IMPLEMENTATION OF A GENERAL-PURPOSE AURALIZER ON A FLOATING-POINT SIGNAL PROCESSOR

Juha Merimaa and Matti Karjalainen
Laboratory of Acoustics and Audio Signal Processing, Helsinki University of Technology,
P.O.Box 3000, FIN-02015 HUT, FINLAND
Tel: +358 9 4512496; fax: +358 9 460224
`e-mail: Juha.Merimaa@hut.fi and Matti.Karjalainen@hut.fi`

ABSTRACT

The design and implementation of a real-time general-purpose auralization device is presented in this paper. The goal has been to make a stand-alone device that can also be controlled from a PC. Several auralization and spatialization methods are included, such as binaural headphone and loudspeaker auralization, and vector base amplitude panning for multi-channel loudspeaker sound reproduction. Combining these to a single device allows easy comparison and choosing the best method for any purpose. The different auralization principles and their implementations using the SHARC digital signal processor are described.

1 INTRODUCTION

Perception of spatial attributes of sound is an important part of the hearing process. The term localization means the human ability to relate the location of a sound source to some attributes of a sound event. The goal of auralization is to simulate these attributes by adding localization cues to a sound signal.

Directional localization relies mainly on two frequency dependent cues: interaural level difference (ILD) and interaural time difference (ITD). Distance localization, on the other hand, is much more complicated and usually requires that the listener is familiar with the sound signal and the current acoustical environment. [2]

Directional auralization can be implemented with several different methods. The most straightforward way to produce localization cues is to use filters approximating measured head-related transfer functions (HRTF). For best auralization, individual HRTFs of a subject should be used. In practice these are seldom available. Satisfactory auralization can, however, be attained with dummy-head HRTFs [12].

HRTF filtering results in a binaural signal, i.e., a separate channel for each ear. The easiest way to reproduce these channels is to use headphones and some equalization. A binaural signal can also be reproduced with two loudspeakers but additional crosstalk canceling is needed to compensate for the sound travelling from each loudspeaker to both ears. Crosstalk canceled binaural auralization can also be called transaural [3].

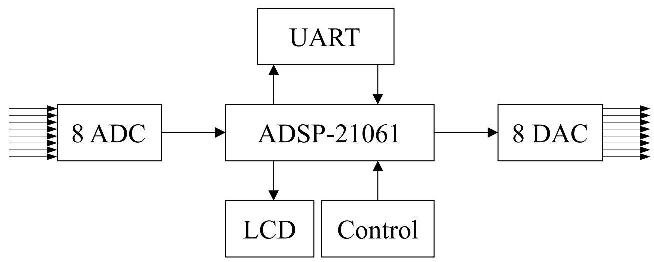

Figure 1: A block diagram of the hardware.

Amplitude panning is a different auralization and spatialization method. It creates localization cues by applying the same signal to two or more loudspeakers with different amplitudes. Vector Base Amplitude Panning (VBAP) allows two- or three-dimensional auralization using any number of arbitrarily placed loudspeakers [9]. Amplitude panning restricts the virtual source directions to the area between loudspeakers but it works on a larger listening area than binaural loudspeaker auralization.

This paper describes a real-time general-purpose auralizer implemented on the Analog Devices [1] ADSP-21061 (SHARC) digital signal processor. The objectives of the project were to create a stand-alone 8-channel device that is capable of using any of the previously mentioned auralization methods.

2 HARDWARE

The hardware is based on an EZ-KIT Lite evaluation board from Analog Devices. The board has been modified to have 8 input and output channels and a user interface with an LCD-display as well as push buttons to control the auralization. Additionally, the device can be controlled with a PC through its serial port. The basic hardware is illustrated in figure 1.

The ADSP-21061 EZ-KIT Lite board operates on a 40-MHz clock frequency. Without pipeline conflicts the processor can execute one instruction on every clock cycle. Thus with a sampling rate of 44.1 kHz, about 900 instructions are available per each set of eight input samples.

3 HRTF FILTERING

Every stable filter can be decomposed into a minimum-phase filter and an all-pass filter. The minimum-phase filter has the same magnitude response as the original filter with a minimal phase delay and the all-pass filter realizes the "excess phase". In the case of HRTFs, the excess phase is almost a linear function of frequency and thus can be approximated with a pure delay. The ITD error introduced in the approximation has been shown to be perceptually negligible [8].

This kind of filter decomposition has two significant advantages. First, if the delay is rounded to the nearest sample, it can be realized efficiently with a simple delay line. With a sampling rate of 44.1 kHz the maximum ITD error resulting from the rounding corresponds to an azimuth error of less than 3° [6] which is less than the localization blur of an average listener.

