eelmets

University of Southampton, Faculty of Engineering and the Environment
{djw1g12@soton.ac.uk; dm23g12@soton.ac.uk; fdcs1g12@soton.ac.uk; go1g12@soton.ac.uk; ee2g11@soton.ac.uk}

Abstract

A 3D virtual audio system optimized for in-car use is presented. The proposed system is based on cross-talk cancelation techniques to control the sound arriving at each of the listener's ears. The low acoustic power requirements and well defined listener locations are beneficial to the optimisation of the system. However, the system must account for limitations such as restricted transducer positioning within the car interior as well as the reflections and reverberation inside the car.

The system is based on a discretization of the Optimal Source Distribution (OSD) method, consisting of 3 pairs of transducers placed in the head lining and footwells of the vehicle. Crosstalk cancellation filters are produced from simulated Head Related Transfer Functions, providing flexibility of implementation in various car interiors and seating positions as well as avoiding the laborious and time-consuming process of measuring HRTFs adequate for use as crosstalk cancellation filters.

Subjective evaluation of the system's localization performance has been carried out as a means to assess its potential for binaural reproduction.

1 Introduction

The purpose of all Virtual Audio systems is to produce an acoustical illusion to a listener or group of listeners that sound is arriving from locations not occupied by physical loudspeakers. Such systems employ a diverse range of technologies and techniques to produce this effect, from the familiar "phantom centre" image produced between a pair of stereo loudspeakers [1], to loudspeaker arrays featuring motion-tracking cameras, which provide personalized virtual audio [2]. 3D Audio is a key sector of Virtual Audio research; it is tasked with the reproduction of virtual sound sources at a range of heights and distances from the listener using a small array of loudspeakers.

Although commercial virtual audio products are available for installation in rooms, in-car virtual audio is still in its infancy. An automotive interior is an acoustical space which poses a number of challenges to virtual audio systems. The compact dimensions produce a sound field dominated by acoustic modes at low- to mid- frequencies. While the car interior is not especially reverberant, reflections from windows have been found to interfere with the performance of systems. In addition to this, there is a restricted set of possible loudspeaker positions, as the interior surfaces of the car must serve many other functions. Currently available systems for surround sound reproduction in automotive interior spaces use additional speakers in ceiling of the car, which helps to add the dimension of height to the listening experience [3]. In these systems however, the quality 3D sound reproduction relies on well-distributed loudspeaker positioning.

Crosstalk cancellation (XTC) technology presents a solution to the problem of unsuitable loudspeaker positioning. With XTC, spatial information in the recordings is maintained by means of controlling the sound arriving at each ear independently. This is comparable to the listener wearing headphones, but without the associated isolation from the acoustic environment.

This allows the reproduction of any audio channel format e.g. 2-channel stereo, 5.1 surround, binaural audio etc. by the synthesis of virtual sound sources around the listener. Critically, the spatial cues required by listeners to localize sound in 3D space are present in the binaural signal. In the presented work, the importance of improving the stereo listening experience inside the car is emphasized. However, virtual audio systems for vehicles need not be limited to in-car entertainment. A range of driver aids could also harness the potential of 3D audio to provide intuitive warnings and information.

2 Theory

In order to locate sound sources, the human brain exploits the time and level differences between the sounds arriving at the ears, as well as the spectral variations in the incoming sounds caused by pinna, head and torso. Each of these variations is affected by the direction of the sound source, forming the necessary acoustic cues for sound localization. Understanding and exploiting the principles of spatial sound perception aids the design of various spatial audio systems from simple stereo to advanced three-dimensional sound systems. In all of these cases, the aim is to mimic spatial audio cues during playback using physical sources in a way that the human brain perceives various virtual sources around the head.

2.1 Crosstalk Cancellation

The principle of XTC-based 3D audio is fundamentally different from the stereo and multichannel surround techniques. Instead of using a number of loudspeakers to mimic sound arriving from different directions, XTC aims to use all loudspeakers to control the interaction between the sources and the ears directly. This system aims to suppress the crosstalk audio path that is naturally present in stereo systems. Once sound pressure can be precisely controlled and crosstalk paths are cancelled at the ears, binaural cues encoded in the audio intended to the left and right ears separately can lead to a convincing spatial audio effect, which is now not limited by loudspeaker positions.

