Experimental validation of stereo dipole systems inside car cockpits

A. Bellini ◊, A. Cannizzaro ◊, G. Cibelli ⋆, E. Ugolotti ⋆, A. Farina ∗∗, C. Morandi ◊

◊ Dip. di Ing. dell'Inform. - Università di Parma - Parco Area delle Scienze, n.181A, I-43100 Parma - ITALY

⋆ ASK Industries S.p.A. - Via F.lli Cervi, 79, I-42100 Reggio Emilia - ITALY

∗∗ Dip. di Ing. Industriale - Università di Parma

1 Introduction

Sound reproduction within a car inside is a difficult task. Reverberation, reflection, echo, noise and vibration are some of the issues to account for. A first step in the direction of increasing sound comfort is that of equalizing the acoustic pressure response in the frequency domain. To accomplish this task the inversion of the measured Sound Pressure Level (SPL) should be performed [1]. Therefore two main step are involved. The former is composed by the acoustic characterization of the target car cockpit, in terms of amplitude and phase of the acoustic pressure. The latter is based on the synthesis of the filter which accurately reproduces the desired shape with a finite number of taps, and on the implementation of the audio processor on a DSP platform. Another issue to be considered is the harmonization of the sound image, namely the nice balancing of sounds coming from different instruments, and the increase of low frequencies, an option which is necessary to compensate running car noise, as taught by experience. Research on the acoustical and psychological bases of human sound localization suggests that the primary acoustical cues used to determine sound source position are the interaural differences in the time of arrival of a sound wave at the listener ears and the interaural difference in overall density of the sound. Another critical issue is the cross-talk which is the reproduced sound at a location where it was not intended to be heard. For instance the sound emitted from the right loudspeaker and heard at the left ear is a cross-talk. In order to eliminate this disturbance a common solution is to filter both channel, so that undesired sounds are compensated. Most researchers have concentrated on systems using loudspeaker arrangements spanning an angle of about 60 degrees as seen by the listener. However recently it was found that a system using two closely spaced loudspeaker is surprisingly robust with respect to listener movements. In this framework we define "stereo dipole" system such a virtual imaging system where the inputs to the two closely spaced loudspeakers are approximately out of phase over a wide frequency range, (i.e. similar to a point dipole source) [2]. To accomplish this task a digital processing structure can be developed using commercial DSPs.

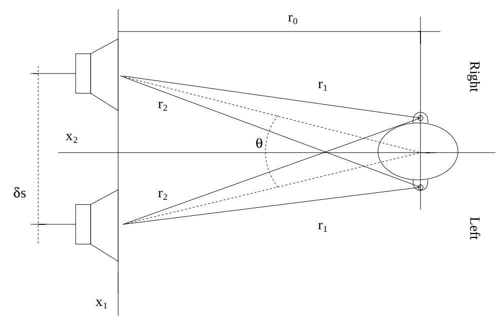

Figure 1: Stereo system topology.

Provided that a complete cross-talk cancellation is unrealizable, it is far easier for the loudspeakers to create the signal as if their virtual source was placed in a single point. It was noticed that better subjective results can be obtained processing the signal so that a virtual source appears in front of the listener, at a distance so that sound spatiality and harmonization is preserved.

Binaural localization of sound is allowed by the stereophonic conformation of human hearing system, exploiting the different sound intensity and delay (phase) at the different ears. The sound intensity is higher at the ear nearest to the source, while it is lower and delayed at the other, due to the masking of the head itself and to the longer path. These effects depend on the frequency of the reproduced sound. At low frequencies there is a critical threshold, at which the listener is blind in stating the sharp direction of sound, since the sound wavelength is longer than the distance between the ears. The system geometry is depicted in fig. 1.

cated in a symmetric way with respect to the listener, which is in front of them, and r_0 is the distance to the source. Two microphones symmetrical with respect to axis x_2 mimic ears behavior. Loudspeakers angle with respect the the listening direct axis is θ, and the shorter path is referred to as direct path r_1, while the other as cross-talk path r_2.

In the following subscripts are used to state variable characteristics. Namely r means that the variable under consideration is referred to the right channel, l to the left channel, rr to the right channel versus the right one, rl to the right channel versus the left one, lr to the left channel versus the right one, ll to the left channel versus the left one, lv to the left virtual source, rv to the right virtual source.

