

A METHOD FOR ENLARGING A LOCATION WITH OPTIMAL THREE-DIMENSIONAL AUDIO PERCEPTION

CROSS REFERENCE TO RELATED APPLICATIONS

5

This application includes references to matter disclosed in US 12/246,491, filed on 6 October 2008.

FIELD OF INVENTION

10

The present invention relates to audio signal processing processes. Specifically, the present invention relates to a method for processing audio signals.

BACKGROUND

15

Stereo signals may be decoded into multi-channel audio to provide a user with a sense of immersion and realism when experiencing the multi-channel audio through a plurality of speakers. The decoding of signals into multi-channel audio may be carried out using techniques disclosed in US 20 12/246,491, which is another patent application filed by Creative Technology Ltd.

It should be noted that a cinema hall typically includes a plurality of speakers distributed in a wide spread loudspeaker layout throughout the cinema hall with the plurality of speakers being directed at cinema goers seated in the cinema hall such that a spatial sound effect is experienced by the cinema goers.

25
30 Unfortunately, arranging a plurality of speakers in a wide spread loudspeaker layout in a relatively smaller enclosed area compared to the cinema hall, such as, for example, a room in a home is not convenient due to constraints in the size of the enclosed area and the fact that the presence of the plurality of

speakers would appear odd. However, it would be highly desirable if spatial sound effects could be reproduced in the home. Furthermore, given the prevalence of compact speaker-array units being found in homes, it would be desirable if spatial sound effects may be reproduced in homes using compact speaker-array units.

In addition, it would also be desirable if the compact speaker-array units could reproduce spatial sound effects over an enlarged location as it is unlikely that persons in a home remain seated at a single location unlike movie-goers in a cinema hall.

The present invention aims to address the aforementioned situations.

SUMMARY

There is provided a method for enlarging a location with optimal three-dimensional audio perception. Optimal three-dimensional audio perception may relate to a fully spatial sound effect.

The method includes deriving three-dimensional encoded localization cues from an audio input signal having a first channel signal and a second channel signal; decoding the first channel signal and the second channel signal into a plurality of decoded channel signals, the plurality of decoded channel signals being equal to a number of speaker units; performing crosstalk cancellation on the plurality of decoded channel signals to eliminate crosstalk between the plurality of decoded channel signals; and outputting the plurality of decoded channel signals which have been subjected to crosstalk cancellation to each of the number of speaker units. It is advantageous that the crosstalk cancellation includes further processing to generate a smoothed frequency envelope.

The smoothed frequency envelope may be reconstructed from truncated cepstrals derived from converting each of the plurality of decoded channel

signals into the cepstrum spectrum. The smoothed frequency envelope also minimizes timbre artifacts, the timbre artifacts being high peaks and low valleys in the cepstrum spectrum of each of the plurality of decoded channel signals.

5

The localization cues may include at least for example, an up-down dimension, a left-right dimension, a front-back dimension, an azimuth angle, an elevation angle and so forth. The derivation of the three-dimensional encoded localization cues may be based on providing a listener with a fully spatial sound effect.

10

The enlarged location with optimal three-dimensional audio perception advantageously allows a listener to move about as the enlarged location relates to a boundary which encompasses a plurality of positions with optimal three-dimensional audio perception.

15

The method may preferably further include summing the plurality of decoded channel signals which have been subjected to crosstalk cancellation before output to each of the number of speaker units. Each speaker unit may include at least one speaker driver. Preferably, the crosstalk cancellation may be performed to cause a listener to perceive audio to be emanated from virtual speakers.

20

DESCRIPTION OF DRAWINGS

25

In order that the present invention may be fully understood and readily put into practical effect, there shall now be described by way of non-limitative example only preferred embodiments of the present invention, the description being with reference to the accompanying illustrative drawings.

30

Figure 1 shows a process flow for a method of the present invention.

Figure 2 shows a schematic view of a system used for carrying out the method of Figure 1.

Figure 3 shows a visual representation of 3D audio reproduction using two loudspeaker arrays.

Figure 4 shows an illustration of a smoothed frequency envelope in a cepstrum spectrum.

- 5 Figure 5 shows a visual representation of 3D audio reproduction using one loudspeaker array.