In order to have a basic control of sound pressure at the two ear positions, one requires at least two independently controlled acoustic sources. The plant of this acoustical system C is introduced as a 2x2 matrix, containing the transfer functions between the left and right source to the ipsilateral ears (diagonal terms), and similarly to the contralateral ears (off diagonal terms).

Since the plant matrix represents the physical sound propagation, direct manipulation is impossible. Therefore an additional 2x2 matrix is introduced on the signal processing domain denoted as H, which is calculated by the direct inversion of the plant matrix.

$$H = C^{-1} \tag{1}$$

This matrix can be adjusted to achieve the required off-diagonal cancellation. Denoting the received signals at the ear positions as $v(\omega)$ and the input signal as $u(\omega)$, the basic system presented is described as

$$\begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{bmatrix} \begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} \tag{2}$$

which is represented as a block diagram in Figure 1 below.

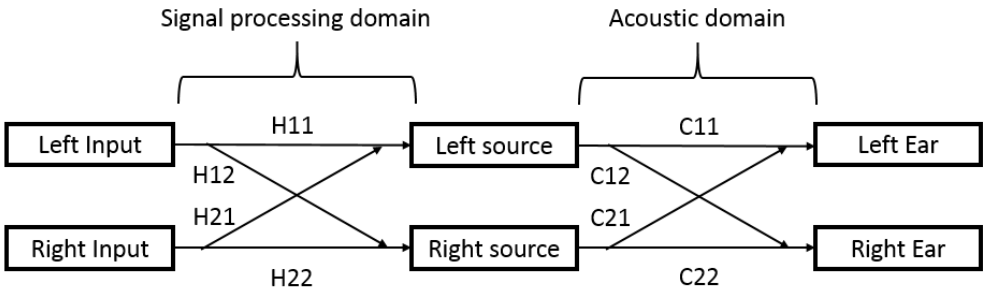

Figure 1. Basic block diagram of XTC system.

2.2 Limitations of XTC and Optimal Source Distribution

The performance of XTC system is limited by various issues, mainly arising from the *half-wavelength problem*. This occurs at frequencies whose wavelength is close to an integer multiple of the path length difference Δr, between the two sources and each ear. Around these frequency regions, the system will have difficulties in controlling the phase relation between receiver positions, i.e. it will be difficult to reproduce in-phase signals at the receivers with in-phase signals at the sources when $\Delta r = \lambda/2$, similarly for out-of-phase signals in the out-of-phase case [4]. This gives rise to a loss of dynamic range and a lack of robustness to errors in the estimate of the plant matrix or in the presence of reflections or reverberant acoustic environments.

The condition number $\kappa(C)$ of the plant matrix C can be used as an indicative measure of the crosstalk cancellation performance, pointing out where problematic frequencies will lie with a specific sources-receivers geometry. This parameter is given by

$$\kappa(C) = \|C\|\|H\| = \|H^{-1}\|\|H\| = max\left(\frac{\sigma_1}{\sigma_2}, \frac{\sigma_2}{\sigma_1}\right) \tag{3}$$

where σ_1 and σ_2 are the singular values of the plant matrix C, obtained via the Singular Value Decomposition method [5].

Where $\kappa(C)$ is large, the reproduced signals at the receiver w are going to be less robust to small changes in the inverse of the plant matrix (C^{-1}), i.e. the inverse filter H is likely to contain large errors due to small errors in C. Frequencies where $\kappa(C)$ is low will result in a nearly ideal inverse filter H, indicating the potential for good XTC performance.

Although these limitations can never be completely overcome, they can be minimized by using the Optimal Source Distribution method (OSD). This method relies on using combinations of source positions and frequency bandwidth such that the condition number of the system is minimized over the widest frequency range possible. Such a system consists of a finite number of loudspeaker pairs, with different source span angles and frequency bandwidths, that will radiate only reasonably well-conditioned frequencies. The colour plot in Figure 2 below illustrates, in a simplified manner, the process of optimization, in this case for a 4-way OSD system.

