CENTER CHANNEL SEPARATION BASED ON SPATIAL ANALYSIS

Dae-young Jang
Broadcasting Telecommunications Media Research Department
ETRI, Daejeon, Korea
dyjang@etri.re.kr

Jeong-pyo Hong
Dept. of Electronics
ICU, Daejeon, Korea

Hoe-Kyung Jung
Dept. of Computer Engineering
Paichai Uni., Daejeon, Korea
hkjung@pcu.ac.kr

Kyeongok Kang
Broadcasting Telecommunications Media Research Department
ETRI, Daejeon, Korea
kokang@etri.re.kr

ABSTRACT

This is a brief description of audio channel or sound source separation algorithm using spatial cues. Basically inter-channel level difference (ICLD) is used for discriminating sound sources in a spatial grid for each channel pair and analysis subband. Inter-channel cross-correlation (ICC) is also used for determining sound source location area and contribution factor for the considering composite sound source. In this paper, the center and side channel separation of stereophonic music signal using spatial sound source discrimination method is introduced. This is simply implemented by using given information of center channel location and derived spatial cues. The separated center channel signal is well matched with separated side channels when reproduced simultaneously.

1. INTRODUCTION

Audio channel or sound source signal separation is an open problem up to now. Furthermore the demand of signal separation is quite strong in speech and music applications. It is because of the sound recording environment that cannot isolate sound sources from each other. Only anechoic and sound proof room can isolate sound source from other back-ground sounds, however this environment is not common in various sound recording fields. In case of music recording in studio, original sound track signal is hardly maintained in storage after mixing and mastering, so it's very difficult to get a sound source of old fashioned music.

As a typical sound source separation method, we could consider BSS (Binaural Source Separation); however it is only for multiple mixtures same as the number of mixed sound sources. There have been also many trials of extraction or separation of sound sources from limited mixtures like stereo contents. We can list up several research examples of underdetermined BSS, over-complete presentation, sparse decomposition [1], time-frequency masking [2]. These research activities cannot obtain proper separation for most high quality audio industry up to now. These technologies are just being used in the area of speech intelligibility enhancement for speech recognition and telecommunications.

On the other hand, trials of channel separation (Upmix) technologies have also been increased. And the reason is the strong demand of channel reduction and extension technology from digital broadcasting, movie and DVD like high quality media service that basically use multi-channel audio format. Generally, matrix decoding like as Dolby Pro Logic is widely used; however this method makes several artefacts of level or sound image change due to the phase change. There are also adaptive filter approaches to reduce level change, while maintaining cross-correlation between channels; however, it also cannot be a proper solution. Above mentioned approaches use time domain processing, and recently frequency domain approach is being increased. J. M. Jot et al. introduced downmix and upmix method for multi-channel audio signal using spatial cues in frequency domain [3-5]. C. Faller also presented frequency domain channel separation technologies using spatial cues and BSS [6, 7].

This paper introduces a new concept of spatial domain sound source or channel separation method especially for stereo down-mixed contents. First we perform analysis of spatial sound image between two channel sources, and then find out the location of sound sources and wideness of sound sources from ICC. Furthermore we propose signal separation method for the sound sources that are located in a given position, and channel separation with known channel location information. The followings describe theoretical background of spatial analysis and channel separation method and several results of derivation.

2. SPATIAL ANALYSIS

Most music contents are composed of 2 or more channel signals, and they make certain acoustic space according to the loudspeaker layouts, when they are reproduced with loudspeakers. There are many sound sources in the acoustic space, and currently those sound sources are not allowed to be separated into each other.

Spatial analysis is a kind of frequency domain approach; however it actually extends spatial domain processing. It is known that the location of sound source of stereophonic signal can be figured out using spatial cues like ICLD and ICTD (Inter-channel Time Difference). In case of mixed sound sources, it is

very difficult to extract location information of interested sound sources. The following shows sound source detection and sound source separation method according to spatial analysis.

2.1. Spatial cues

We can estimate the location of sound source with ICLD and ICTD. As a matter of fact, studio music recording simply renders sound source in a location using CPP (Constant Power Panning). Therefore ICLD is more important for the sound source location estimation of music contents than ICTD. ICLD can be easily calculated from level difference of the two channel signals.

CPP can be explained by figure 1. A sound image can be generated when a certain ratio of power is allocated to each channel. On the contrary, we can estimate the location of sound source from comparing levels of the two channel signals according to equation (1).

