Low Power DSP Implementation of 3D Sound Localization

Noriaki Sakamoto†, Wataru Kobayashi††, Takao Onoye†††, Isao Shirakawa†
†Dept. Information Systems Engineering, Osaka University
2-1 Yamada-Oka, Suita, Osaka, 565-0871 Japan
Phone: +81-6-6879-7807, Fax: +81-6-6875-5902, E-mail: {sakamoto, sirakawa}@ise.eng.osaka-u.ac.jp

††Arnis Sound Technologies, Co. Ltd.
2-7-9 Kita-Senzoku, Ota, Tokyo, 145-0062 Japan
Phone: +81-3-5749-3621, Fax: +81-3-5749-3625, E-mail: kobachan@arns.com

†††Dept. Communications & Computer Engineering, Kyoto University
Yoshida-Honmachi, Sakyo, Kyoto, 606-8501 Japan
Phone: +81-75-753-4803, Fax: +81-75-753-4804, E-mail: onoye@kuee.kyoto-u.ac.jp

Abstract: This paper describes a DSP implementation of a real-time 3D sound localization algorithm with the use of a low power embedded DSP. A distinctive feature of this implementation is that the audible frequency band is divided into three, in accordance with the sound reflection and diffraction phenomena through different media from a certain sound source to human ears, and then in each subband a specific implementation procedure of the 3D sound localization is devised so as to operate real-time at a low frequency of 50MHz on a 16bit fixed-point DSP. Thus our DSP implementation can provide a listener with 3D sound effects through a headphone at low cost and low power consumption.

1. Introduction

In recent years, according to the progress of image processing and virtual reality technologies, a great number of studies have been carried out on three-dimensional (3D) visualization [1,2]. On the other hand, in the field of acoustic signal processing, several sophisticated approaches have been attempted for realizing 3D sound effects [3, 4], which are based mainly on the so-called Head-Related Transfer Function (HRTF) [5].

The frequency response characteristic of an HRTF can be measured by recording sounds through dummy head microphones at left and right ears. In this case, however, the HRTF measured by such binaural recording is so complicated in terms of peaks and dips that, in order to realize the accurate characteristic, a considerable number of digital filters with much flexibility in setting the parameters of frequency, gain, quality factor (Q), etc., are necessary.

To cope with this difficulty, a novel 3D sound localization algorithm has been developed [6], which can attain low computational complexity by extracting key factors necessary for perceiving the 3D localization through the human auditory sense.

This paper describes a DSP implementation on the basis of this real-time 3D sound localization algorithm with the use of a low power embedded DSP, for which the computer complexity and power consumption are to be evaluated.

2. Conventional 3D Localization

A conventional method for the 3D sound localization is summarized as follows [3, 4]: First, compute the HRTFs necessary for the 3D sound localization on the outside of the human head in the auditory senses, which can be attained with the use of two equations, one representing the signal output from the dummy head microphone for the sound transferred from a certain sound source and the other indicating the signal output from the stereo-headphone. Then, a monaural input signal is processed real-time by these HRTFs, and the results are superposed to be reproduced. In other words, the 3D sound localization can be realized by the stereo sound which is reproduced by processing through HRTFs a sound of no localization.

Usually, the frequency response characteristic of an HRTF is so complicated in terms of peaks and dips that a great number of digital filters are necessary to realize the characteristic accurately, and moreover, in the case when a parametric equalizer (PEQ) is employed for such a digital filter, all the parameters of frequency, gain, and quality factor (Q) must be taken into account for each filter. Furthermore, the frequency response characteristics at both ears for any localized sound source are different each other, and hence the structures of filters to realize these characteristics at both ears should be different. Therefore, a great amount of operations are necessary for the whole set of filters to realize the 3D sound localization.

Consequently, associated with such a conventional 3D sound localization algorithm, there still remains much room to remove the barrier of tremendous computational complexity so as to realize the real-time 3D sound localization.

3. Real-time Algorithm

A new real-time algorithm [6] of the 3D sound localization intends to extract the major factors necessary for lowering computational complexity.

First, this algorithm divides the audible frequency band into three so that an effective scheme of modeling the diffraction effects can be devised in each of these frequency subbands.

The architecture of this algorithm is outlined in what follows: In the case of achieving 3D sound localization for a monaural sound source, audio digital signals input according to time series are processed in this order. First, an audio digital input signal is divided into two: left and right channels. Second, each audio digital signal is divided into three frequency subbands, and the direction of a sound source and the distance to the source are computed to be added to the audio signal in each frequency subband. Finally, audio digital signals of three subbands are mixed and output.