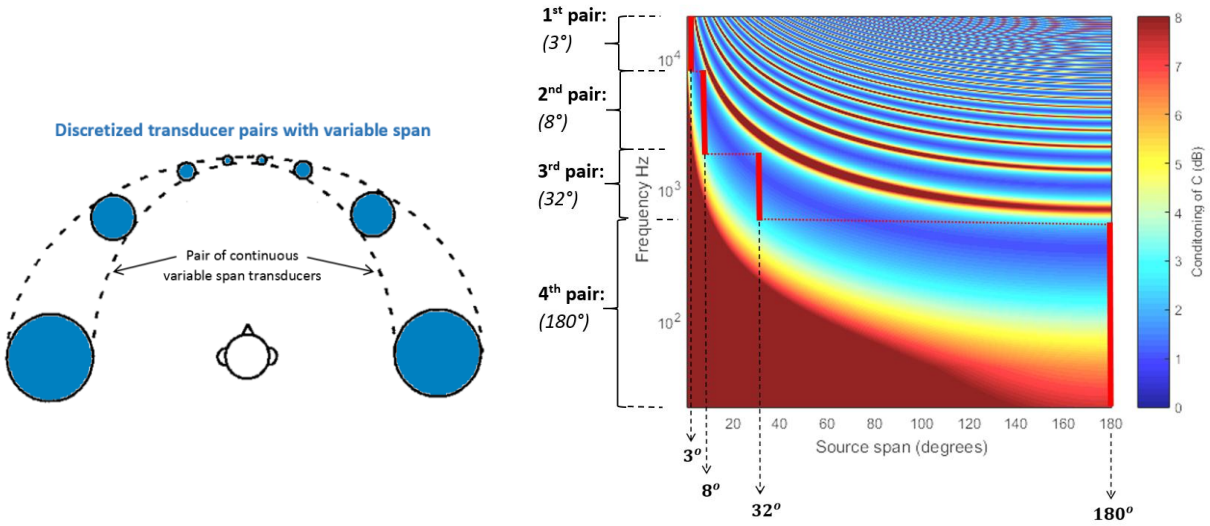

Figure 2. (Left) Diagram illustrating the process of designing a practical system discretized from the ideal OSD. (Right) The colour plot shows the condition number as a function of source span and frequency for a free-field model with two monopole sources and receivers. The red lines denote the chosen span angle and frequency range for each transducer pair.

3 In-car acoustics

The confined interior space in vehicles leads to a number of acoustical effects that can be detrimental to spatial sound reproduction from regular in-car entertainment systems [6]. The low reverberation time in vehicles leads to a reduced spatial perception – this is perhaps coupled with non-acoustic cues; listeners in physically small spaces do not expect to perceive acoustically large spaces [7]. Further concerns related to the small size of automotive interiors are present in the frequency response of such spaces. At low frequencies, the acoustic response of any cavity is dominated by acoustic modes, which begin to overlap significantly at the Schroeder frequency, which lies around 500 Hz for an average sized car. This indicates that the frequency response of a vehicle below this frequency, although predictable, will vary significantly.

Loudspeaker positioning is constrained by the vehicle's geometry, and will have further constraints imposed on it for optimal source distribution, leaving little flexibility to take into account the car's modal behavior as well.

While the interior of the car has a low reverberation time, the presence of nearby reflective surfaces such as the dashboard and windows may cause a comb filtering effects in the frequency domain, degrading the performance of the system. The data of the car interior lining and seats exhibit high sound absorption characteristics, whereas the window proves to be a strongly reflective surface. [8] This finding expresses a concern of the achievable XTC performance in the car interior since the listener positions are inherently placed in close proximity to the side window.

