Enhance spatial sound of stereo audio by upmixing

**吳柏澔**

**Ⅰ. Introduction:**

Nowadays, virtual reality develops quickly, so does audio. There are a host of methods to make listeners feel the sound is like surrounding. For instance, movie theaters apply multi-channel audios, such as 5.1 or 7.1 surrounding sound and virtual reality takes advantage of ambisonic to present audio in 360 degrees. However, in people’s daily lives, they invariably listen to stereo audio to enjoy the music. Thus, strengthening the spatial sound of stereo audio is quite fascinating. People can enjoy more real stereo sound in simpler ways. The approach can be completed by upmixing stereo sound to 5.1 surrounding sound with simple method, low pass filter and so forth. Another approach is completed by a much more complicated algorithm, Primary-Ambient Extraction (PAE). After that, by testing different weights of each channel, stereo sound can be generating by combining different weighted channels in order to enhance spatial feeling.

**Ⅱ. Method:**

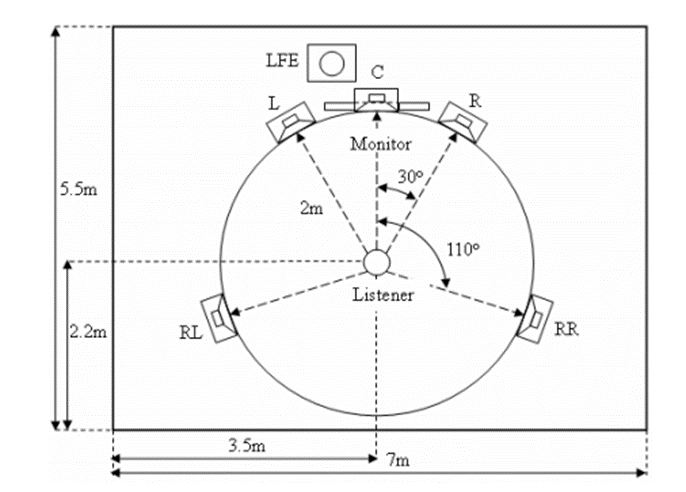
First of all, we need to decide relative positions for 5.1 surrounding channels. The general locations are shown in Figure.1. Then, the architecture for each channel is shown in Figure.2.

For the Center channel, it is generated by filtering the input using an FIR bandpass filter with cut-off frequencies 100 Hz and 4 kHz. In this channel, the main purpose is to emphasize human voice.

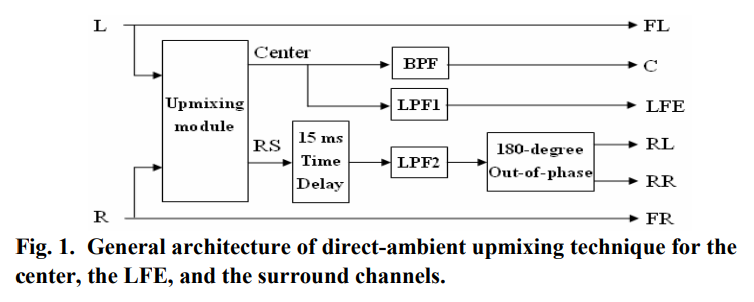
For the Rear Left (RL) and the Rear Right(RR) channels, both of them are created by adding a 15ms delay, and then filtering with a 7 kHz FIR lowpass filter. Both channels are intended to generate ambient sound, that is background sound. In addition, the rear channels are 180-degree out-of-phase with one another in order to increase the spaciousness of the field.

For the Low Frequency Enhancement (LFE) channel is derived by filtering the input with a lowpass filter to keep the signals below 120Hz. The goal of this channel is to enhance the bass sound.

In addition, there is another way, PAE, to acquire Center, RL, and RR channels. From [2], it can extract the direct sound and ambient sound by using Principal Component Analysis. The equations of primary and ambient are:



**Figure.1. The standard 5.1 configuration for multichannel loudspeaker reproduction.**



**Figure.2. General architecture of direct-ambient upmixing technique for the center, the LFE, and the surround channels**

where 0 means left channels, 1 means right channel, p0 means dominant source of original input of left channel, p1 means dominant source of original input of the right channel. are the extracted primary and ambient components in the two channels respectively.

Ⅲ. Result:

Since most of the results are audio files, there is a video to demo the result. Please see the appendix file. Before watching the video, each audio is divided into two parts. First part is music with instruments. The second part is music with the singer’s voice.

Ⅳ. Conclusion:

From the result, the ambient component of PCA can filter more human voice than audio in RL and RR channels. However, since the final goal is to apply to electronic devices, the PCA method may be too complicated to achieve. Thus, we decide to utilize the first approach. It is more possible to implement. Besides, for the first approach in the demo video, it still needs a great deal of testing to choose optimum weights. What’s more, those optimum weights may be different with different types of music. As a result, for future work, we need to classify different types of music, so that we can set proper standards to adjust weights. Moreover, we have lots of works needed to do for the cutoff frequencies of filters as well.

Ⅴ. Reference:

[1] Bai, Mingsian & Shih, Geng-Yu. (2007). Upmixing and Downmixing Two-channel Stereo Audio for Consumer Electronics. Consumer Electronics, IEEE Transactions on. 53. 1011 - 1019. 10.1109/TCE.2007.4341580.

[2] He, Jianjun & Tan, Ee-Leng & Gan, Woon-Seng. (2014). Linear Estimation Based Primary-Ambient Extraction for Stereo Audio Signals. IEEE Transactions on Audio Speech and Language Processing. 22. 505-517. 10.1109/TASLP.2013.2297015.