Stereo widening system using binaural cues for headphones

S M A Basha[1], Abhinav Gupta[2], Anshul Sharma[3],

[1,3]Advanced Technologies Division, Samsung India Software Operations Private Ltd, Bangalore-India.

[2]Software Lab 1, Digital Media, Samsung Electronics Corporation Limited, Gueonggi-Do, Suwon City, South Korea.

{ [1]basha.shaik, [2]gupta.abhinav, [3]anshul } @ samsung.com

***Abstract*-This paper proposes a novel system for achieving stereo widening using an efficient approach for headphones. Our system is based on centre signal to side signal ratio and uses suitable delay lines and tuneable gain parameters to explore binaural cues. An extensive study is performed on centre to side phantom image shift analysis for better audio externalization effect. The lateralization effect is highly minimized and appreciable externalization is obtained through implementation of IID, ITD and ICC binaural cues. This paper also proposes the use of one dominant early reflection circuit to generate spaciousness in the stereophonic sound. The results obtained are satisfactory and encouraging to proceed for real-time audio post processing applications for headphones.**

Keywords -
Binaural Cues, stereo widening, externalization, delay, headphones.

I. INTRODUCTION

Stereophonic headphones are common in our daily life. Stereophonic system is an audio device with two channels of information over headphones. In stereophonic reproduction the listener perceives sound images along a line between the headphones [1]. In general, stereophonic headphones create a spatial impression of 60 degrees. Exposure to this unnatural sound effect for prolonged period of time may eventually cause sound appearing from with-in (in-head localization) and fatigue [2]. Low frequency components in almost any music (e.g. bass guitar) always appear as coming from inside the head. This in-head localization is unique to recorded music.[3] Several solutions have been proposed by researchers to externalize the sound with the use of binaural cues[4] and the research is still underway. While evaluating the widening effect, the broader externalization of the audio phantom image and minimum spectral coloration are major factors to be considered.

II. APPROACH AND METHODS

Basic stereo widening can be achieved by generating decorrelated signals in the left(L) and right (R) channels with the help of binaural cues by using phase shift network or delay network [5]. Previous attempts to improve headphone listening have been made by reversing the phase of stereo channels and introducing the delays between them of the order of 1 ms [2]. A majority of existing schemes use a decorrelator and crossfeed network to explore ITD to widen the sound image. In this paper we propose a new approach which uses dominant binaural cues like ILD and ICC along with ITD for realistic sound. The proposed stereo widening scheme involves boosting of side signal (i.e., L-R, difference signal) and reducing the correlation between the two channels. Our system comprises of three modules, Difference circuit, Crossfeed circuit and the Early Reflector circuit. The Difference circuit boosts the side signal in order to make the side components of the L and R signals more prominent. The crossfeed circuit simulates the loudspeaker environment over headphones using inter-channel time difference (ITD) and inter-channel level difference (ILD). The early reflector circuit extracts one dominant early reflection and adds it back to the original signal with different delay lengths for both the channels. This helps to lower the inter-channel coherence (ICC) by making the signals out of phase. The detailed description of all blocks is given in section V. Our proposed model reduces the in-head localization and widens the stereo phantom image towards the left and right ears. This algorithm is expected to be useful as an audio post processing unit and could fit inside the headphones or can be used independently for a rich widened stereo listening experience.

III. BINAURAL CUES

Binaural cues are the parameters that measure the extent of the 'spaciousness' of the sound. ILD, ITD, and ICC are the three major binaural cues which play an important role in the localization and externalization of the sound.

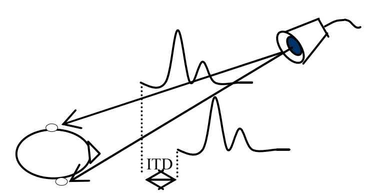

Fig.1. Inter-Channel Time Difference

The time delay between the sounds arriving at the left and right ears differs depending on the location of source as shown in figure(1), called inter-channel time difference (ITD). Inter-channel Time Difference (ITD) is a primary horizontal localization cue and is a function of frequency. As the sound intensity in free field is inversely proportional to the square of distance from the source, the sound appears

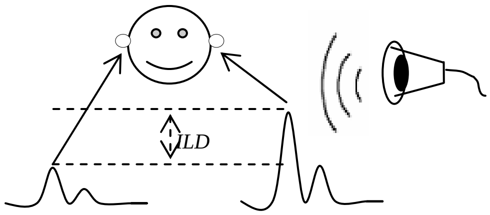

Fig.2. Inter-Channel Level Difference

to be louder to the ear which is close to it. This phenomenon is called ILD as shown in figure (2). Similar to ITD and ILD, Inter-channel coherence (ICC) also plays an important role in the creation of more spaciousness in the sound. ICC is the cross-correlation function between the impulse responses at the two ears. Low ICC results into more spacious sound and a greater feeling of immersed 'fuller' sound image. Along with these binaural cues head shadow, torso reflections also play a major role for realistic listening scenario [6].