A generic system can be represented by the stereo input signals $x_l(t)$ and $x_r(t)$, which are the electric signals which drive the loudspeaker in open field, and $y_l(t)$ $y_r(t)$, which are the electric signals measured at microphones left and right respectively. Their Fourier transformation are indicated with caps letters and the following relationships hold.

$$\begin{bmatrix} Y_r \\ Y_l \end{bmatrix} = \begin{bmatrix} C_{rr} & C_{lr} \\ C_{rl} & C_{ll} \end{bmatrix} \times \begin{bmatrix} X_r \\ X_l \end{bmatrix} \tag{1}$$

where $[Y_r,\ Y_l]$ is the signal at microphones, and $[X_r,\ X_l]$ is the signal at loudspeakers. The transmission channel, which characterizes the listening environment, can be modeled with the C matrix. Assuming that geometric and physical symmetry are respected the matrix can be simplified with $C_{ll}(f) = C_{rr}(f)$ and $C_{lr}(f) = C_{rl}(f)$.

Notice that, since we are considering binaural sound, even with a single sound source S the signal at the two microphones is different. In fact, if $x_r(t) = 0$, we have

$$Y_l(f) = X_l(f) * C_{ll}(f)\ ;\ Y_r(f) = X_l(f) * C_{lr}(f) \tag{2}$$

where $C_{lr}(f)$ is the transfer function of the cross-talk channel.

2 Cross-talk cancellation and virtual sound source

In order to achieve an exact cross-talk cancellation the signal of a channel must be perfectly reproduced by the corresponding microphone, and completely neglected by the other, i.e. each channel is completely unbounded from the other. In general the cross-talk condition for the left channel can be expressed forcing the signal at microphones to be equal to

$$\begin{bmatrix} 0 \\ Y_l \end{bmatrix} e^{\frac{-j2\pi f r}{c_0}} \tag{3}$$

after anti-transformation.

Forcing the (3) in (1) we have:

$$\begin{bmatrix} 0 \\ Y_l \end{bmatrix} e^{\frac{-j2\pi f r}{c_0}} = \begin{bmatrix} C_{rr} & C_{lr} \\ C_{rl} & C_{ll} \end{bmatrix} \times \begin{bmatrix} X_r \\ X_l \end{bmatrix} \tag{4}$$

If we assume that S is the mono signal which drives the filters we have

$$X_r = S * H_r\ ;\ X_l = S * H_l \tag{5}$$

and replacing in (4) we obtain

$$S(C_{rr}H_r + C_{lr}H_l) = 0 \tag{6}$$

$$S(C_{rl}H_r + C_{ll}H_l) = Y e^{\frac{-j2\pi f r}{c_0}} \tag{7}$$

This relationship allows to determine the transfer function of the filters H_l and H_r to be added in order to achieve cross-cancellation, as a function of the transfer channels considered [3]. Considering a virtual sound source as sketched in fig. 2, $y_{lv}(t)$ and $y_{rv}(t)$ are the electric signals at microphones produced by this artificial source, and $Y_{lv}(f)$, $Y_{rv}(f)$ their Fourier transformations. The aim of our digital system is to design a filter structure H so that the sound at microphones is equal to that of the virtual source even if it is produced by the loudspeaker system depicted with solid lines in fig. 2. To this aim the following constraints are forced

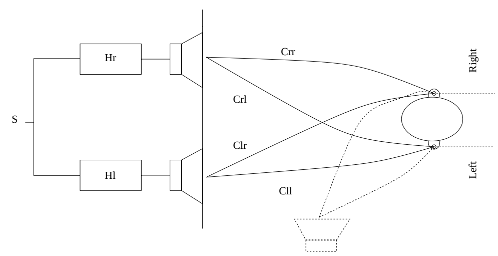

Figure 2: Virtual sound imaging topology.

$$Y_l(f) = Y_{lv}(f)\ ;\ Y_r(f) = Y_{rv}(f) \tag{8}$$

Therefore we obtain

$$\begin{bmatrix} Y_{rv} \\ Y_{lv} \end{bmatrix} e^{\frac{-j2\pi f r}{c_0}} = \begin{bmatrix} C_{rr} & C_{lr} \\ C_{rl} & C_{ll} \end{bmatrix} \times \begin{bmatrix} X_r \\ X_l \end{bmatrix} \tag{9}$$

The solution of the relationship (9) produces the desired shape of the filters H_r and H_l.

