Personal Spatial Audio in Cars

Development of a loudspeaker array for multi-listener transaural reproduction in a vehicle

C. House University of Southampton, ISVR
S. Dennison University of Southampton, ISVR
D. G. Morgan University of Southampton, ISVR
N. Rushton University of Southampton, ISVR
G. V. White University of Southampton, ISVR
Dr. J. Cheer University of Southampton, ISVR
Prof. S. Elliott University of Southampton, ISVR

1 INTRODUCTION

High quality consumer audio products are well established, however, research into improving the performance and functionality of car audio systems is still growing. This project attempts to enhance the capabilities of in-car audio by combining personal and spatial audio technologies via an array of loudspeakers to reproduce separate spatial audio for both the driver and the front passenger of a car.

A MATLAB model of the array has been developed, including source directivity and first order reflections, and the parameters obtained from this acoustic model were used to aid the design of the array. The physical array was CNC milled out of aluminium to create a sleek, high quality product and contained 27 nominally identical loudspeakers. The array was tested using filters created from both modelled and measured transfer functions, created using the least mean squares (LMS) algorithm to control the pressure at four control points.

The array was tested in both an anechoic chamber and a car to assess its performance in terms of crosstalk cancellation between control points corresponding to the positions of the listeners' ears. It was found that the array performed very well in anechoic conditions creating 25 - 45 dB of Crosstalk Cancellation (CTC) between 300 Hz and 11 kHz, and even with the addition of several reflections in the car environment, it was still found to perform well with 18 - 40 dB of CTC over the same frequency range.

2 BACKGROUND

In-car audio systems are developing rapidly, with many car manufacturers collaborating with specialist audio companies to realise improved sound quality. BMW & B&W,[1,2] Porsche & Burmester[2,3] and Audi & B&O[4,5] are three examples of such collaborations, each using loudspeakers distributed throughout the vehicle to introduce an element of surround sound to their vehicles. This would allow for hazard/collision warnings, parking sensors and even hands-free mobile phone conversations to come from specific areas in the car, as well as passengers of the car to watch a film in surround sound.

Other automotive companies are focussing on zonal audio. Harman's Individual Sound Zones,[6] Dyn Audio,[7] Bose NEAR[8] and Philips Music Zone technology[9] are all automotive sound systems that claim to create separate audio zones around the car, allowing the driver and passenger(s) to simultaneously listen to different streams of audio.

This project aims to combine these systems, as well as incorporating previous research from the ISVR,[10,11] to develop a loudspeaker array which can simultaneously reproduce two streams of binaural content into two distinct zones; the front two seats of the vehicle.

3 THEORY

3.1 Least Mean Squares Optimisation

The least mean squares (LMS) algorithm is an invaluable active control tool designed to adapt a set of system filters to minimise an output error. A uniform linear array of L sources driven by signals v_1 to v_L is considered, as shown in Figure 1.

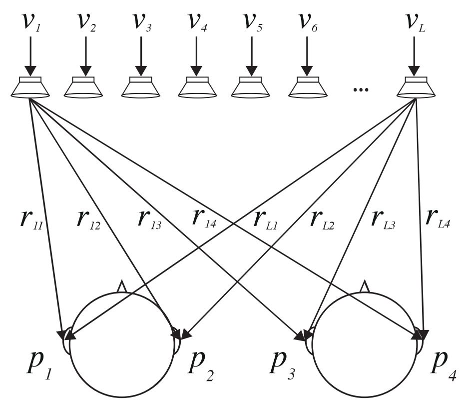

Figure 1: Array geometry.

The pressure p at the listeners' ears can be expressed in the frequency domain as

$$\underbrace{\begin{pmatrix} p_1 \\ p_2 \\ p_3 \\ p_4 \end{pmatrix}}_{\mathbf{p}} = \underbrace{\begin{pmatrix} C_{11} & \dots & C_{L1} \\ C_{12} & \dots & C_{L2} \\ C_{13} & \dots & C_{L3} \\ C_{14} & \dots & C_{L4} \end{pmatrix}}_{\mathbf{C}} \underbrace{\begin{pmatrix} v_1 \\ \vdots \\ v_L \end{pmatrix}}_{\mathbf{v}}, \tag{1}$$

where $\mathbf{C}$ is the plant matrix between each loudspeaker in the array and each ear of the two listeners.