IV. Challenges in Audio Externalization

The major challenges for the successful implementation of audio externalization involve:

a. Decorrelation of the Left and Right Channel
b. Increasing the Side signal (S) to Center signal (C) ratio and determining the widening measure
c. The spectral coloration
d. Cross talk for spatial envelopment

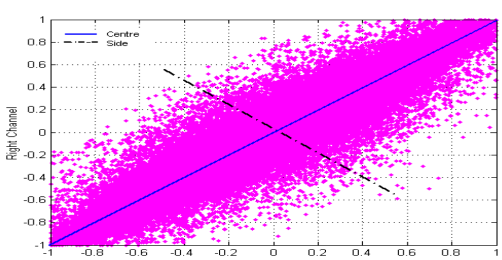

Fig.3. Scatter plot showing the centre signal (L+R) and side signal (L-R) components

As the above problems are not related to each other directly, different solutions exist. In general, correlation coefficient can be minimized by using static decorrelation technique in which stereo channels are passed to two different all pass filters either with random phase characteristics or with complementary phase characteristics. Dynamic decorrelation can be obtained by calculating all pass filter with random phase response for every new time frame [7]. Preferably IIR filters are used in dynamic decorrelation as the filter coefficients can be easily updated by randomly varying the distances of the poles and zeros from the unit circle [8]. In the conventional stereo, centre signal appears to come from middle of the listeners head. The side signal is a sound that comes from the surrounding rather than directly from the sound source. The centre signal components are 0.707*(L+R) and side signal components are 0.707*(L-R). The S/C ratio can be increased by adding the attenuated delayed side signal to Left and Right channels. These two signals C(shown in solid line) and S(shown in dashed dot line) are perpendicular to each other and form a new coordinate system, as shown in figure (3). The ratio of S/C is a measure to determine the extent of widening of signal. A highly widened signal possess high S/C ratio and vice versa. Almost all of the stereo widening algorithms create a little or some spectral coloration to the signal because of the feed-back/feed-forward delay networks [9]. Since the early reflections are the major cause for the coloration effect on the audio signal, our model uses only one dominant early reflection to minimize the spectral coloration. In the loudspeaker setup, the sounds which are intended to be heard on only one channel leak onto the other channel and is called as crosstalk. This phenomenon is frequency dependent in which lower frequencies gets diffracted and high frequencies get attenuated by almost nearly 20 dB. The crosstalk that is present in normal hearing is mimicked by mixing a time-delayed and low pass filtered signal of one channel to the other channel.

V. Implementation

Our method involves time domain approach for audio externalization, as shown in figure (4). The detailed description of our algorithm is explained in this section.

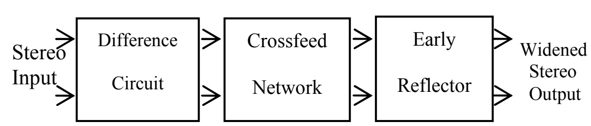

Fig.4. Stereo Widening system

A. *Difference Circuit*

The block diagram of the difference circuit is shown in figure (5). Let L_{in} and R_{in} be the stereo audio channels in time domain. The input signals are passed through the optional decorrelator, consisting of two all-pass filters of linear phase characteristics that introduce complementary phase shifts in the left and right audio channel [10]. The decorrelator is highly useful if the stereo channels are highly correlated. Decorrelation reduces the timbral coloration when applied to individual channels [8]. By this method the cross correlation coefficient of left and right channels is highly reduced as shown in figure (6). The widening is achieved by enhancing the directional information in each channel. The side signal is delayed in the range of 10 ms to 25 ms. Empirically, the delay of 15 ms was selected as optimum as the delay exceeding 25 ms may incorporate echo artefacts and more timbral coloration.