As found by Morkholt et. al, the average reverberation time inside cars is around 0.05 s above the estimated Schroeder frequency as well as up to approximately 0.1 s in the modal frequency region of the car interior [8]. This measured result can be used to determine the critical distance r_c. This is the radial distance from the source at which the sound pressure due to the direct field is equivalent to the sound pressure due to the reverberant component of the sound field. The critical distance will be evaluated using the highest reverberation time occurring in the modal frequency region. The total

volume of the car is estimated to be 5 m^3 and the source is assumed to be omnidirectional. With these assumptions in mind, an approximate value for the critical distance in the car is $r_c \approx 40$ cm.

Bearing this in mind, it is anticipated that if the mid- and high-frequency drivers (operating above f_{Sch}) are placed within this distance of the listener's ears, the detrimental effect of reverberation on XTC performance can be reduced as the direct field from the loudspeakers will dominate over the reverberant field.

4 Design, Methodology and Results

During the course of our work, several systems were designed, built and tested (both objectively and subjectively). Following the optimization process described in Section 2.2 above, filters were designed using both measured and simulated Head Related Transfer Functions (HRTFs), which were then assessed subjectively inside the car, based on their localization performance.

4.1 Measured HRTFs

Initial considerations suggested that the most feasible option would be a 3-way system with tweeters and mid-range pairs placed in a soundbar near the sunvisor and woofers placed in the footwells. The first HRTF measurements (using a KEMAR dummy head) were carried out in a small anechoic chamber with a variable span loudspeaker cabinet, in order to find the optimal positions for the tweeters and mid-ranges in the soundbar. However useful as an indication of where the drivers should be placed within the soundbar, the measurements were found inappropriate for the production of inverse filters. The presence of noise during measurement sessions, however small, contributed to the poor quality of these measurements, as their absolute conditioning was relatively high and contained exporadic peaks. Attempts to invert the measurements resulted in audible artefacts introduced by errors in the inversion process when the conditioning of the plant matrix estimate is poor. Similarly, measurements of the HRTFs inside the car (outside the anechoic chamber) led to even higher conditioning due to the noisier conditions and the effects of car's modal behavior and reflections.

4.2 Simulated Head Related Transfer Function (HRTF)

The inability to obtain HRTF measurements adequate to the design of the XTC filters, led us to consider using modelling HRTFs. A range of models were developed from a simple free-field based on point sources and receivers to more complex models using sound scattering around a rigid sphere and simulated pinna reflections. Using modelled HRTFs has the advantage of being easily adaptable for various car interiors or seating positions, by simply changing the parameters in the software. It also avoids the laborious and time consuming process of measuring HRTFs.

The final crosstalk cancellation filters used in the project are based on an analytical model of scattered pressure field around a rigid sphere (Simulated HRTFs). This model calculates the pressure p on the surface of a rigid sphere of radius r_h due to insonification by a point source at a particular radial distance r_0 from the centre of the head [9]. p is ascertained from the total volume potential V, a linear combination of the incident field from the point source $V_i = \frac{e^{jkr_0}}{kr_0}$ and the scattered field from the sphere V_s. Specifically,

$$p = \rho \frac{\partial}{\partial t} V. \tag{3}$$

$$V(k,\theta_i,r_0,a) = V_i + V_s = -\frac{1}{(ka)^2}\sum_{n=0}^{\infty}(2n+1)P_n(\cos\theta_i)\frac{h_n^{(1)}(kr_0)}{h_n^{(1)\prime}(ka)} \quad (4)$$

where P_n is a Legendre polynomial, and $h_n^{(1)}$ is a spherical Hankel function.

In this formulation, a rotated spherical polar coordinate system is used, as shown in Figure 3, with the pole of the sphere oriented to coincide with the source position, rather than being oriented with the physical features of the head. The volume potential, and thus the pressure on the surface of the sphere, is axisymmetric about this pole and can therefore be described by the single angular coordinate $\theta_{i\,(L,R)}$. For a general source position, the left and right ears lie on different circular arcs around this pole, subtending an angle of θ_{iL} and θ_{iR} respectively. This allows for the frequency dependent interaural level and time difference to be calculated simply by evaluating Eq. (4) at these two angles. In the practical algorithm, it was found that 45 iterations are sufficient to approximate the infinite sum to frequencies above 20 kHz.