Assuming the sources act as point monopoles, the plant matrix $\mathbf{C}$ can be defined as

$$\mathbf{C} = \frac{j\omega\rho_0}{4\pi} \begin{bmatrix} \frac{e^{-jkr_{11}}}{r_{11}} & \dots & \frac{e^{-jkr_{L1}}}{r_{L1}} \\ \frac{e^{-jkr_{12}}}{r_{12}} & \dots & \frac{e^{-jkr_{L2}}}{r_{L2}} \\ \frac{e^{-jkr_{13}}}{r_{13}} & \dots & \frac{e^{-jkr_{L3}}}{r_{L3}} \\ \frac{e^{-jkr_{14}}}{r_{14}} & \dots & \frac{e^{-jkr_{L4}}}{r_{L4}} \end{bmatrix}, \tag{2}$$

where the air density ρ_0 = 1.28 kgm^{-3}, the speed of sound in air c_0 = 343 ms^{-1}, the wavenumber $k = \omega/c_0$ and r_{mn} is the distance between source m and the receiving point n.

A matrix of crosstalk cancellation filters $\mathbf{H}$ can be calculated in the frequency domain and multiplied with the vector of desired audio signals $\mathbf{a}$ such that the acoustic pressure at each ear replicates the desired audio signal:

$$\underbrace{\begin{pmatrix} p_1 \\ p_2 \\ p_3 \\ p_4 \end{pmatrix}}_{\mathbf{p}} = \underbrace{\begin{pmatrix} C_{11} & \dots & C_{L1} \\ C_{12} & \dots & C_{L2} \\ C_{13} & \dots & C_{L3} \\ C_{14} & \dots & C_{L4} \end{pmatrix}}_{\mathbf{C}} \underbrace{\begin{pmatrix} H_{11} & H_{12} & H_{13} & H_{14} \\ \vdots & \vdots & \vdots & \vdots \\ H_{L1} & H_{L2} & H_{L3} & H_{L4} \end{pmatrix}}_{\mathbf{H}} \underbrace{\begin{pmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \end{pmatrix}}_{\mathbf{a}} = \underbrace{\begin{pmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \end{pmatrix}}_{\mathbf{a}}. \tag{3}$$

Equation 3 holds true when the matrix of crosstalk cancellation filters $\mathbf{H}$ is equal to the exact inverse of the plant matrix $\mathbf{C}$, however for non-invertible matrices the pseudo-inverse can be used and Equation 3 can be considered approximate. Bai *et al.*[12] explain how the problem can be expressed using the minimisation of the Frobenius norm of an equation in the form

$$\underbrace{\begin{pmatrix} p_{t1} \\ p_{t2} \\ p_{t3} \\ p_{t4} \end{pmatrix}}_{\mathbf{p_t}} = \underbrace{\begin{pmatrix} C_{11} & \dots & C_{L1} \\ C_{12} & \dots & C_{L2} \\ C_{13} & \dots & C_{L3} \\ C_{14} & \dots & C_{L4} \end{pmatrix}}_{\mathbf{C}} \underbrace{\begin{pmatrix} H_{1t} \\ \vdots \\ H_{Lt} \end{pmatrix}}_{\mathbf{H_t}}, \tag{4}$$

where the target pressure at control point 1 (for example) is specified as: $\mathbf{P_t} = \begin{bmatrix} 1 & 0 & 0 & 0 \end{bmatrix}^T$. $\mathbf{H_t}$ is a single column of $\mathbf{H}$ corresponding to the chosen control point.