DESCRIPTION OF PREFERRED EMBODIMENTS

- 10 Referring to Figures 1 and 2, there is provided a process flow for a method 20 for enlarging a location with optimal three-dimensional audio perception (also known by the theoretical concept of "audio sweet spot"), and a schematic view of an apparatus 40 used for carrying out the method 20 respectively. Figures 1 and 2 will be referred to in subsequent paragraphs when describing the
- 15 method 20 and apparatus 40 respectively. It should be appreciated that the method 20 and the apparatus 40 are described herein for illustrative purposes and should not be construed to be limiting in any manner. Optimal three-dimensional audio perception relates to a fully spatial sound effect. It should also be appreciated that the enlarged location with optimal three-dimensional
- 20 audio perception allows a listener to move about as the enlarged location relates to a boundary which encompasses a plurality of positions with optimal three-dimensional audio perception.

- The method 20 for enlarging a location with optimal three-dimensional audio
- 25 perception includes deriving three-dimensional encoded localization cues from an audio input signal having a first channel signal and a second channel signal (22). The audio input signal with the first channel signal and the second channel signal may be known as a stereo signal. The techniques for deriving the three-dimensional encoded localization cues may relate to audio signal
- 30 processing techniques described in US 12/246,491 or any other known audio signal processing technique. The derivation of the three-dimensional encoded localization cues is an essential step to reproduce a fully spatial sound effect. The localization cues includes, for example, an up-down dimension, a left-

right dimension, a front-back dimension, an azimuth angle, an elevation angle and so forth.

The method 20 also includes decoding the first channel signal and the second
5 channel signal into a plurality of decoded channel signals (24), the plurality of decoded channel signals being equal to a number of speaker units. Each speaker unit may include at least one speaker driver. Subsequently, crosstalk cancellation may be performed on the plurality of decoded channel signals (26) to eliminate crosstalk between the plurality of decoded channel signals.

10 Crosstalk cancellation is performed to cause the listener to perceive audio to be emanated from virtual speakers. Crosstalk cancellation eliminates the crosstalk between channels. Crosstalk cancellation also includes further processing to generate a smoothed frequency envelope 100 as shown in Figure 4. The smoothed frequency envelope 100 is reconstructed from
15 truncated cepstrals derived from converting each of the plurality of decoded channel signals into the cepstrum spectrum (labeled as "raw" 102). The smoothed frequency envelope 100 minimizes timbre artifacts, the timbre artifacts being high peaks and low valleys in the "raw" 102 graph in the cepstrum spectrum of each of the plurality of decoded channel signals.

20 Consequently, the method 20 further includes summing the plurality of decoded channel signals (30) which have been subjected to crosstalk cancellation before output to each of the number of speaker units. Finally, the method 20 includes outputting each of the summed decoded channel signals
25 (32) which have been subjected to crosstalk cancellation to each of the number of speaker units such that the listener is able to enjoy the fully spatial sound effect with an enlarged location with optimal three-dimensional audio perception. The concept of the enlarged location will be described in further detail in the subsequent paragraphs.

30 Referring to Figure 5, there is shown a visual representation of 3D audio reproduction using one loudspeaker array with four speakers. It should be noted that the region between E_1 and E_4 represents the enlarged location

(area where lines from the virtual speakers v1, v2, v3, v4 intersect) with optimal three-dimensional audio perception. Head related transfer functions (HRTFs) describe time and amplitude differences that are imposed on a listener's binaural responses to any sound event. These differences are attributed to the listener's head and pinnae structure and are used by ears to detect where sound emanates from. Loudspeaker/headphone virtualization is designed using HRTFs to provide the listener with the perception of sound emanating from virtual rather than actual speakers.

- 10 Mathematical representations will now be provided to illustrate the concept of the enlarged location with optimal three-dimensional audio perception:

X is the multichannel audio produced by deriving three-dimensional encoded localization cues from an audio input signal (22 in method 20).

15

Y is the transaural audio perceived by the listener.

H_c is a HRTF matrix from the real audio sources to the listener.

- 20 H_v is a HRTF matrix from the virtual audio sources to the listener.

$\hat{X}$ is the virtualization output sent to the real audio sources.

ifft relates to "inverse discrete fourier transform".