The second advantage is related to interpolating measured HRTFs. In a FIR implementation the interpolation can be easily performed by linear combination of adjacent HRTF impulse responses. With mixed-phase filters this corresponds to linear amplitude panning between two or more sound sources and may result in notable comb-filtering effects. With minimum-phase filters, however, coefficent interpolation is equivalent to interpolating in the frequency domain and doesn't produce comb-filtering effects. [7]

Multi-channel HRTF filtering is computationally the most demanding part of the binaural auralization. To avoid any unnecessary overhead, the filtering algorithm has been carefully programmed using the SHARC assembly language. The directional filters are implemented with minimum-phase FIR filters and frequency-independent delays with a sampling rate of 44.1 kHz.

The common overhead of auralizing all channels is no more than about 20 instructions. Crosstalk canceling or headphone compensation (see sections 4 and 5) require at most a little more than 100 instructions. This leaves about 750 instructions or 90 per channel for directional filtering. Of these about 70 can be used for FIR implementation and thus filters of order of 35 can be used. For dummy-head HRTFs, an order of 40 has been shown sufficient for 75% of people to find no difference when compared to full-length filters [5].

The directional filters are based on KEMAR dummy-head HRTFs [4]. The auralizer's HRTF database consists of altogether 77 filter pairs with 10 degrees azimuth intervals and elevations of -30, 0, 30, 60 and 90 degrees. The database size has been reduced by assuming the head to be symmetrical. This enables the auralizer to create left-side virtual sources with right-side HRTFs by swapping the output channels. Intermediate directional filters are interpolated linearly from four adjacent directions. Individual HRTFs can also be used by downloading them from a PC.

4 HEADPHONE COMPENSATION

In order to sound similar to loudspeaker, commercial headphones are equalized to include the non-flat transfer charasteristics from loudspeakers to the listener's ears. Two equalization methods are generally used; In free-field equalization, sound of a loudspeaker directly in front of the listener in a free field is used as a reference. In the second method the reference signal is measured in a diffuse field. Diffuse-field equalization has generally been found to sound more natural. [11]

A binaural signal already includes the transfer characteristics from a sound source to the listener's ears. Thus the effect of headphone equalization has to be compensated with an additional filter. The auralizer includes an inverse diffuse field filter for this purpose. In addition, custom filters are provided for some headphone models and the user can download his/her own filters from a PC. Headphone compensation is implemented with FIR filters of the order of 50. Similar filters are used for both ears.

5 CROSSTALK CANCELING

Using stereo loudspeakers for binaural sound reproduction introduces some additional problems. In a normal listening situation, the sound from both loudspeakers arrives at both ears. In the case of binaural auralization we would like to be able to reproduce completely independent signals to the listener's ears. This can be done with some extra filtering.

Let $H_i(z)$ denote the transfer function from either loudspeaker to the ear on the same side and $H_c(z)$ the transfer function to the ear on the opposite side of the loudspeaker in a symmetrical setup. The propagation of sound can be described with the following equations:

$$y(n) = H(z)\hat{x}(n) \tag{1}$$

where

$$y(n) = \begin{bmatrix} y_l(n) \\ y_r(n) \end{bmatrix}, \hat{x}(n) = \begin{bmatrix} \hat{x}_l(n) \\ \hat{x}_r(n) \end{bmatrix} \tag{2}$$

and

$$H(z) = \begin{bmatrix} H_i(z) & H_c(z) \\ H_c(z) & H_i(z) \end{bmatrix} \tag{3}$$

$\hat{x}_l$ and $\hat{x}_r$ are the loudspeaker signals and $y_l(n)$ and $y_r(n)$ the resulting signals in the listener's left and right ear, respectively. If $x_l(n)$ and $x_r(n)$ are the signals to be delivered to the listener's ears, an inverse matrix $G(z)$ must be found such that $G(z) = H(z)^{-1}$ and $\hat{x}(n) = G(z)x(n)$. It can be shown that $G(z)$ can be written as

$$G(z) = D \begin{bmatrix} \frac{1}{H_i(z)+H_c(z)} & 0 \\ 0 & \frac{1}{H_i(z)-H_c(z)} \end{bmatrix} D \tag{4}$$