The LMS solution of Equation 4 using the pseudo-inverse technique with Tikhonov regularisation is given as[13]

$$\mathbf{H_t} = [\mathbf{C}^H\mathbf{C} + \beta\mathbf{I}]^{-1}\mathbf{C}^H\mathbf{P_t}, \tag{5}$$

where the superscript H is the Hermitian transpose and β is a regularisation parameter.

The LMS algorithm has been used a number of times for the generation of optimal CTC filters; for example, Simón-Gálvez's 16-channel array.[14] This array was designed using 16 separate control signals to allow for soundfield control at two control points; a single listener. With the inclusion of head tracking software, this system was shown to be an effective array and is a good example of efficient use of the LMS algorithm.

3.2 Spatial Aliasing

Spatial aliasing describes the physical limitation of an array that occurs when the spacing between loudspeakers in an array is larger than a wavelength. Because of the inherent periodicity in the array pattern, symmetric side lobes may appear at angles other than the target angle when the array is working at high frequencies; these are known as grating lobes. Although in theory these grating lobes do not intersect with a listener position, and therefore will not be heard, they waste acoustic energy, and in a reverberant environment, can start to cause issues.

The amplitude of these grating lobes increases with frequency, and at the point that the main lobes and the grating lobes have equal amplitude, it is said that spatial aliasing occurs. Mathematically this is written as

$$\frac{d}{\lambda} > 1, \tag{6}$$

where d is the spacing between loudspeakers and λ is the wavelength. Equation 6 should not be satisfied to avoid producing fully developed grating lobes.

Grating lobes can be controlled to a certain extent by using a non-uniformly spaced array (at which point the average source spacing becomes relevant), or by applying array shading techniques, similar to windowing used when controlling the side lobes in a Fourier transform.

It can be seen from Equation 6 that the upper frequency limit of the array increases as the spacing d decreases, however the size of the loudspeakers ultimately governs how closely spaced they can be. This is why the decision to adopt a double line array was taken, as well as the decision to remove one corner of the loudspeaker housing (both discussed in Section 4).

The final array loudspeaker spacing of 35.2 mm will result in a far-field upper frequency limit of 9.8 kHz.

3.3 Crosstalk Cancellation (CTC)

CTC is a key parameter in analysing the level of attenuation between consecutive sound zones; either between left and right ears of a single person (as is the case for binaural reproduction), or between multiple distinctly separate sound zones. In the former case, it is generally assumed that 20 dB of CTC will achieve a reasonably convincing spatial envelopment for a listener, however the latter will require greater levels of CTC to reproduce different programme material in each zone.[15]

CTC is normally defined as the ratio of the mean square pressure received at the desired ear to the mean square pressure received at the undesired ear. This definition has now been extended to include a summation of all the undesired signals:

$$CTC = \frac{1}{N}\sum_{m=1}^{N}\frac{|p_{mm}|}{\sum_{n\neq m}^{N} p_{mn}} \tag{7}$$

where N is the number of control points and $|p_{mn}|^2$ is the mean square pressure received at the control point n when the signal is sent to control point m. This intuitively is the mean square pressure sent to the desired control point divided by the sum of the mean square pressures received at the other control points. This is then averaged for all N control points.

3.4 Numerical Modelling

To investigate the filter design process, a 2 dimensional numerical model of the loudspeaker array was developed, based on the following initial assumptions:

- Free-field propogation
- Monopole source directivity

From this model, the pressure field generated by the array could be predicted, as shown in Figure 2. This enabled the performance of various array configurations to be visualised and benchmarked for comparison, including number of loudspeakers, loudspeaker spacing, and loudspeaker layout.

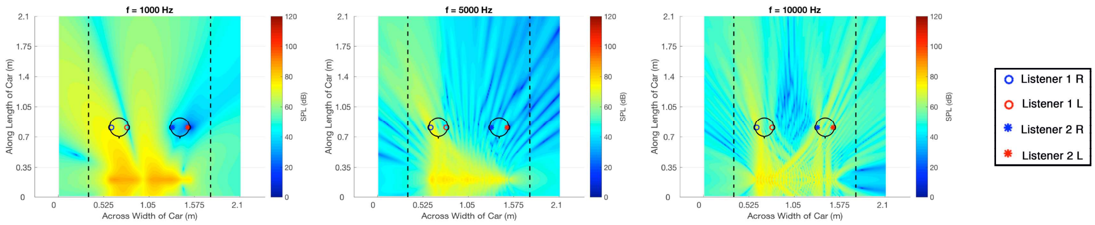

Figure 2: Predicted Pressure Field.