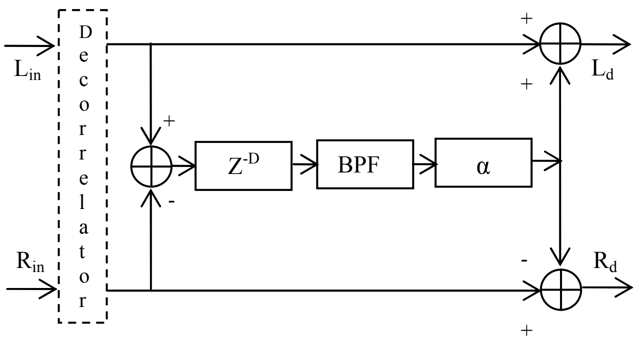

Fig.5. Block diagram of difference circuit

The delayed difference signal is passed through the band pass filter (BPF) for frequency selective spatial enhancement. Default is the band pass filter with cut-off frequencies 250Hz and 12 KHz for lower bound and upper bound respectively. The output of the BPF is attenuated by a widening depth parameter 'α' and added to the left channel and the 180 degrees phase shifted signal is added to the right channel. This circuit increases the side to centre signal ratio, thus enhancing the spatial fidelity of the signal.

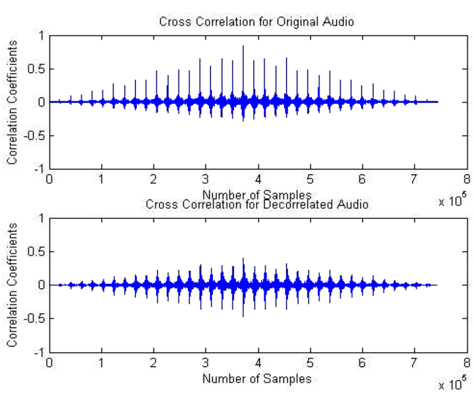

Fig.6. Cross correlation plot of original stereo channels (top) and the decorrelated stereo channels (bottom)

B. Cross-feed circuit

It is a well known phenomenon that crosstalk occurs when the left ear hears a little of the right channel signal and vice versa in a loudspeaker set-up. To create similar listening effect on headphones a crossfeed network can be used to recreate inter-aural crosstalk [11]. The low frequency sounds gets diffracted as they go behind the head. The high frequency sounds above 1.5 KHz are blocked as the sound wavelength is short compared to the head dimension, also termed as head shadow effect. ITD and ILD cues are prominent cues below 1.5 kHz and head shadowing becomes prominent above 1.5 kHz [12]. The cross-feed circuit is shown in figure (7) which simulates the inter-channel cross talk. The delay (Δd) is chosen as 15 samples (We assume that it takes 0.3 ms for sound to reach from loudspeaker to listener at 44.1 KHz sampling frequency). The delay introduced in cross talk path is Δx = (Δd+itd) where itd is the inter-aural time difference which is approximately 9 samples (0.2 ms). The direct path signal is passed through an all pass filter (Hd). Cross path signal is passed through a cross talk filter (Hx) which is a low pass filter of cut-off frequency 2 KHz. Delayed cross talk filtered output from the left channel is multiplied by cross-feed gain parameter 'β' and added to right channel and vice versa. The cross-feed gain parameter controls ILD between the cross path and direct path. The gain parameter is chosen in the range of 0.3 to 0.7.

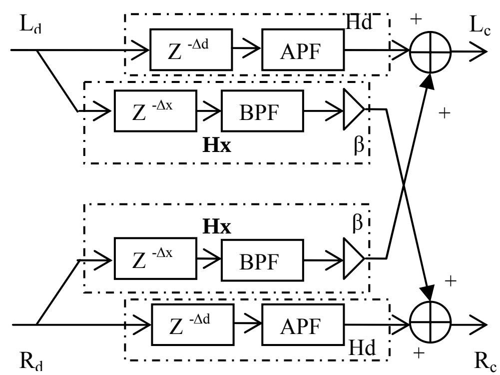

Fig.7. Block diagram of Crossfeed circuit

C. Early Reflection Circuit

Out of head localization can be achieved by adding the appropriate amount of indirect sound to direct sound. The system shown in figure (8) simulates the addition of a single, dominant early-reflection by using a feed-forward delay network.