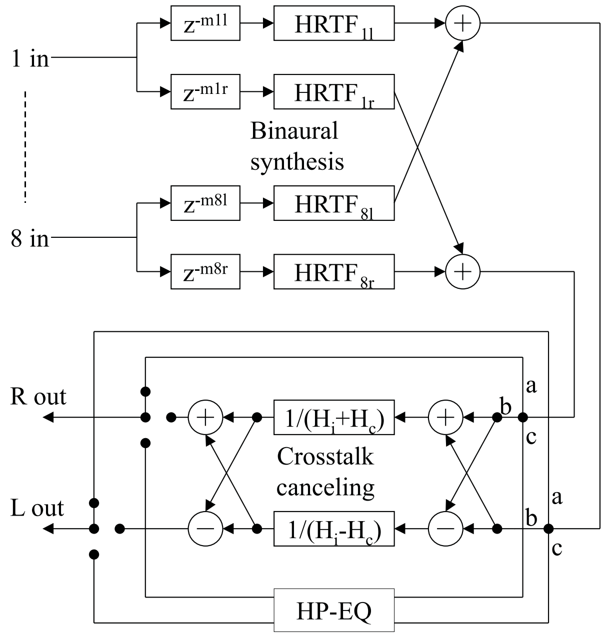

Figure 2: A block diagram of the binaural part of the auralizer. The binaural signal is optionally routed a) directly to the output b) through the crosstalk canceling filters or c) through headphone compensation filters.

where

$$D = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \tag{5}$$

This can be implemented with a shuffler structure filter illustrated in the crosstalk canceling part of figure 2. [6, 7] The figure represents the binaural part of the auralizer including HRTF filtering and headphone compensation. The normalization gains $1/\sqrt{2}$ of the crosstalk canceling haven't been included.

The sum and difference filters $H_i(z) + H_c(z)$ and $H_i(z) - H_c(z)$ have been shown to be of joint minimum-phase [3], that is, they have a common excess phase, and can be normalized to minimum-phase filters. Thus, stable inverse filters can easily be designed.

In an anechoic room, $H_i(z)$ and $H_c(z)$ equal to HRTFs of the loudspeaker directions. In a real room, however, using HRTFs is only a crude approximation. Since it is not possible to find generic transfer functions $H_i(z)$ and $H_c(z)$, the binaural loudspeaker auralization quality is always dependent on the acoustics of the listening room.

In the auralizer, the inverse filters have been implemented with 25th order IIR filters designed with Prony's method from $\pm 30°$ horizontal plane HRTFs. This seems to produce quite good localization for virtual sources in front of the listener but sources in the back are usually not localized correctly.

The listening area of this kind of auralization is restricted to a small "sweet spot". According to informal listening tests, the sweet spot can be enlarged and the localization and the timbre of virtual sources in unideal listening conditions can be improved by windowing the crosstalk canceling filters in the frequency domain to have unity gain at high frequencies. By substituting $H_i(z) + H_c(z) = H_i(z) - H_c(z) = 1$ in (4) this modification can be seen to reduce the crosstalk canceling to normal stereo reproduction of the high frequencies.

The improvements can be understood by considering the listener's small displacements from the sweet spot. At high frequencies the wavelength of sound is short and even small displacements can cause the canceling to turn into summing. This effect disturbs the localization and emphasizes high frequencies.

6 VBAP

In amplitude panning virtual sources in between loudspeakers are produced by applying the same sound signal to two or more loudspeakers at the same time. This kind of auralization doesn't produce as sharp virtual sources as binaural auralization at its best. However, amplitude panning works satisfactorily on a much larger listening area and worse acoustic environments than binaural loudspeaker auralization. The auralization accuracy is also dependent on the number of loudspeakers and their placement.

Vector Base Amplitude panning (VBAP) [9] is a three-dimensional vector base reformulation of the traditional tangent law panning. The idea is to represent a unit vector pointing to the direction of the desired virtual source as a linear combination of base vectors pointing to two or three loudspeakers depending on the dimensionality of the algorithm. The coefficients of the base vectors are used as gain factors for loudspeaker signals.

The VBAP implementation differs from other signal processing software in the project since it is written in C. The algorithm is computationally so efficient that the small overhead produced by the C-runtime environment can be allowed.

The algorithm uses separate modes for two and three-dimensional loudspeaker placements. In the 2D-mode, VBAP reduces to amplitude panning between two nearest loudspeakers on both sides of the desired virtual source. In the 3D-mode, each input signal is panned inside a loudspeaker triangle.

The triangles are chosen so that the formed virtual sources are always as sharp as possible. Optimal choices for non-overlapping triangles to cover all possible auralization angles [10] are calculated each time the loudspeaker configuration is changed. From these, an adequate triangle for any virtual source direction can quickly be found when the auralization directions are changed.