2.1 Virtual source with head related transfer function

Accurate measurements and implementation of the virtual sound system described in the previous section allow the reproduction of a virtual sound source, movements of the source. However different listeners have different sound sensations and their hearing system affects the effectiveness of this process. Some listeners could be more sensititive to the position of the virtual image, therefore a more accurate model should be used. A real listener can be considered and

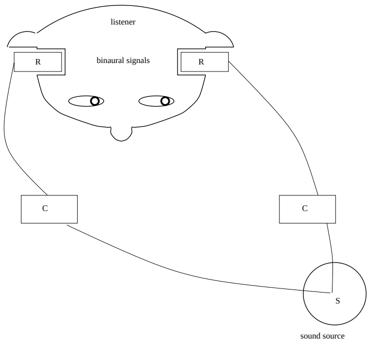

Figure 3: Stereo dipole topology accounting for Head Related Transfer Functions.

described as in fig. 3, and the head related transfer functions can be introduced. With a monoaural sound source S the binaural signals Q at the eardrum are

$$\begin{bmatrix} Q_r \\ Q_l \end{bmatrix} = \begin{bmatrix} C_r & 0 \\ 0 & C_l \end{bmatrix} \times \begin{bmatrix} R_r & 0 \\ 0 & R_l \end{bmatrix} \times \begin{bmatrix} S \\ S \end{bmatrix} \quad (10)$$

where C as usual is the environment transfer function description, R is the matrix of the Head Related Transfer Function (HRTF), i.e. the transfer function of the hearing system.

In the audio frequency range the sound waves travel in the hearing channel as flat wave, therefore the matrix of the HRTF is independent of the source position [4]. The sound virtual system aims at reproducing as accurately as possible the signals Q (fig. 3) with two loudspeakers. The simplest way to accomplish this task is to record the real source with two microphones placed within a dummy head.The effect of the transducers will be present even in the recorded signals, and can be modeled by the M matrix. Therefore the signals F after the microphones can be expressed as

$$\begin{bmatrix} F_r \\ F_l \end{bmatrix} = \begin{bmatrix} C_r & 0 \\ 0 & C_l \end{bmatrix} \times \begin{bmatrix} R_r & 0 \\ 0 & R_l \end{bmatrix} \times \begin{bmatrix} M_r & 0 \\ 0 & M_l \end{bmatrix} \times \begin{bmatrix} S \\ S \end{bmatrix} \quad (11)$$

The topology of the system consistent with these specification is reported in fig. 4. The signals $F(f)$ recorded by the dummy head with the described configuration are processed by the digital filters $H(f)$.

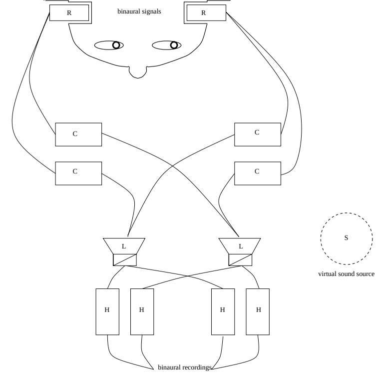

Figure 4: Virtual sound system topology.

The vector so obtained is multiplied by the diagonal matrix $L(f)$ which contains the electroacustics transfer function of the transducers. Moreover the transfer functions of the transmission channel $C(f)$ and of the hearing channel $R(f)$ are considered. In summary at the eardrum of the listener we obtain

$$\begin{bmatrix} W_r \\ W_l \end{bmatrix} = \begin{bmatrix} H_{rr} & H_{lr} \\ H_{rl} & H_{ll} \end{bmatrix} \times \begin{bmatrix} C_{rr} & C_{lr} \\ C_{rl} & C_{ll} \end{bmatrix} \times \begin{bmatrix} R_r & 0 \\ 0 & R_l \end{bmatrix} \times \begin{bmatrix} L_r & 0 \\ 0 & L_l \end{bmatrix} \times \begin{bmatrix} F_r \\ F_l \end{bmatrix} \quad (12)$$

Where the H matrix represent the filters needed to obtain the desired virtual sound source accounting for HRTF. The shape of the filters can be computed forcing that the signal at the eardrum is equal to that of the virtual source and measuring L, R and C.

$$W_l(f) = W_{lv}(f) \; ; \; W_r(f) = W_{rv}(f) \quad (13)$$

Then the system transfer function of the transmission channel $C(f)$ is measured using the same transducers and microphones used for binaural recording and through Aurora MLS signal, processing the time domain impulse response. A key point is the angle θ between the loudspeaker (fig. 1), which must be around 10 degrees for an optimal performance [5].