The diffraction characteristic in each frequency subband is outlined below.

A: Low Frequency Subband

First, consider the sound diffraction by a human head. Suppose a human head to be of spherical shape with diameter of about 150~200mm, although there are small differences among individuals. If the half of a sound wavelength is larger than the head diameter, the effect of sound reflection and diffraction by the head are negligibly small. In this case, given an audio digital input signal, the 3D localization can be achieved only with the use of the difference between sound volumes and that between arrival times of the sound emanating from the sound source and entering into right and left ears.

Assuming that the sound velocity and the head diameter are given by $v = 340m/s$ and $d1 = 150{\sim}200mm$, respectively, the boundary frequency is given by $f1 = v/(d1/2) = 850{\sim}1{,}100Hz$.

B: High Frequency Subband

Next, consider the sound diffraction by a human pinna. Suppose a human pinna is a cone with base diameter of 35~55mm, and if the half of sound wavelength is shorter than the diameter of base, the pinna physically has a great influence on the sound diffraction. It is verified through the measurement by dummy head microphones that the frequency response characteristic of a sound in this high frequency subband can be approximated by a comb filter.

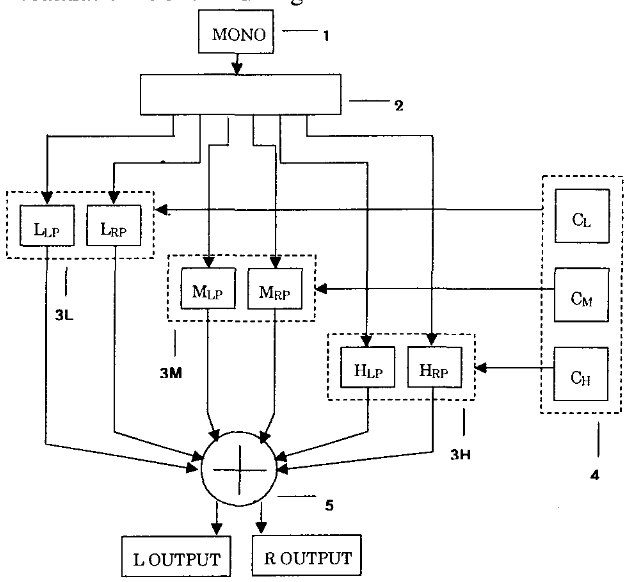
Suppose that the base diameter of a human pinna is given by $d2 = 35{\sim}55mm$, the boundary frequency is obtained as $f2 = 3{\sim}5kHz$.

C: Intermediate Frequency Subband

The intermediate frequency subband between f1 and f2, the 3D sound localization is performed by a parametric equalizer (PEQ) in the same way as in the conventional method. Thus the frequency response characteristic of an HRTF has to take account of the influence of the sound diffraction by the head and the pinna.

In this frequency subbands, PEQs are used only in the intermediate subband and a great number of operations can be cut in comparison with the usual operation, which uses PEQs in the low and high frequency subbands.

An outline of this real-time algorithm of 3D sound localization is shown in Fig.1.



1 MONO: a monaural source
2 Filters dividing frequency subbands
3L LLP: a low frequency subband processing for a left ear
LRP: a low frequency subband processing for a right ear
3M MLP: an intermediate frequency subband processing for a left ear
MRP: an intermediate frequency subband processing for a right ear
3H HLP: a high frequency subband processing for a left ear
HRP: a high frequency subband processing for a right ear
4 CL: parameters for a low frequency subband processing
CM: parameters for an intermediate frequency subband processing
CH: parameters for a high frequency subband processing
5 mix and delay
L OUTPUT: processed sound for a left ear
R OUTPUT: processed sound for a right ear

Fig.1. The Outline of the real-time algorithm of sound localization.

4. DSP Implementation of Real-time Algorithm

A low-power DSP implementation of this real-time algorithm for the 3D sound localization is outlined in Fig. 1, where a sound source is assumed to be static.

Our implementation flow of the real-time algorithm of 3D sound localization by a Texas Instruments 16-bit fixed-point DSP, TMS320C54x, is illustrated in Fig. 2.

First, the implementation procedure of this real-time algorithm is divided into three phases, (i) frequency division, (ii) sound localization, and (iii) mixing. Each phase is summarized in what follows.