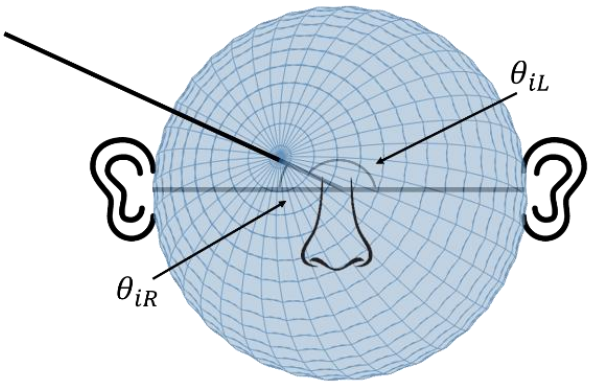

Figure 3. Rotated spherical coordinate system for a source above and to the right of a forward-facing listener, indicating the angles of each ear.

In real HRTFs the reflections from the pinna forms a significant role in spatial hearing. The exact timing and strength of pinna reflections are individual to particular listeners, though a tapped delay line filter (diagram in Figure 4) with n taps at time τ_{pn} has been theorised [10].This filter mimics some of the most prominent effects, using a KEMAR mannequin's standardised pinna. The expression for τ_{pn} is:

$$\tau_{pn}(\varphi,\delta) = A_n\cos\left(\frac{\varphi}{2}\right)\sin\left[D_n\left(\frac{\pi}{2}-\delta\right)\right] + B_n \quad (5)$$

where ρ_n, A_n, B_n and D_n are the pinna reflection model coefficients. Table 1 below reports the numerical values of these coefficients for the n^{th} delay.

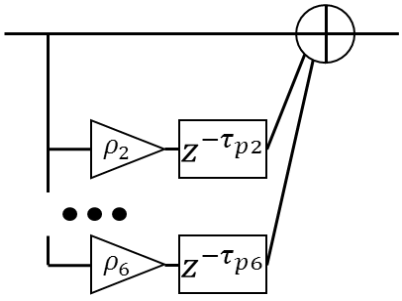

Figure 4. Time domain block diagram of pinna reflection filter.

Table 1. Pinna reflection model coefficients.

n	ρ_n	A_n	B_n	D_n (set 1)	D_n (set 2)
2	0.5	1	2	1	0.85
3	-1.0	5	4	0.5	0.35
4	0.5	5	7	0.5	0.35
5	-0.25	5	11	0.5	0.35
6	0.25	5	13	0.5	0.35

Due to the short delays required relative to the sampling interval $1/F_s$, these filters are implemented using interpolation between samples. Either of the two sets of D_n can be used in the algorithm, allowing a potential to achieve some customization for different listeners.

4.3 Conditioning of plant matrices of the simulated acoustic system

As described in Section 2.2, the conditioning of the acoustic plant of the system provides guidance on chosing an optimal combination of source span and frequency range for each loudspeaker pair. In order to achieve adequate spatial audio performance, the conditioning of the system is required to be as low as practically possible. Figure 5 shows the conditioning of the simulated plat matrix as a function of source span and frequency, evaluated for sources 35 cm away from the listener.