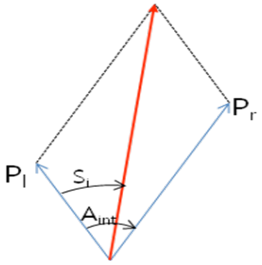

Figure 1: Source direction estimation by inverse CPP

$$S_i\,[\text{deg}] = A_{int} * (P_r / (P_l + P_r)) \tag{1}$$

In case of one source stereo sound, location estimation is satisfactorily accurate as shown in figure 2. However when it contains two or more sources, location estimation becomes difficult at least for overlapped area as shown in figure 3. Also we can see in the figure 3 that sound image blurring is relatively changed according to ICC. That is, when ICC becomes 1 sound image is correctly estimated, and when ICC becomes lower sound image blurring is more serious. Here we can assume sound source location can be estimated as a function of ICC and the location of estimated sound image.

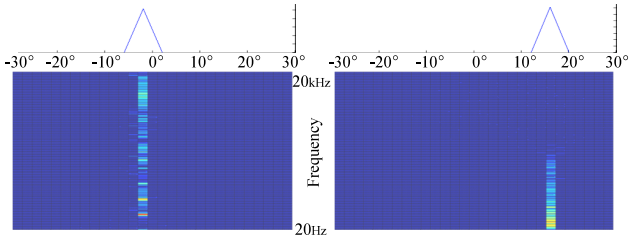

Figure 2: *Sound location estimations for one source*

ICC can be calculated with equations (2) and (3) in time domain as a maximum value of cross-correlation function. ICC also can be obtained in frequency domain with equations (4) and (5) as a degree of cross-correlation value and auto-correlation of each channel signal [8]. In the equations, index *k* represents time and *n* represents frequency band.

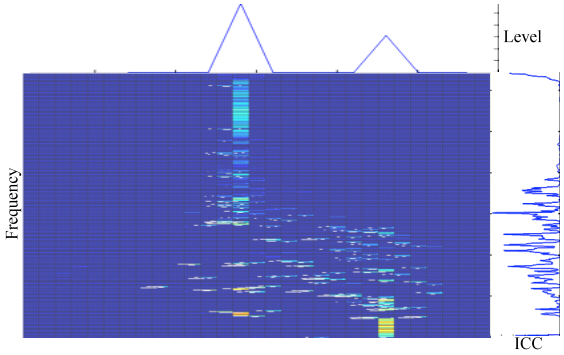

Figure 3: *Sound location estimation for two sources*

$$\bar{\Gamma} = \max_d |\bar{\Phi}_{12}(d)| \tag{2}$$

$$\bar{\Phi}_{12}(d) = \lim_{\ell\to\infty} \frac{\sum_{k=-\ell}^{\ell} x_1(k)x_2(k+d)}{\sqrt{\sum_{k=-\ell}^{\ell} x_1^2(k) \sum_{k=-\ell}^{\ell} x_2^2(k)}} \tag{3}$$

$$|\Gamma(n,k)|^2 = \frac{\Phi_{ij}(n,k)\Phi_{ji}^*(n,k)}{\Phi_{ii}(n,k)\Phi_{jj}(n,k)} \tag{4}$$

$$\Phi_{ij}(n,k) = \alpha \tilde{S}_{i,n}(k)\tilde{S}_{j,n}^*(k) + (1-\alpha)\Phi_{ij}(n,k-1) \tag{5}$$

Theoretically sound source width within loudspeakers is determined according to the ICC values as depicted in the left side of figure 4 [8]. That is, when ICC value is 1, sound image is concentrated at one point, and otherwise sound image is blurred in the area that is generated by loudspeakers. As shown in the right side of figure 4, on the other viewpoint, we can regard two sound sources are located outside of estimated sound image according to the ICC values.

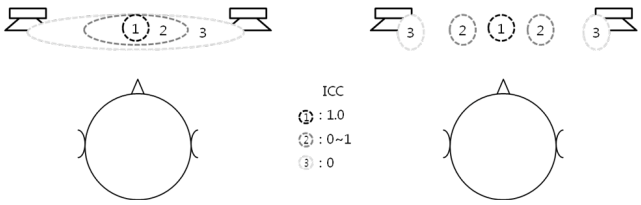

Figure 4: *Sound image for ICC change*

2.2. Spatial domain separation

The previous chapter explained the possibility of sound image estimation with ICLD and ICC. Furthermore in case of multiple sound sources mixture, we can roughly decide sound source location by using ICC, in spite of blurred sound image estimation.

Figure 5 is showing two sound sources mixture image again. It shows classification of principal components for each sound source by masking with spatial filter. Of course, in overlapped frequency bands, some components cannot be avoided to be disregarded. By regarding ICC value, we can roughly determine

sound source levels of given frequency bands for blurred sound image.