To provide a more realistic simulation, the numerical model was further developed to include the effects of:

- First order side window reflections
- Piston source directivity
- Spherical head scattering
- Height differences between each listener and array (effectively making it a 3D model)

Due to the complexity of these results they are not shown here, however they have been used in the later filter design study.

4 ARRAY DESIGN & MANUFACTURE

Thiele-Small analysis was conducted on the loudspeakers, and an optimum cabinet volume was calculated to realise a low frequency roll off of 250 Hz without needing to have an excessively large array cabinet.

As can be seen in Figure 4, it was decided to adopt a ‘double line’ array, with two separate lines of loudspeakers. This decision was fuelled by the desire to create the longest array practically possible with the smallest loudspeaker spacing, as discussed previously. After choosing a specific loudspeaker unit, the spacing between the loudspeakers and the overall height were minimised through careful design. This required one corner of the loudspeaker casing to be removed which wouldn’t affect the acoustic performance of the units. Given the internal dimensions of the car, this allowed for a 27 loudspeaker array with a horizontal spacing of 35.2mm.

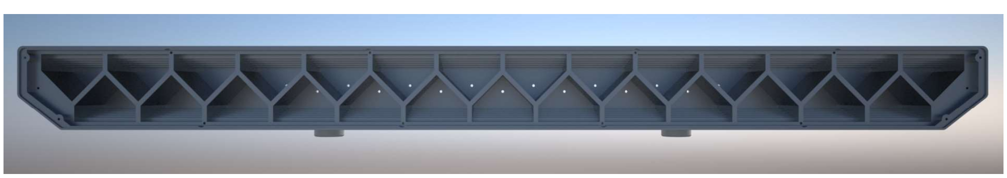

Figure 3: Render of the main array body, showing the stepping to increase machine efficiency.

To increase machining efficiency, the walls of the internal cavities were stepped and the curved edges of the array were squared off, as shown in Figure 3. Aluminium was chosen as the material for its strength and machining suitability, however this required significant acoustic treatment inside the array body to minimise resonances; thus wadding and damping material were applied to the interior faces. Even after these adjustments, it still took approximately a week of 8hr days to machine all parts of the array.

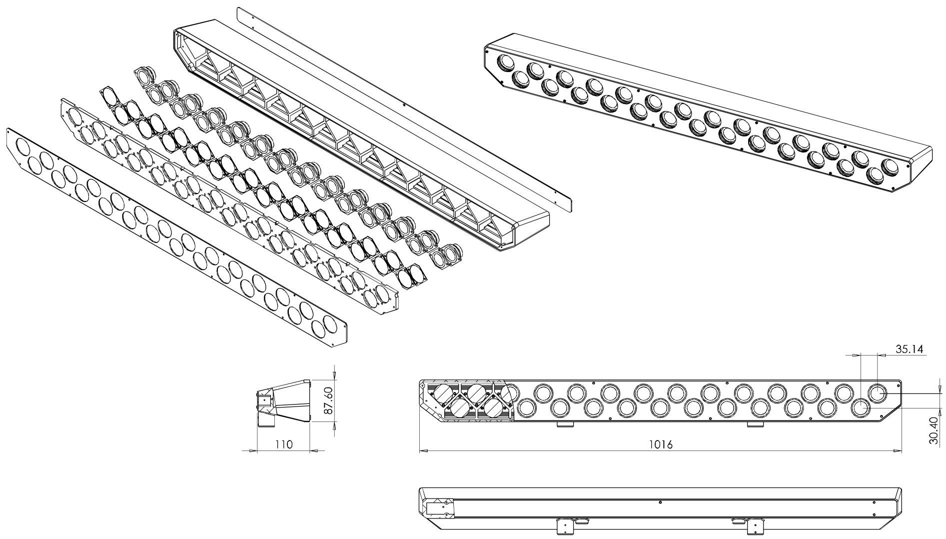

Figure 4: Drawings of the Loudspeaker Array.