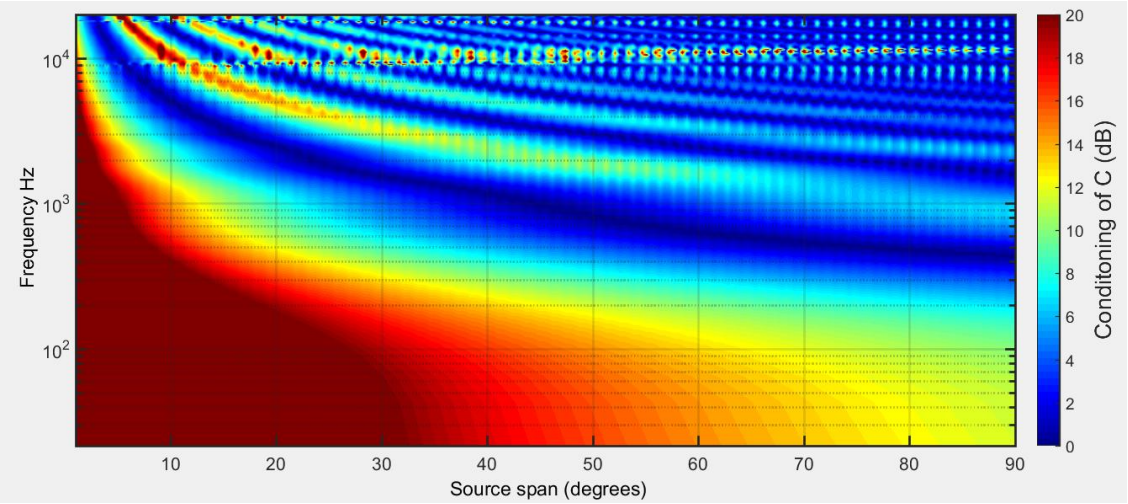

Figure 5. Conditioning of the plant matrix against span angle and frequency.

The half wavelength problem of matrix inversion can be generalized to more complex models. When the two transfer functions between a single loudspeaker and a pair of ears at a particular frequency are equal in magnitude and phase, high conditioned plant matrices are constructed. Especially above 10 kHz, as well as down to approximately 3 kHz, small regions of high conditioning are observable. It is understood that nature of this behavior is physical, a result of the combination of comb filtering effect of the analytical reflections from the rigid sphere and the resultant equal transfer functions at the two ears.

By considering the possible source locations in the automotive interior, loudspeaker pairs are positioned at span angles which provide minimum conditioning. A 3-way system has been developed to suit a car interior with the parameters as shown in Table 2.

Table 2. Parameters for the proposed 3-way system.

Example 1	Span angle θ	Elevation angle δ	Radial distance to listener	Frequency range
Woofer	34 ^{o}	-43.4^{o}	87.4cm	100-300 Hz
Midrange	44.7^{o}	17.0^{o}	34.1cm	300-6000 Hz
Tweeter	5.4^{o}	18.4^{o}	31.7cm	6000-20000 Hz

In this arrangement, the tweeters and midrange speakers can be mounted in the position of the sun-visor of the car allowing some flexibility. The woofers are mounted in the footwell of the car, restricting the possible positions severely, leading to non-optimal but maximum span angle of 34 degrees.

4.4 Modelled vs. Measured HRTF

Figure 6 shows the conditioning of the simulated HRTF filters against those measured in the anechoic chamber. The conditioning of the filters produced from the measurements are here smoothened to show their general trend, thus shadowing the effects of the noise. The overall conditioning of the analytical plant matrices for the system design is below 10 dB from 900 Hz to ~13 kHz. There is good agreement between the measured conditioning of the system in anechoic conditions and that from the analytical model. This is a strong indicator that the modelled HRTF is a good approximation of the actual acoustic plant.