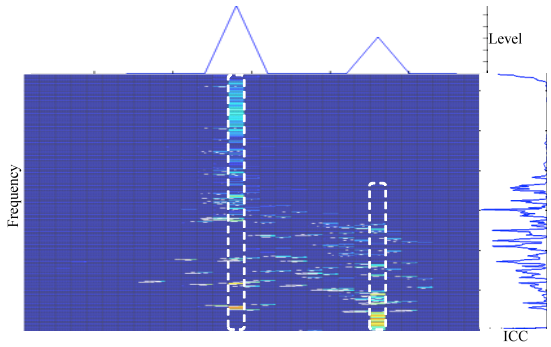

Figure 5: Spatial filter for apparent sound source

Determination of sound source level in a given subband is described in equation (6). Subband level is determined in proportion to ICC value and inverse proportion to distance between expected sound location and estimated sound location. Estimated sound location is limited to search area that is decided in inverse proportion to ICC as described in equation (7).

$$S_n(n) = ICC(n) * f(1/d) * S(n) \quad (6)$$

$$D(n) = \alpha(1 - ICC(n)) + \beta \quad (7)$$

Here $S_n(n)$ is an estimated sound source level for subband n, d is distance between expected and estimated sound location, $D(n)$ is search area for estimating sound source. Figure 6 (white line) shows search area for each subband as a function of ICC.

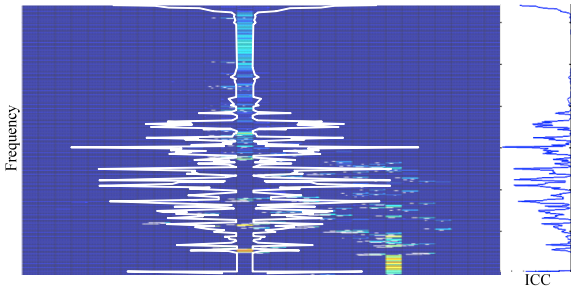

Figure 6: *Spatial filer according to ICC*

3. CHANNEL SEPARATION

3.1. Methodology

As mentioned above, according to ICC variation, we can generate spatial filter that can extract sound sources for a mixture containing several sound sources that are spatially separated. However actual music contents have many sound sources that are not even spatially separated, and therefore it is so difficult to discriminate apparent sound source in a given subband. To achieve more accurate sound source extraction, existing solution like BSS will have to be also considered. Here we would investigate centre and side channel separation method using spatial analysis.

Centre channel signal generally contains vocal signal and other rhythmic instruments. Centre channel extraction method is almost same as sound source extraction. The only difference is sound source location is fixed in centre position, that is in front of a listener. Figure 7 shows spatial filter for centre channel extraction, and it limits minimum and maximum range for search area.

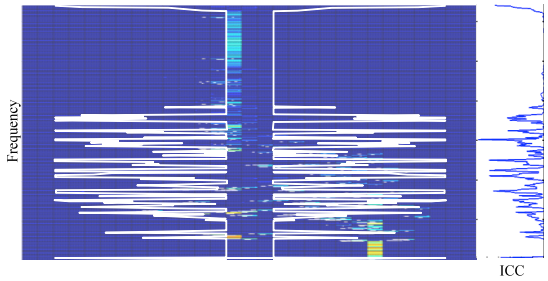

Figure 7: *Spatial filter for center channel image*

$$S_c(n) = ICC(n) * f(1/d) * (S_{lt}(n) + S_{rt}(n)) \quad (8)$$

$$S_{l,r}(n) = (1 - ICC(n) * f(1/d)) * S_{lt,rt}(n) \quad (9)$$

$$S_s(n) = S_l(n) - S_r(n) \quad (10)$$

Equations (8) to (10) describe channel separation for centre (8), left and/or right (9), and rear (10) channel individually. Left and right channel signals are obtained as a remainder of centre channel extraction. Rear channel signal is generated as a difference of extracted left and right channel signals. A proper window of spatial filter for each subband according to the distance from centre position can be designed by various distribution curves. This paper considers simple cosine window.

Centre and side signals essentially make blocking error, due to spectrum discontinuity by spatial domain channel extraction filter. Therefore de-blocking filter has to be adopted for block boundaries of each channel signal.