Two struts supported the array in the car, slotting into the underside of the array and sitting either side of the central column in the vehicle. As seen in Figure 5, in this configuration the array barely reaches the rear-view mirror and if it was to be supported from the roof rather than the ground it is believed that the impact on a driver would be minimal.

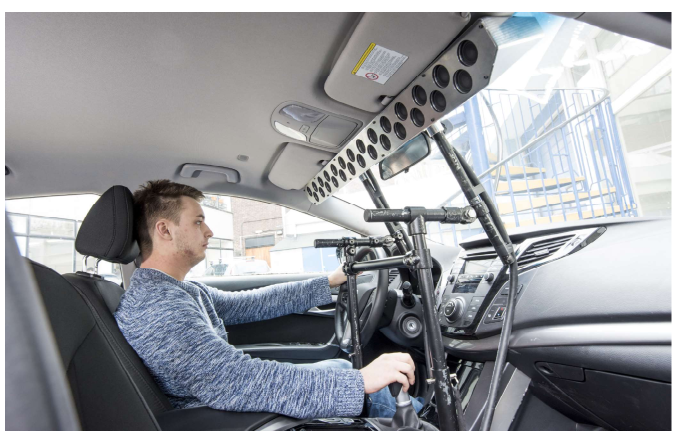

Figure 5: Array Mounted in the Car.

The loudspeakers were rear-mounted and countersunk into the front face to allow sufficient airflow into the loudspeaker units, and to avoid blocking any vents. Laser-cut rubber gaskets were also used to create a firm seal around the loudspeakers, ensuring maximum efficiency at low frequencies.

Due to the complexity of the design shown in Figure 4, and the requirement for each cavity to be identical, it was decided that CNC machining would be the optimal way to manufacture the array. This allowed for a high-precision prototype to be realised, and minimised the potential for human error in the construction.

5 REALTIME DSP

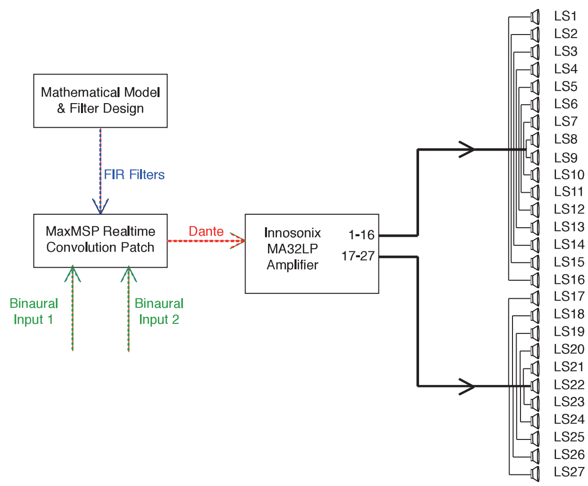

Figure 6: Hardware System Schematic.

5.1 Hardware

A schematic of the hardware system layout can be seen in Figure 6. An Innosnix MA32 amplifier[16] was used to drive each loudspeaker via DANTE. MaxMSP was used to develop a realtime implementation of the array processing, allowing passengers in the car to easily select what they wish to listen to; including time critical sources such as the radio or a mobile phone which cannot be pre-processed.

5.2 Filters

The mathematical model discussed previously was used to generate a set of CTC filters. Each filter set consisted of 108 FIR filters, each 2048 taps long. Varying levels of regularisation (including Frequency Dependent Regularisation) were used, and their performance was compared both objectively in terms of measured CTC and Array Effort, and subjectively through informal listening assessments.