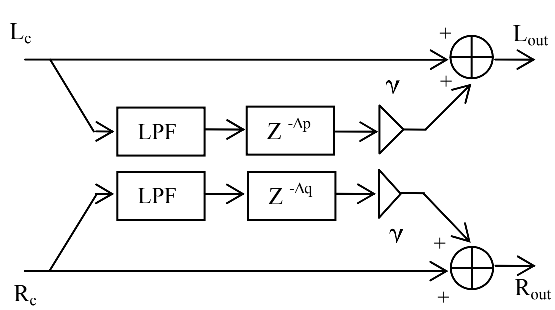

Fig.8. Early Reflection Circuit

The early reflections provide most of the spatial information, recognizable directionalities as well as distinct arrival times of an environment [13], [14]. A low pass filter is inserted in the delay loop as suggested by Schroeder [15] and Moorer [16] as the walls tend to absorb more high frequencies than low frequencies in a room environment. In our implementation, each channel signal is added to a delayed attenuated version of the channel itself, to reduce the inter-channel coherence while preserving the quality of the signal. The two slightly different delays are used in both the channels to minimize the Inter channel coherence. Empirically we have chosen the gain parameter, 'γ' as 0.5 and the 'Δp' and 'Δq' delays can lie in the range of 5 ms to 10 ms and the difference between two delays was selected as 3 ms for good results. Thus, the lower ICC improved the spatial legibility and immersiveness in the stereo audio channels.

VI. RESULTS

To evaluate the model more systematically, the subjective listening tests and objective tests were performed. In subjective listening tests, the task of the listener was to perceive the extent of pleasing widening effect and sensation of the externalization with reference to the input audio. A variety of music genre (Pop, Rock, Hip-hop etc.,) of CD quality dataset was collected and subjectively evaluated by 25 listeners independently. They are asked to rate the quality by giving a 1 - 5 score. The rating is as follows (with respect to the input audio):

5. Widening and sensation of externalization
4. Widening
3. Better than Original
2. No difference
1. Poor quality

The average rating for each dataset corresponding to various music genres is shown in figure (9). From the results, as shown in figure (9), we notice that for most of the music genres, our system gives good widened output. The listeners felt an appreciable amount externalization for genre's containing more high frequency signal components.

In objective tests, statistically we have measured the extent of widening by plotting the histogram of the difference signals as shown in figure (10). Most part of the difference signal of the conventional stereo is concentrated at the centre while the difference signal of the widened audio is more spread towards left and right ears. The centre signal is almost reduced one third when compared to the original stereo.

VII. CONCLUSION

We conclude that ILD, ITD and ICC are the dominant cues for externalization giving the sensation of the widened sound stage on headphones which further helps to lessen fatigue. Our experimentation shows that the appreciable extent of externalization is achieved without affecting the timbral quality of original signal. Since our approach was limited to time domain, it opens up many possibilities for future extensions using frequency domain analysis. Flexible modifications are possible within our approach, as HRTF processing and use of additional cues like pinna cues etc., along with late reverberation could result in better realistic listening.

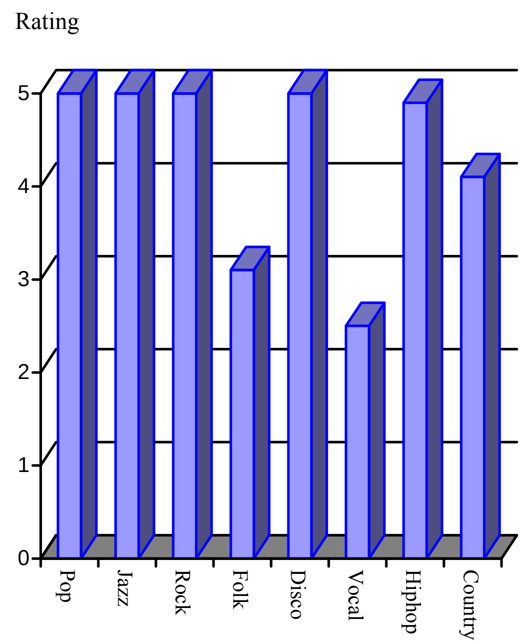

Fig.9. Results of subjective listening tests

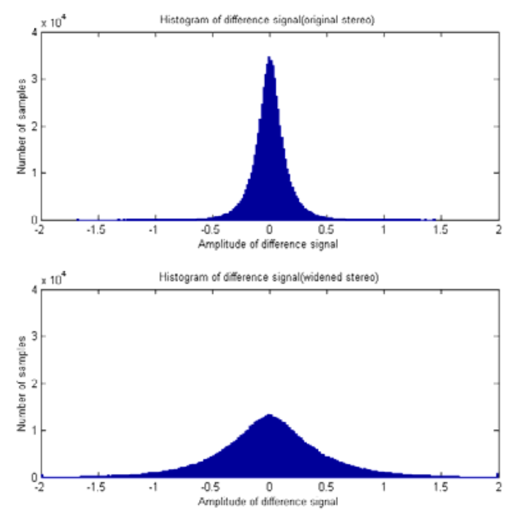

Fig.10. Histogram of difference signal - Conventional stereo (top) and widened stereo (bottom)

ACKNOWLEDGMENT

The authors thank Raj Narayana Gadde for his suggestions and encouragement in all the time of research work. This research was supported by Advanced Technologies Division of the Samsung India Software Operations, Bangalore, India.