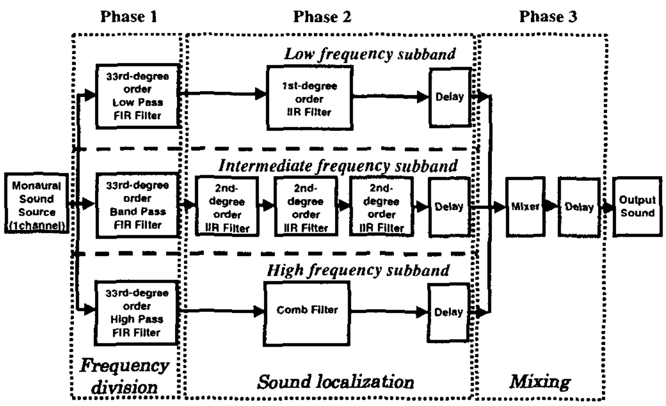


Fig. 2. Implementation flow of 3D sound localization.

[PHASE1: Frequency Division] First, the real-time algorithm of the 3D sound localization executes the frequency division for monaural audio signals. The low, intermediate, and high frequency subband are obtained by using the 33rd-order low pass FIR filter, band pass FIR filter, and high pass FIR filter, respectively. Since the parameters used in FIR filters are dependent on the location of a sound source, DSP stores the parameters calculated in advance in an external memory. The audio data attained in each subband are stored in an internal memory.

[PHASE 2: Sound Localization] The 3D localization is invoked for audio data in the three frequency subbands.

In the low frequency subband, we have only to consider the parameters of delay and volume differences between the left and right channels. Thus we have

$$y[n] = (x[n] - A*d[n-1])*C + B*C*d[n-1] \qquad (1)$$

which shows a first-order IIR filter, called a Shelving filter, as illustrated in Fig. 3, where the parameters of A, B*C, and C indicate gains. It should be noticed that the 3D sound localization can be realized with the use of only one IIR filter.

In the intermediate frequency subband, in order to reproduce the frequency response characteristic of an HRTF with high precision, three second-order IIR filters of Fig, 4 are used, each of which is represented by the following equation

$$y[n] = (x[n] + b1*d[n-1] + b2*d[n-2])*a0 + a1*d[n-1] + a2*d[n-2], \qquad (2)$$

where it should be added that the series connection of three IIR filters needs the memory area less than the parallel connection, and that totally 15 parameters of a0~a2 and b0~ b1 are necessary.

In the high frequency subband, a comb filter of Fig. 5 is used to realize the 3D sound localization, where totally 4 parameters of direct gain, effect gain, feedback gain and delay, are necessary.

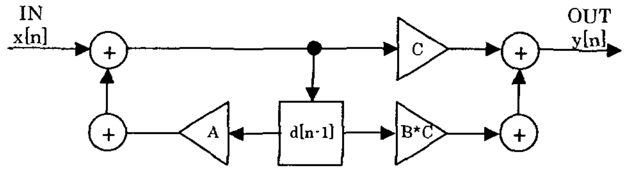


Fig. 3. Structure of the Shelving IIR filter.

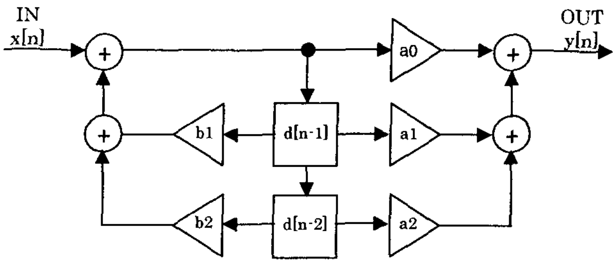


Fig. 4. Structure of the second-order IIR filter.

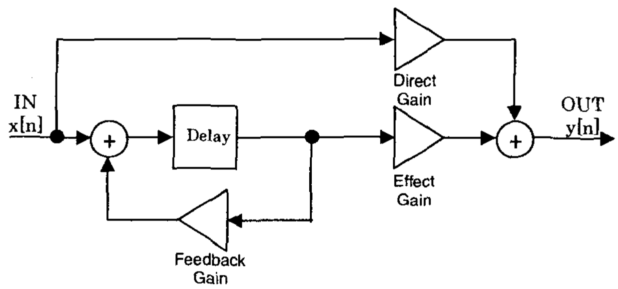


Fig. 5. Structure of the comb filter.

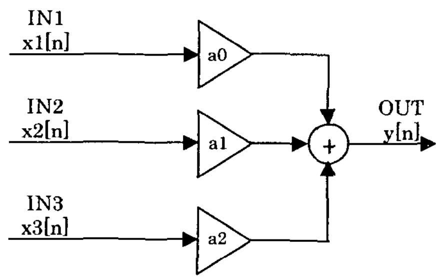


Fig. 6. Structure of mixer.