To compare this, filters calculated from measured data were also used to drive the array. The array and a pair of dummy heads were placed in an anechoic chamber, and the response between each loudspeaker and each ear was measured using the exponential sine-sweep method. This setup is shown in Figure 8. This measurement of the plant response was then inverted in a similar way to the modelled plant to generate filters.

5.3 Software

MaxMSP was used to convolve the designed filters with the input signals in real-time. Each of the four input signals was convolved with 27 different filters to generate 108 separate signals. The contributions to each loudspeaker were then summed to create 27 individual outputs; one for each loudspeaker, as shown in Figure 7.

5.4 iPad App

As this array is ultimately intended for automotive applications, it was felt that a Graphical User Interface (GUI) was important for easy control of the array. Most modern vehicles have touchscreen

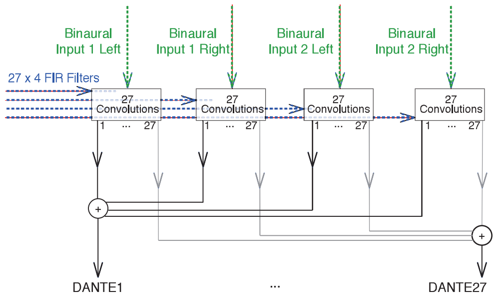

Figure 7: Software System Schematic.

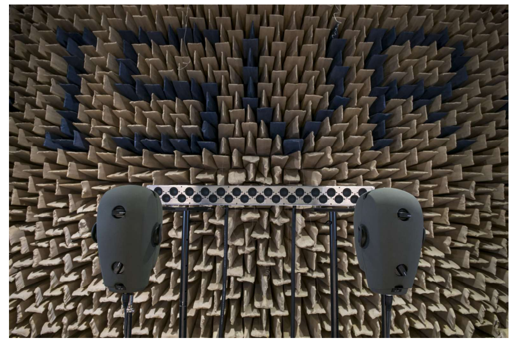

Figure 8: Measuring Array Performance.

infotainment systems for this purpose, and therefore an iPad control surface was developed using the MaxMSP MIRA protocol,[17] as shown in Figure 9.

This control surface allowed both users to easily switch between input channels and adjust levels (both overall and in each zone), as well as including output metering and filter selection controls.

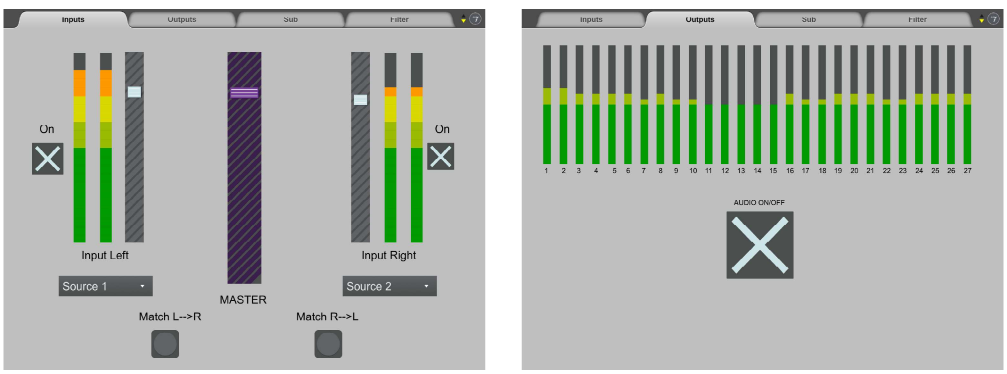

Figure 9: Screenshots of the iPad Control App.

6 COMPLETED SYSTEM & CONCLUSIONS

6.1 Measured Performance

The results in Figure 10 show the performance of the array in both anechoic and car environments, using the CTC filters designed from the responses measured in the anechoic chamber, and a constant frequency independent regularisation value. These filters were objectively found to provide the highest level of CTC, and subjectively provided the most convincing binaural experience.