3.2. Considerations

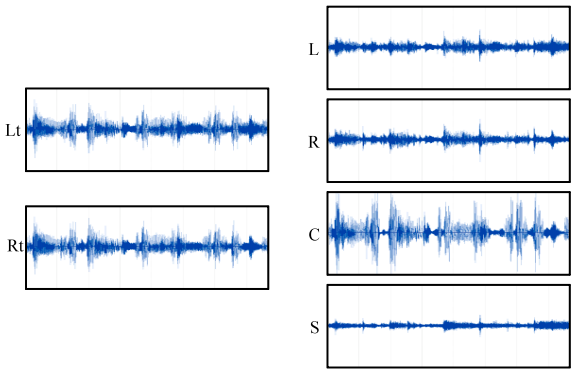

Figure 8: Original and separated 4 channel signals

Figure 8 shows original stereo music signal and separated 4-channel signals. For the most part, centre channel signal get quite portion of energy and left and right channel signals are easily observed that centre channel signal is effectively excluded. Cur-

rently spatial filter and de-blocking filter were not fully optimized, so tiny artefacts were heard. However this spatial domain approach is a new concept of signal analysis and when it is applied with other approach it can get better separation performance.

Extracted multichannel signal has been compared with original stereo music signal in viewpoint of spatial auditory image. The comparison was executed by using headphone generating virtual speakers with binaural sound processing. It made a little blocking noise and sound image change, but produced better sound image than that of matrix based upmixer. Also, centre channel image was more stable and externalized in 3-dimensional space. The reason is depicted in figure 9 representing the difference of two channel and three channel sound presentation. Rear sound also does not cause sound image alteration, but gives enough spaciousness impression.

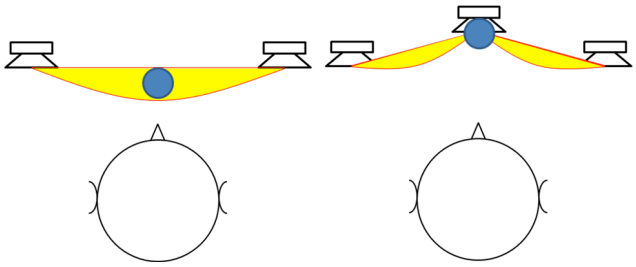

Figure 9: Center Images for 2 and 3-ch virtual loudspeakers

4. CONCLUSIONS

Audio source and channel separation is very important in audio industries for high quality and flexible services. Especially object based interactive audio service and 3-dimensional sound applications strongly demand the audio signal separation technology. Up to now many approaches have been tried out; however there's no appropriate solution yet. Currently several solutions have been integrated for better performance of sound source separation.

Here we introduce a new concept of sound separation based on spatial sound image analysis. As the first usage example, channel separation method is proposed and evaluated. This new approach of channel separation performs better sound image compared with matrix based upmixer. Also, this method is verified that it has an advantage for especially 3-dimensional sound application with more externalization effect.

Spatial analysis is very useful for sound image detection and sound source or channel extraction. Currently the concept of spatial filter was just made, and it is necessary to tune up and optimize to get a proper performance and applicable quality. Furthermore, as mentioned above separated sound source and/or channel can be also applied to enhance 3-dimensional sound effects by handling extracted sound source in acoustical space.

5. ACKNOWLEDGMENTS

This work was supported by the IT R&D program of MKE/IITA. [2008-F-011-01, Development of Next-Generation DTV Core Technology]

6. REFERENCES

[1] Laurent Daudet, "Sparse and Structured Decompositions of Audio Signals in Overcomplete Spaces," DAFX'04 proceedings, Naples Italy, Oct. 2004.

[2] O Yilmaz, S. Richard, "Blind Separation of Speech Mixtures via Time-Frequency Masking," IEEE Transactions on Signal Processing, Vol. 52, No. 7, pp. 1830-1847, Jul. 2004.

[3] C. Avendano, J.M. Jot, "Ambience Extraction and Synthesis from Stereo Signals for Multi-channel Audio Up-mix," ICASSP'02 proceedings, Orlando, FL, USA, Mar. 2002.

[4] J.M. Jot, J. Merimaa, M.M. Goodwin, A. Krishmaswamy, and J. Laroche, "Spatial Audio Scene Coding in a Universal Two-Channel 3-D Stereo Format," 123th AES Convention Paper 7276, New York, NY, USA, Oct. 2007.

[5] J. Usher, J. Benesty, "Enhancement of Spatial Sound Quality: A New Reverberation-Extraction Audio Upmixer," IEEE Transactions on Audio, Speech, and Language Processing, Vol. 15, pp. 2141-2150, Sep. 2007.

[6] C. Faller, "Parametric coding of spatial audio," DAFX'04 proceedings, Naples, Italy, pp.151-156, Oct. 2004.

[7] A. Favrot, M. Erne, C. Faller, "Improved Cocktail-Party Processing," DAFX'06 proceedings, Montreal, Canada, Sep. 2006.

[8] C. Faller, F. Baumgarte, "Binaural Cue Coding-Part II: Schemes and Applications," IEEE Transactions on Speech, and Audio Processing, Vol. 11, No 6, pp. 520-531, Nov. 2003