After the filter processing in each frequency subband, the volume control and the delay control are performed for the left and right channels.

[PHASE 3: Mixing] After the 3D sound localization, audio data in three subbands are mixed by adjusting a ratio, and the total delay control is applied again to the mixed sound.

In this mixing phase, as shown in Fig. 6, a specific frequency subband can be emphasized by multiplying audio data by a gain.

In each PHASE, the audio input data are treated in the form of 16-bit integer, the intermediate data are calculated in the 32-bit precision including the fraction part, and the output data after processing is rounded off to the form of 16-bit integer.

The computational complexity of each PHASE is summarized in Table 1, where: 1) TMS320C54x is used for a monaural-sound input with one channel and stereo-sound output with right and left channels at a sampling frequency of 48kHz; 2) the filter parameters and the volume and delay difference are loaded by DSP from an external memory area to an internal memory area in advance; and 3) the overhead of setting initial values is not taken into account. In addition, the process of a band pass FIR filter is executed by the sequence process of a high pass FIR filter and a low pass FIR filter, where it should be added that a series connection of low and high pass FIR filters can keep the flatness in the intermediate subband, which can avoid the management of flatness by PEQ.

As can be seen from Table 1, 15.4M operations per unit time are invoked in the frequency division, 25.3M in the sound localization, 0.96M in the mixing, and totally 42M. This means that the 3D sound localization can be achieved by a 50MHz DSP, even if we take account of the initial overheads necessary in each processing phase such as transferring filter parameters and sound data from/to external of the DSP.

TABLE I
COMPUTATIONAL COSTS IN EACH PHASE

	Frequency division	Sound localization	Mixing	
Low subband	3.65M	3.46M	0.96M	
Intermediate subband	8.16M	18.6M		
High subband	3.65M	3.17M		
Total	15.4M	25.2M	0.96M	41.7M

The power consumption of the DSP implementation of this real-time 3D sound localization algorithm is 135mW, which includes 45mA current of DSP core in FIR processing and 3V supply voltage from the datasheet of TMS320C54x [7].

The implementation of this real-time 3D sound localization run on a 16bit fixed-point DSP is fully capable of operating at a low frequency of 50MHz. Consequently, our implementation can effectively provide a listener with the 3D sound effects through a headphone at low cost and low power consumption. This 3D sound localization scheme can be applied to a headphone, a handy-phone, and so on, with an embedded DSP.

5. Conclusion

This paper has described a DSP implementation of a real-time 3D sound localization algorithm with the use of a low power embedded DSP. Specifically, by using a Texas Instruments 16-bit fixed-point DSP, TMS320C54x, the real-time 3D sound localization can be realized at a low frequency of 50MHz. In addition, the experimental result of our DSP implementation indicates that the power consumption is 135mW at 3V supply voltage. Consequently, our new approach can effectively provide a listener with the 3D sound localization through a headphone at low cost and low power consumption.

References

[1] T. Naemura, M. Kaneko, and H. Harashima: "Orthographic approach to representing 3-D images and interpolating light rays for 3-D image communication and virtual environment," *Signal Processing: Image Communication*, vol. 14, pp. 21-37, Nov. 1998.

[2] V. Blanz and T. Vetter: "A morphable model for the synthesis of 3D faces," in *Proc. Computer Graphics (SIGGRAPH' 99)*, pp. 187-194, 1999.

[3] M. Okamoto, I. Kinoshita, S. Aoki, and H. Matsui: "Sound image rendering system for headphones," *IEEE Trans. Consumer Electronics*, vol. 43, pp. 689-693, Aug. 1997.

[4] C. Phillip Brown and Richard O. Duda: "A Structural model for binaural sound," *IEEE Trans. Speech Audio Processing*, vol. 6, pp. 476-488, Sept. 1998.

[5] E. M. Wenzel, M. Arruda, D. J. Kistler, and F. L. Wightman: "Localization using nonindividualized head-related transfer functions," *J. Acoust. Soc. Amer.*, vol. 94, pp. 111-123, July 1993.

[6] W. Kobayashi, N. Sakamoto, T. Onoye, and I. Shirakawa: "3D acoustic image localization algorithm for embedded DSP processing," submitted to *ITC-CSCC2000*.

[7] Texas Instruments, "Calculation of TMS320LC54x Power Dissipation, Application Report," June 1997.