7 USER INTERFACE

The stand-alone user interface of the device consists of 8 push buttons and an LCD-display. The push buttons can be used to change the current auralization method and headphone compensation as well as to control the auralization directions and gains of the input channels and the directions of the loudspeakers used in the VBAP auralization.

The control software for PC is written in C++ and runs under Microsoft Windows. All the parameters of the auralizer can be changed from the PC and it is also possible to download customized filters to the device. The control of both user interfaces consumes some processor time on the SHARC, but it can be distributed to the short idle times of the processor between processing consecutive input samples.

8 FUTURE DEVELOPMENTS

Possible future developments of the auralizer include porting the code to a more efficient signal processor. This would allow the device to use more accurate HRTF approximations or more input channels. Analog Devices has already introduced the ADSP-21160M processor that includes two computational units, operates on a 100-MHz clock frequency and is code compatible with the ADSP-21061. Binaural auralization could also be improved by connecting a head tracker to the device. This would allow virtual source directions to be independent of listener's head movements.

9 CONCLUSIONS

A real-time general-purpose auralization tool has been designed and described. The auralizer combines several auralization and spatialization methods to a single stand-alone device. Headphone auralization has been implemented with minimum-phase FIR filters and frequency independent delays approximating measured dummy-head HRTFs.

The resulting binaural signal can optionally be routed through a headphone compensation filter or a shuffler structure crosstalk canceling system. In informal listening tests, the headphone auralization seems to produce quite accurate localization cues, especially in the horizontal plane. In binaural loudspeaker auralization, however, simulating virtual sources behind the listener is quite difficult. In both cases, individualized filters can improve the auralization quality.

The VBAP algorithm uses a different approach to auralization. Instead of trying to faithfully reproduce all localization cues, it utilizes amplitude panning. VBAP restricts virtual source directions to angles between loudspeaker pairs or inside loudspeaker triangles. Thus, a multi-channel sound reproduction system is needed to cover the same range of virtual source directions that can be attained with two channels in binaural methods. VBAP, however, has some distinct advantages compared to binaural loudspeaker auralization: the practical listening area is much wider and the listening room acoustics has less effect on auralization quality.

10 ACKNOWLEDGEMENTS

This work has been supported by Maanpuolustuksen tieteellinen neuvottelukunta (MATINE). The multichannel expansion of the DSP board was designed and implemented by Kari Hautio.

References

[1] Analog Devices, Inc. http://www.analog.com/.

[2] J. Blauert, Spatial Hearing. Revised Edition. Cambridge, Massachusetts: The MIT Press. 1997.

[3] D. H. Cooper, J. L. Bauck, "Prospects for Transaural Recording," J. Audio Eng. Soc., vol. 37, no. 1/2, pp. 3-19. 1989.

[4] B. Gardner, K. Martin, HRTF Measurements of a KEMAR Dummy-Head Microphone, MIT Media Lab Perceptual Computing, technical report #280. 1994.

[5] J. Huopaniemi, M. Karjalainen, "Review of Digital Filter Design and Implementation Methods for 3-D Sound," Proceedings of the 102nd Convention of the Audio Engineering Society, Preprint 4461, Munich, Germany. 1997.

[6] J. Huopaniemi, "Virtual Acoustics and 3-D Sound in Multimedia Signal Processing," Ph.D. Thesis, Helsinki University of Technology, Espoo, Finland. 1999.

[7] J-M. Jot, V. Larcher, O. Warusfel, "Digital signal processing issues in the context of binaural and transaural stereophony," AES 98th Convention, Paris, 1995. Preprint 3980.

[8] A. Kulkarni, S. K. Isabelle, H.S. Colburn, "Sensitivity of human subjects to head-related transfer-function phase spectra," J. Acoust. Soc. Am., vol. 105, no. 5, pp. 2821-2840. 1999.

[9] V. Pulkki, "Virtual Sound Source Positioning Using Vector Base Amplitude Panning," J. Audio Eng. Soc., vol. 45, no. 6, pp. 456-466. 1997.

[10] V. Pulkki, T. Lokki, "Creating Auditory Displays with Multiple Loudspeakers Using VBAP: A Case Study with DIVA Project," International Conference on Auditory Display, Glasgow, England. 1998.

[11] G. Theile, "On the Standardization of the Frequency Response of High-Quality Studio Headphones," J. Audio Eng. Soc., vol. 34, no. 12, pp. 956-969. 1986.

[12] E. M. Wenzel, M. Arruda, D. J. Kistler, F. L. Wightman, "Localization using nonindividualized head-related transfer functions," J. Acoust. Soc. Am., vol. 94, no. 1, pp. 111-123. 1993.