It can be seen that between 300Hz and 11kHz, the system is achieving 25-45dB of CTC under anechoic conditions and 20-40dB of CTC in the car environment. In both cases, this is generally considered to be adequate CTC performance for high quality transaural reproduction. It is also clear that the designed low-frequency roll off of 250Hz discussed previously has been realised, and the measured filters equalise the response of the desired signal to avoid unwanted coloration of the sound.

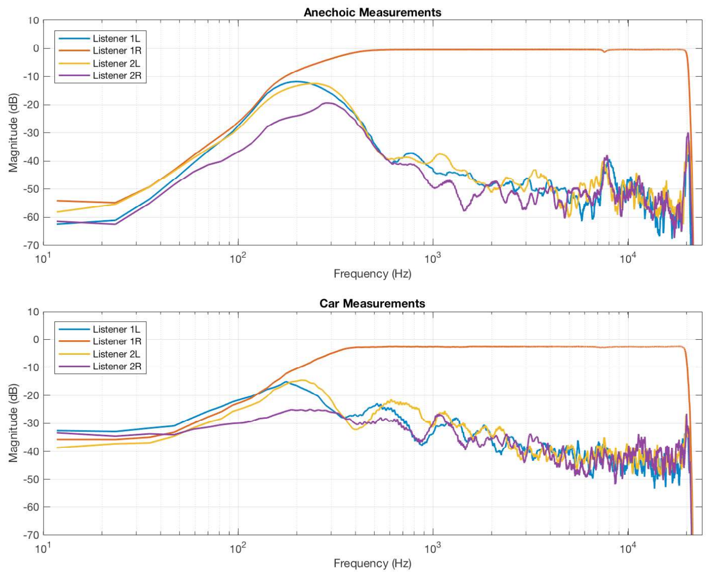

Figure 10: Measured CTC Performance.

6.2 Informal Subjective Testing

Although a full-scale formal listening test has not been conducted, informal subjective testing was carried out to support these measurements, and with the same binaural signal being played to both zones a convincing binaural experience was realised. However it was noted that in the car environment, if different content was played to each zone the crosstalk can become distracting. If simple stereo playback is required, rather than binaural, and the two audio streams are similar in terms of level and spectral content then the cancelation between zones is acceptable, however the strong reflections present in the car make it very difficult to reproduce two independent binaural experiences.

6.3 Conclusions & Further Work

A loudspeaker array has been designed to create four separate audio zones in the front of a car; enabling the driver and front passenger to have simultaneous binaural reproduction with separate programme material. Novel manufacturing techniques have been used to build the array, and real-time DSP has been implemented using MaxMSP.

Currently the array requires the listener's head to be positioned very precisely - the sweet spot is limited and therefore the binaural reproduction can break down quickly with head movement. The introduction of head tracking such that the filters adapt in real-time to both listeners' head positions would decrease this issue, and create a more robust system.

MaxMSP was used for all DSP tasks, however these could be implemented more efficiently if the convolutions were implemented using a lower-level programming language such as C. This would reduce the CPU load and could allow for longer filters, higher sampling rates, or just a cheaper processor.

The filter design process has a number of areas that could be improved upon. Frequency dependent regularisation was briefly investigated during this project, however further work into optimising the coefficients to minimise the matrix condition number and improve both robustness and performance could be beneficial.

It can be seen from the Figure 10 that the loudspeaker array has reasonable CTC performance from approximately 500 Hz upwards. As it is primarily high frequencies that humans use to localise sound, this performance may be acceptable and through informal listening has been shown to provide a realistic binaural experience. It has been shown, however, that headrest based loudspeakers are a highly effective and efficient transaural system,[18] with previous work at the ISVR[19] finding that the optimal frequency range for a headrest based transaural system is approximately 200 Hz to 1 kHz. Although headrest speakers are effective above this, head scattering begins to have a significant impact which causes a non-flat frequency response at the ears.

The combination of a loudspeaker array, as discussed here, high pass filtered to only reproduce sounds above approximately 1 kHz, and a headrest based transaural system as discussed in,[18,19] band-passed to only reproduce sounds between 200 Hz and 1 kHz, may provide a highly effective system with a flat CTC performance over a wider bandwidth.