3 Digital Implementation of stereo dipole

The hardware platform adopted for a prototype implementation of stereo dipole system is the DSP Analog Device EZ-LITE which relies on the processor AD21061. It features a clock speed of 40 MHz, with 120 MFLOPS of peak performance, 64Kx32 bit on-chip Dual Access RAM and 32 bit registers. Moreover the prototype development board EZ-LITE is can sample analog signals from 5.5 KHz up to 48 KHz. The designed stereo dipole architecture samples input signals at 44.1 KHz and is driven by electric signals on the range of 750 mV, which are the audio CD standard values, and its gain is of 0 dB.

As discussed in the previous section it is possible to design a filter structure H_{rr}, H_{lr}, H_{rl}, H_{ll}, which realizes the desired stereo dipole system, fig. 5. Namely, in order to respect real-time constraints

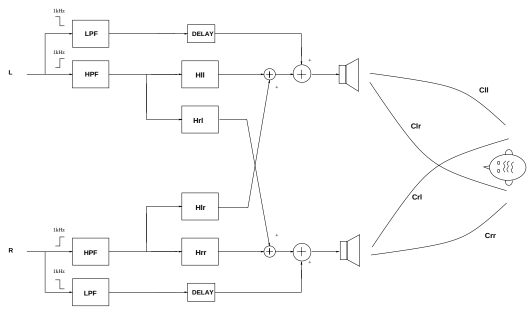

Figure 5: Block diagram of a stereo dipole system.

the diagram sketched in fig. 9 is modified adopting a "multi-filter" architecture, where the signal is divided in the low-pass and high-pass components. In fact human earing system sensitivity to phase delay can be grossly estimated, assuming that signals whose delay is around 100 μ s are considered coherent. The high-pass part of the signal is processed by the filter, while the low-pass part is only delayed and summed to the other, but the gain of the low-pass filter used to split the signals is 3 dB. This is made to increase the low frequencies, a feature which is very useful to mask running car noise. The shape of the filters described in the previous section is synthesized by a set of FIR filters.

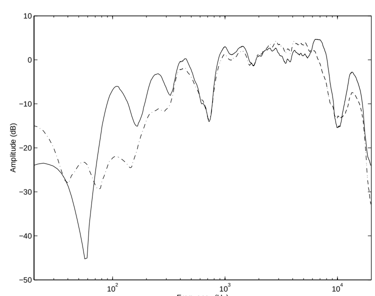

Figure 6: Acoustics response within the car with B&K dummy head with (solid line) and without (dashed line) stereo dipole system.

The choice of the cut-off frequency of the low-pass filter is a critical issue, since the spectrum below

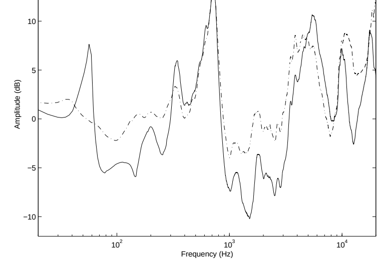

Figure 7: Cross-talk within the car with B&K dummy head with (solid line) and without (dashed line) stereo dipole system.

this limit is kept unchanged. Therefore several experimental tests were performed adopting different choices, namely 100 Hz, 300 Hz, 1000 Hz. The former resulted to be the best trade-off between performance and cost. The shape of the filters H is computed relying on the time impulse responses of the transmission channels $c_{rr}(t)$, $c_{lr}(t)$, $c_{rl}(t)$, $c_{ll}(t)$ and on the relationship (12). This relationship however provides only two constraints, while four filters must be computed. A possible choice is to determine the shape of the H_{rr}, H_{ll} filter, assuming that they are flat on all the audio frequency range. On the other hand, a more accurate procedure could be investigated, in order to further improve sound sensation and equalization. Namely for car cockpits it was experienced that an improvements of the low-part contents is desirable and a more fitted comfort curve is preferable [6]

As detailed in the previous section acoustics measurements are needed to compute the shape of the H filters. The transmission channels C_{ll} and C_{lr} measured in a Citroën Evasion cockpit by means of the AURORA toolbox with MLS signal [7]. The MATLAB signal processing toolbox was used to synthesize the computed shape of the filters. In order to perform subjective tests a Brüel & Kiaer 4100 dummy head was used, equipped with Falcon 4190 $50mv/Pa$ microphone and suitable pre-amplifier.

In the car environment a prototype loudspeaker system was used, where dedicated loudspeakers are used and placed in front of the car driver, so that the θ angle is around 10 degrees, which is the optimal value.