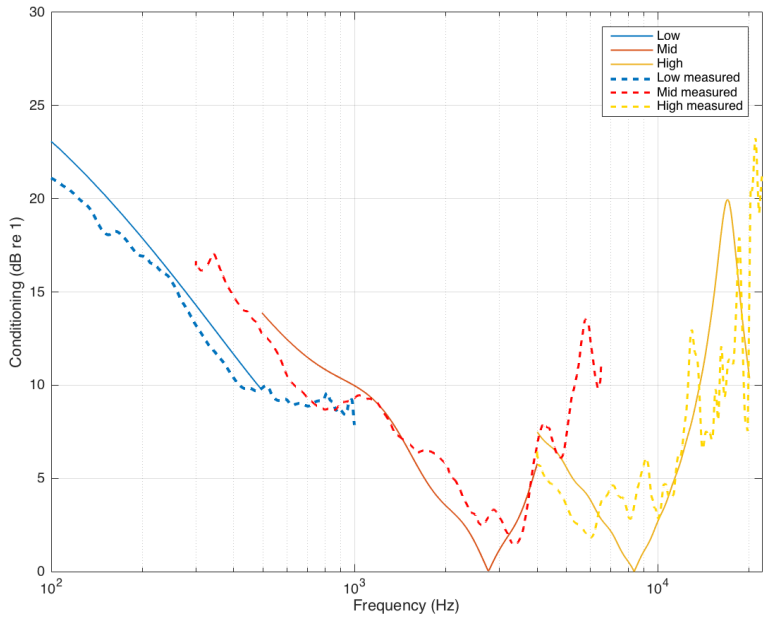

Figure 6. Comparison of measured and analytical conditioning of the demonstration system as described in Table 2.

4.5 Subjective Tests

Alongside objective experiments, subjective experiments were carried out. The aim was to investigate how well our 3D audio system was able to convincingly reproduce virtual images based on the response of a subject. A necessary (but not sufficient) descriptor of a spatial audio system is its subjective sound localisation performance, which was deemed to be informative enough for the scope of this work.

The experiment attempted to establish a relationship between the test condition choice (the independent variable) and the mean absolute localisation error D (the dependent variable). The test conditions were three: listening over headphones (perfect cross-talk cancellation), listening through the designed 3D audio system with XTC based on free-field model ("model 1"), and listening through the same system but with XTC based on simulated HRTF ("model 3"). Twenty-two subjects were asked to locate the virtual sound source of a 250ms-long white noise signal, which was convolved with seven HRTFs associated with azimuthal angles of -80, -55, -30, 0, 30, 55 and 80 degrees.

The collected data was linearly interpolated to obtain the mean absolute localisation error *D* for each test configuration per subject. Initial observations on *D* can be made by means of boxplots and descriptive statistics (mean *M* and standard deviation *SD*) in Figure 7. A statistically significant difference can be expected between listening over headphones and the remainder conditions, since there are large differences in means *M* with no noticeable overlaps in standard deviations *SD*. For the same reasons, no significant difference can be expected between conditions "model 1" and "model 3". The outcome of the statistical analysis (a repeated measure analysis of variance) confirms the above observations, especially when a pairwise comparison of conditions (Bonferroni analysis) is considered, as shown in Figure 7; in fact, we can appreciate that the *significance* p is less than 0.05 for the pairs "headphon"-"model 1" and "headphon"-"model 3", but not for "model 1"-"model 3".

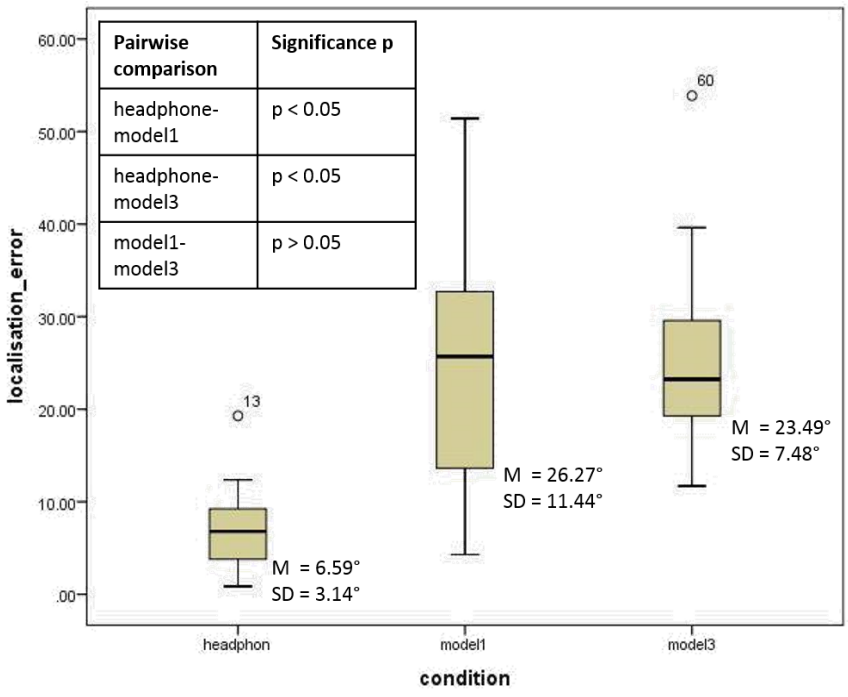

Figure 7. Boxplots and statistics of mean absolute localization error D for the tested conditions