A system such as this would be very effective, but may have limited applications for music reproduction because of the poor low-frequency performance. A low-frequency control system as described in previous personal audio publications[6,11] could be implemented to improve this.

References

[1] B&W. (2015) BMW 7 Series. [Online]. Available: http://www.bowers-wilkins.co.uk/car-audio/car-audio/BMW/7-Series.html

[2] V. Mischi. (2017) Seven of the best in-car systems from the Geneva Motor Show 2017. [Online]. Available: https://www.whathifi.com/news/7-best-in-car-systems-geneva-motor-show-2017

[3] Burmester. (2015) Burmester high-end 3D surround sound system. [Online]. Available: http://www.porsche.com/international/models/panamera/panamera/comfort-audio/burmester-3d-high-end-surround-sound-system/

[4] B&O. (2015) Bang & Olufsen 3D advanced sound system. [Online]. Available: http://www.bang-olufsen.com/en/solutions/automotive/audi/q7

[5] V. Bodelius. (2016) Bang & Olufsen and Audi just took high fidelity car audio to a whole new level - in 3 dimensions. [Online]. Available: http://nordic.businessinsider.com/with-a-brand-new-three-dimensional-sound-system-bang–olufsen-and-audi-now-set-new-standards-for-high-fideliy-audio-in-cars-2016-5/

[6] Harman. Individual sound zones. Harman. [Online]. Available: http://www.individualsoundzones.com/

[7] N. Odantzis, “Why the future of in-car audio will save ears and lives.” 10 2015. [Online]. Available: http://www.t3.com/features/why-the-future-of-in-car-audio-will-save-ears-and-lives

[8] W. Cunningham. (2016) Bose NEAR makes your car sound safe. CNET. [Online]. Available: https://www.cnet.com/news/bose-near-makes-your-car-sound-safe/

[9] “Philips carstudio car audio system,” Tech. Rep. [Online]. Available: https://www.download.p4c.philips.com/files/c/ce139dr_05/ce139dr_05_pss_enggb.pdf

[10] X. Liao, J. Cheer, S. Elliott, and S. Zheng, “Design of a loudspeaker array for personal audio in a car cabin,” *Journal of the Audio Engineering Society*, vol. 65, no. 3, pp. 226–238, March 2017. [Online]. Available: https://eprints.soton.ac.uk/403236/

[11] J. Cheer, S. J. Elliott, and M. F. S. Glvez, “Design and implementation of a car cabin personal audio system,” *J. Audio Eng. Soc*, vol. 61, no. 6, pp. 412–424, 2013. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=16832

[12] M. R. Bai, J.-G. Ih, and J. Benesty, *Acoustic array systems: Theory, implementation, and application*. United States: Wiley-Blackwell, 12 2014.

[13] M. F. Simón-Gálvez and F. M. Fazi, “Room compensation for binaural reproduction with loudspeaker arrays,” in *European Acoustics Association*, 06 2016.

[14] M. F. Simó-Gálvez and F. M. Fazi, “Loudspeaker arrays for transaural reproduction,” in *International Congress on Sound and Vibration*, Florence, Italy, 07 2015, pp. 1–8.

[15] J. Francombe, R. Mason, M. Dewhirst, and S. Bech, “Determining the threshold of acceptability for an interfering audio programme,” in *Audio Engineering Society Convention 132*, Apr 2012. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=16277

[16] Innosonix. (2017) Innosonix MA32-LP. [Online]. Available: http://www.innosonix.de/ma32-lp.html

[17] (2017) Max Mira. [Online]. Available: https://cycling74.com/products/mira/

[18] S. Elliott, C. House, J. Cheer, and M. Simón-Gálvez, “Crosstalk cancellation and equalisation for headrest sound reproduction,” in *Soundfield Control*. Audio Engineering Society, 07 2016.

[19] C. House, “Head tracking for local sound reproduction in a seat,” BEng Thesis, Southampton, ISVR, 05 2016.