The measurements performed with the dummy head allow to compute the HRTF too, therefore a more accurate stereo dipole system can be synthesized. Fig. 6 reports the frequency-domain measurements of the car cockpit, without and with the stereo dipole system, while in fig. 7 cross-talk capability is sketched: the left channel amplitude is lower when signal is only on the right channel.

Subjective tests were performed too, recording the answers of ten listeners to 7 questions, concerning sound quality, localization, and harmonization. The value reported is the mean score, where 8 correspond to the maximum evaluation (fig. 8). A general increase of the score occurs with pre-elaborated sound (white histogram), and this confirms the effectiveness of the proposed approach. The test was composed by the evaluation of the following quality benchmarks: Q1 Initial sound sensation; Q2 musical scene location (voice and instrument position sharpness); Q3 Amplitude of sound front; Q4 Natural sound reproduction ; Q5 Low frequency response; Q6 Medium frequency response; Q7 High frequency response.

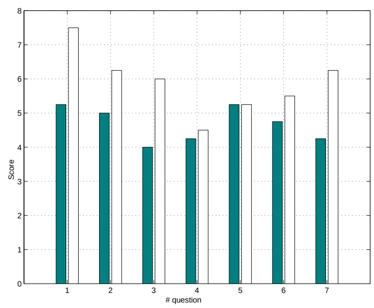

Figure 8: Subjective test results.

4 Conclusion

The stereo dipole system was implemented on a 21061 EZ-LITE DSP and aims at achieving a pleasant sound image harmonization. The following items were accomplished: definition of a model for the “stereo dipole system”, and its implementation with a DSP, characterization of the HRTF, definition of an automatic procedure to synthesize suitable filters, setting up of a “pair matched” Hi-Fi system, when symmetrical conditions are required, setting up of an “automotive oriented” Hi-Fi system, with asymmetric loudspeaker systems.

Listening tests were performed on a car inside using a pre-computed sound track. The car audio system was adapted inserting two adjacent loudspeakers for reproducing frequencies above 1 kHz, in the front of the car driver position. The stereo dipole system was implemented in a single 21061 EZ-LITE for both stereo channels, and used to produce real-time elaboration of the usual audio source.

The subjective results were effective and consistent with the expected behavior. Namely the sound monic. The measurements performed however do not account for the complex interaction between different harmonics and the which a real sound reproduction involves. Moreover relying on the usual commercial CD recording it is difficult to discern the effective recording structure adopted. Therefore it is hard to distinguish between monoaural, binaural and multitrack recording. Therefore since the boundary conditions are not well defined, the analytical problem does not have a fixed solution. The CD recording is probably a monoaural one, the adoption of a stereo dipole structure however produces an increase of the sound spatiality sensation. Every listener has agreed with the marked increase of sound spatiality, especially of accurate DDD audio tracks are concerned. As for car cockpits a marked increase of frequency resolution is reported by listeners. Moreover the gain of the low-pass filter is very effective in increasing the pleasance of bass to mask running car noise.

Future developments could be in the area of designing pre-elaboration filters which account for acoustics reflections and delays, which can further increase the sound quality.

References

[1] A. Bellini, G. Cibelli, E. Ugolotti, A. Farina, and C. Morandi, “Non-linear digital audio processor for dedicated loudspeaker systems,” *IEEE Trans. on Consumer Electronics*, pp. 1024–1031, Aug. 1998.

[2] O. Kirkeby and P. A. Nelson, “Local sound field reproduction using two closely spaced loudspeakers,” *Journal Acoustic Society of America*, vol. 104, October 1998.

[3] O. Kirkeby and P. Nelson, *Local sound field reproduction using two closely spaced loudspeakers.* New–York: Acoustical Society of America, 1998.

[4] P. Nelson, “Influence of individual head related transfer function on the performance of virtual acoustic imaging systems,” *Preprint of the AES*, vol. 104, May 1998.

[5] O. Kirkeby and P. Nelson, *The stereo dipole- a virtual sourse imaging system using two closely spaced loudspeakers.* New–York: Acoustical Society of America, 1998.

[6] G. Cibelli, A. Bellini, E. Ugolotti, A. Farina, and C. Morandi, “Experimental validation of loudspeaker equalization inside car cockpits,” *Preprint 4898 of the AES*, vol. 106, May 1999.

[7] A. Farina and F. Righini, “Software implementation of an mls analyzer, with tools for convolution, auralization and inverse filtering,” in *Preprint of the AES*, vol. 103, (New York), pp. 26–29, September 1997.