5 Conclusion

A set of design guidelines for the implementation of an in-car virtual audio system based on crosstalk cancellation has been developed. The developed method is used to determine a set of optimal loudspeaker positions and crossover frequencies, by considering system conditioning and robustness to reflections. The method is generalised in the sense that it can be applied to a wide range of different vehicle interiors. The guidelines can be summarised by the following:

- To achieve satisfactory crosstalk cancellation over a wide frequency range, a multi-way loudspeaker system is required. A minimum of three pairs of loudspeakers is recommended.
- The mid- to high-frequency components of the system should be reproduced by two or three pairs of small loudspeaker drivers arranged in a soundbar configuration. The soundbar should be located in the headlining of the vehicle, near the sunvisor, and in line with the listener's seat.
- Woofers should be mounted symmetrically in the door and footwell of the vehicle. The loudspeakers should be mounted as close to the listener as is practical, to increase the loudspeakers' span relative to the listener.
- In order to overcome the detrimental effect of early reflections, and to reduce the dynamic range loss in the system, the distance between the loudspeakers and the listener's head should be brought within the *critical distance* of the acoustic space, nominally ~40 cm in most vehicles, where practical.
- Crosstalk cancellation filters should be generated using simulated Head Related Transfer Functions. These should be based on the transfer function between an omnidirectional point source and a point on a rigid sphere.
- By varying the positions and crossover frequencies of the loudspeakers, the condition number of the resulting acoustic plant matrices can be minimised. A system with as low conditioning as possible over a desired frequency range is deemed optimal.

During the course of the project, physical prototypes were designed, built, and tested both objectively and subjectively in order to verify these conclusions.

It has also been found that a crosstalk cancellation filter length of 2048 taps is adequate to achieve cross talk cancellation without noticeable sound quality loss. This relatively short filter length promotes the feasibility of using a real-time convolution engine to achieve a low latency real-time demonstration system.

It is hoped that the outcomes of this Project provide valuable contributions to the design and implementation of virtual audio systems within cars, thus promoting the growth and awareness of this exciting audio technology.

References

[1] A. Blumlein, "Improvements in and relating to Sund-transmission, Sound-recording and Sound-reproducing systems," *UK Patent 394325,* 1933.

[2] M. Song et. al "An Interactive 3D Audio System with Loudspeakers," *IEEE Transactions on Multimedia,* vol. 13, no. 5, pp. 844-855, 2011.

[3] Mercedes Benz, [Online]. Available: https://www.mercedes-benz.com/en/mercedes-benz/innovation/the-burmester-sound-system-in-the-new-s-class/.

[4] T. Takeuchi, "Optimal Source Distribution for Virtual Acoustic Imaging," University of Soutahmpton, Southampton, 2000.

[5] Massachusetts Institute of Technology, "Singular Value Decomposition", *Open Course Ware,* 2011.

[6] R. Mingsian et. al, "Signal Processing Implementation and Comparison of Automotive Spatial Sound Rendering Strategies," *EURASIP Journal on Audio, Speech and Music Processing,* 2009.

[7] B. Gardner, "A Realtime Multichannel Room Simulator," New Orleans, 1992.

[8] J. Mørkholt et. al, "Measurement of reverberation time of a passenger car utilizing the wavelet filter bank," *Inchon: Department of Mechanical Engineering,* 2004.

[9] J. Bowman et. al, "Electromagnetic and Acoustic Scattering from Simple Shapes," Michigan University, 1970

[10] C. Brown et. al, "A Structural Model for Binaural Sound Synthesis," *IEEE Transactions on Speech and Audio Processing,* vol. 6, no. 5, pp. 476-487, 1998.