25

fft relates to "fast fourier transform".

$$Y = H_c X$$

7

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix} = \begin{bmatrix} c_{11} & c_{21} & \cdots & c_{N1} \\ c_{12} & c_{22} & \cdots & c_{N2} \\ \vdots & \vdots & \ddots & \vdots \\ c_{1N} & c_{2N} & \cdots & c_{NN} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix}$$

$$\hat{X} = H_c^{-1} H_v X$$

$$= HX$$

$$= \begin{bmatrix} h_{11} & h_{21} & \cdots & h_{N1} \\ h_{12} & h_{22} & \cdots & h_{N2} \\ \vdots & \vdots & \ddots & \vdots \\ h_{1N} & h_{2N} & \cdots & h_{NN} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix}$$

5

H is converted into cepstrum spectrum,

$$\text{ceps} = \text{ifft}(\log(\text{abs}(H)))$$

10

Subsequently, smoothed spectral envelopes are reconstructed from truncated cepstrals,

$$H_{\text{smooth}} = \exp(\text{fft}(\text{window}(\text{ceps})))$$

15

The smoothed spectral envelopes 100 may be seen in Figure 4.

Referring to Figure 3, there is shown a visual representation of 3D audio reproduction using two loudspeaker arrays. Seven positions of the listener, P1, P2, P3, P4, P5, P6, P7 represent positions where the listener is able to perceive optimal three-dimensional audio perception, where the positions are obtainable from the mathematical processes as detailed in the preceding paragraphs. The seven positions may be deemed to denote a boundary of an area where the listener experiences optimal three-dimensional audio perception.

25

Referring to Figure 2, there is shown a schematic view of a system 40 used for carrying out the method 20. The system 40 allows input of audio input signals in the form of stereo signals (N1 and N2) into a decoder 42 of the system 40. The decoder 42 may process N1 and N2 to derive three dimensional encoded localization cues and decode N1 and N2 into a plurality of decoded channel signals ($x_1, x_2, \dots, x_N$).

The system 40 includes a plurality of audio filters 44 for performing crosstalk cancellation on the plurality of decoded channel signals ($x_1, x_2, \dots, x_N$). Crosstalk cancellation is performed to cause the listener to perceive audio to be emanated from virtual speakers. Crosstalk cancellation eliminates the crosstalk between channels. Crosstalk cancellation also includes further processing to generate a smoothed frequency envelope 100 as shown in Figure 4.

The system 40 includes a plurality of signal summing circuits 46 for summing the plurality of crosstalk cancelled signals. Finally, the plurality of crosstalk cancelled signals which have been summed are output to a plurality of speaker units ($S_1, S_2, \dots, S_N$) such that the listener is able to enjoy the fully spatial sound effect with an enlarged location with optimal three-dimensional audio perception.

Whilst there has been described in the foregoing description preferred embodiments of the present invention, it will be understood by those skilled in the technology concerned that many variations or modifications in details of design or construction may be made without departing from the present invention.

CLAIMS

1. A method for enlarging a location with optimal three-dimensional audio perception, the method including:

5 deriving three-dimensional encoded localization cues from an audio input signal having a first channel signal and a second channel signal;

 decoding the first channel signal and the second channel signal into a plurality of decoded channel signals, the plurality of decoded channel signals being equal to a number of speaker units;

10 performing crosstalk cancellation on the plurality of decoded channel signals to eliminate crosstalk between the plurality of decoded channel signals; and

 outputting the plurality of decoded channel signals which have been subjected to crosstalk cancellation to each of the number of speaker units,

15 wherein the crosstalk cancellation includes further processing to generate a smoothed frequency envelope.

2. The method of claim 1, wherein the localization cues includes at least one selected from a group comprising: an up-down dimension, a left-right
20 dimension, a front-back dimension, an azimuth angle and an elevation angle.

3. The method of claim 1, wherein the enlarged location with optimal three-dimensional audio perception allows a listener to move about as the enlarged location relates to a boundary which encompasses a plurality of
25 positions with optimal three-dimensional audio perception.

4. The method of claim 1, wherein each speaker unit includes at least one speaker driver.

30 5. The method of claim 1, wherein the crosstalk cancellation is performed to cause a listener to perceive audio to be emanated from virtual speakers.

6. The method of claim 1, wherein derivation of the three-dimensional encoded localization cues is based on providing a listener with a fully spatial sound effect.

5 7. The method of claim 1, wherein the smoothed frequency envelope is reconstructed from truncated cepstrals derived from converting each of the plurality of decoded channel signals into the cepstrum spectrum.

8. The method of claim 7, wherein the smoothed frequency envelope
10 minimizes timbre artifacts, the timbre artifacts being high peaks and low valleys in the cepstrum spectrum of each of the plurality of decoded channel signals.

9. The method of claim 1, wherein optimal three-dimensional audio
15 perception relates to a fully spatial sound effect.

10. The method of claim 1, further including summing the plurality of decoded channel signals which have been subjected to crosstalk cancellation before output to each of the number of speaker units.