REFERENCES

[1] http://www.musictechnology.northwestern.edu

[2] B.B.Bauer, Improving headphone listening comfort, Journal of the Audio Engineering Society, Vol.13, No.4, pp 300-302, April 1965.

[3] David Griesinger,"General overview of the spatial impression, envelopm-ent, localization and externalization", Lexicon.

[4] Weinrich, S Gert "Improved Externalization and Frontal Perception of Headphone Signals"AES-1992.

[5] Eargle, J. M., Handbook of Recording Engineering, New York: Van Nostrand Reinhold, 1986.

[6] Yost, William A, "Perceptual models for auditory localization", presented at the AES 12th conference.

[7] Guillaume Potard, Ian Burnett, "Decorrelation techniques for the rendering of the apparent sound source width in 3D audio displays", 7th international conference on Digital Audio Effects (DAFX'04), Naples, Italy.

[8] Gary S Kendall, The Decorrelation of audio signals and its impact on spatial imagery, Computer Music Journal, 19:4, pp:71:87, Winter1995, Massachusetts Institute of Technology.

[9] Jean Mark Jot, Carlos Avendano, "Spatial enhancement of Audio Recordings", AES 23rd Int. Conference, Copenhagen, Denmark, May 23-25, 2003.

[10] Stefan A Davis, Martin Walsh, and David Berners, US patent: 6714652. http://www.freepatentsonline.com/6714652.html

[11] Ole kirkeby, A balanced stereo widening network for headphones, 22nd International conference on Virtual, Synthetic and Entertainment Audio.

[12] J. Blauert, "Spatial Hearing. The psychophysics of human sound localization", revised edition, MIT Press, Cambridge MA, 1997.

[13] Soren H. Nielsen. "Auditory distance perception in different rooms", J. Audio Eng. Soc., 41(10):755–770, October 1993.

[14] Schroeder, M. R., and Atal, B. S., "Computer Simulation of Sound Transmission in Rooms," 1963 IEEE International Convocation Record, Part 7, pp.150-155, March 25-28, 1963.

[15] M.R.Schroeder, "Natural Sounding Artificial Reverberation", *J. Audio Eng. Soc.*,vol. 10, no. 3, pp. 219-223, July 1962.

[16] J. A. Moorer, "About This Reverberation Business", Computer Music Journal, vol. 3, no. 2, pp. 13-28, 1979.

S.M.A. Basha received the MScEE, degree in signal processing from Blekinge Institute of Technology, Sweden in 2005.From 2005 to 2006 he was employed at the Satyam Computer Services Limited, Bangalore, India, working on audio and speech codec's. He is presently working for Advanced Technologies Division, Samsung India Software Operations Private Ltd, Bangalore, India, in research and development of audio and speech processing applications. His main research interests are audio processing, speech recognition, speech enhancement, genetic algorithms and multi-rate filter-bank analysis.

Abhinav Gupta received the BS degree in electronics & communication engineering from the Indian Institute of Technology, Chennai, India in 2005. From 2005-2006, he worked for Advanced Technologies Division, Samsung India Software Operations Private Ltd, Bangalore, India, in research and development of audio and speech codec's. He is presently working for Samsung Electronics Corporation Ltd, Suwon, South Korea, in research and development of DSP applications. His research interests are audio & speech processing, and optimization

Anshul Sharma received the MS degree in Signal Processing and Communication from The Ohio State University, Columbus, USA, in 2000, and a the BS degree in Electrical Engineering from Indian Institute of Technology, Kanpur, India in 1998. He has been working in the domain of multimedia compression technologies and related DSP architectures. He worked as a DSP Software Engineer at Improv System Inc., Beverly, MA, USA from 2000 to 2003 working on reconfigurable architectures for multimedia processing. Since 2003, he has been working in the Advanced Technologies Division in Samsung India in the areas of multimedia compression and transmission.