Abstract: A method for creating a stereophonic sound image out of a monaural signal (200) combines two sub-methods. Comb filters decorrelate the left (210) and right (211) channel signals. Intra-aural difference cues, such as an Intra-Aural Time Difference (ITD) and an Intra-aural Intensity Difference (IID) separated channels. Strict complementary (SC) linear phase FIR filters divide the incoming monaural signal into three frequency band separation. The comb filters and ITD/IID applied to the low and high frequency bands create a simulated stereo sound image for the instruments other than human voice. Listening tests indicate that this invention provides a wider stereo sound image than previous methods, while retaining human voice centralization. Since the comb filter solution and ITD/IID solution can share the same filter bank, the computational cost of this method is almost the same as the previous method.
STEREO SYNTHESIZER USING COMB FILTERS
AND INTRA-URAL DIFFERENCES

The technical field of this invention is stereophonic audio synthesis applied to
enhancing the presentation of both music and voice for more pleasant sound quality.

BACKGROUND

Currently, most commercial audio equipment has stereophonic (stereo) sound
playback capability. Stereo sound provides a more natural and pleasant quality than monaural
(mono) sound. Nevertheless there are still some situations which employ mono sound signals
including telephone conversations, TV programs, old recordings, radios, and so forth. Stereo
synthesis creates artificial stereo sounds from pIain mono sounds attempting to reproduce a
more natural and pleasant quality.

US Application Nos. 11/560,397 and 11/560,387, both filed 16 November 2006,
describe two different synthesis algorithms. The first of these [see, Application No.
11/560,397] applies comb filters [referred to as complementary linear phase FIR filters] to a
selected range of frequencies. Comb filters are commonly used in signal processing. The
basic comb filter includes a network producing a delayed version of the incoming signal and
a summing function that combines the un-delayed version with the delayed version causing
phase cancellations in the output and a spectrum that resembles a comb. Stated another way,
the composite output spectrum has notches in amplitude at selected frequencies. When
arranging separate comb filters to produce allocated notches of at different frequencies for
left and right channels, the outputs from the both channels become uncorrelated. This causes
the band-selected sound image to be ambiguous and thus wider. Typically, the purpose of
band selection is to centralize just the human voices. The second [see, US Application No.
11/560,387] describes the use of an Intra-Aural Time Difference (ITD) and an Intra-Aural
Intensity Difference (HD). This simulates the cultural fact that, in many live orchestras and
some rock bands, the low instruments tend to be located toward the right and the high
instruments on the left. To do this, the incoming mono signal is split into three frequency
bands and then sent to left and right channels with different delays and gains for each
channel, so that the band signals add up to the original, but with ITD and HD in low and high
bands respectively.
FIG. 1 illustrates a functional block diagram of a stereo synthesis circuit using intra-aural time difference (ITD) and an intra-aural intensity difference (HD). The input monaural sound 100 is split into three frequency ranges using high pass filter 101, mid-band pass filter 102 and low pass filler 103. Mid-band frequencies 119 are passed through sample delayA 104 and sample delayD 107. High pass frequencies 121 are passed to sample delayB 105 and low pass frequencies 124 are passed to sample delayC 106. The output of sample delayB 105 supplies the input of high band attenuation 108 which forms signal 123. The output of sample delayC 106 supplies the input of low band 109 which forms signal 126. The resulting six signal components 121 through 126 are routed to two summing networks 110 and 111. Summing network 110 combines high pass output 121, mid-band delayed output 122 and low pass delayed and attenuated output 126. The resulting left channel signal 116 is amplified by left amplifier 112 and passes to left output driver 114. In similar fashion, summing network 111 combines low pass output 124, mid-band delayed output 125 and high pass delayed and attenuated output 123. The resulting right channel signal 117 is amplified by right amplifier 113 and passes to right output driver 115.

SUMMARY

The invention provides a new method for creating a stereophonic sound image out of a monaural signal. The method combines two synthesis techniques. In the first technique comb filters de-correlate the left and right channel signals. The second technique applies intra-aural difference cues. Specifically this invention applies intra-aural time difference (ITD) and intra-aural intensity difference (IID) cues. The present invention performs a three-frequency band separation on the incoming monaural signal using strictly complementary (SC) linear phase FIR filters. Comb filters and ITD/IID are applied to the low and high frequency bands to create a simulated stereo sound image for instruments other than human voice. Listening tests indicate that the method of this invention provides a wider stereo sound image than previous methods, while retaining human voice centralization. Since the comb filter computation and ITD/IID computation can share the same filler bank, the invention does not increase the computational cost compared to the previous method.
BRIEF DESCRIPTION OF THE DRAWINGS

FIG. 1 (Prior Art) illustrates the basic principles of ITD and UD implemented in functional block diagram form;

FIG. 2 illustrates the block diagram of the stereo synthesizer of this invention;

FIG. 3 illustrates the block diagram of each of comb filter pairs used in the stereo synthesizer of this invention; and

FIG. 4 illustrates a portable music system such as might use this invention.

DETAILED DESCRIPTION OF THE EMBODIMENTS

The stereo synthesizer of this invention combines the best features of two techniques employed in prior art. Comb filters provide wider sound image and the combination of ITD/IID gives sound quality more faithfully reproducing the character of the original mono signal. This application describes a composite method that combines the two algorithms creating a wider sound image than the two methods provide individually. Since the two algorithms can share the same filter bank, which is three strictly complementary (SC) linear phase FIR filters, the integrated system can maintain a simple structure and the computational cost does not unduly increase.

FIG. 2 illustrates the block diagram of the stereo synthesizer of this invention. First, the incoming monaural signal 200 is separated into three regions using three SC FIR filters: (a) a low pass filter (LPF) \( H_{l}(z) \) 201; a band pass filter (BPF) \( H_{m}(z) \) 202; and a high pass filter (1IPF) \( H_{h}(z) \) 203. The outputs from \( H_{l}(z) \) and \( H_{h}(z) \) are processed with the respective comb filters 208 and 218 to create left channel 210 and right channel 211 signals with a simulated stereo sound image. The comb filter outputs for each channel are mixed with gains and delays, in order to generate ITD and HD. The output 204 from \( H_{m}(z) \) 202 is added to these simulated stereo signals in summing networks 205 and 206, so that the total output signal sums up to the original, but with frequency-band-partly widened sound. Respective optional equalization (EQ) filters 207 and 217 compensate for the frequencies that might be distorted by the notches of the comb filters 208 and 218. In practice, the low band EQ filter \( Q_{l}(z) \) 207 and high band EQ filter \( Q_{h}(z) \) 217 are designed as respective low and high shelving filters.

In FIG. 2, \( H_{l}(z) \) 201, \( H_{m}(z) \) 202, and \( H_{h}(z) \) 203 are said to be strictly complementary to each other only if:
\[ H_1(Z) + H_J(Z) + H_h(z) = c Z^{N/2} \quad (1) \]

is satisfied, where \( c = 1 \), in particular. Thus just adding all these filter outputs perfectly reconstructs the original signal. It is also important to make these FIR filters be phase linear with an even number order \( N \). With the choice \( N_0 = N/2 \), equation (1) can be written as:

\[ H_1(Z) + H_n(Z) + H_h(z) = z^{N/2} \quad (2) \]

Substituting \( z = e^{j\omega} \) and recognizing that \( H_i(e^{j\omega}), H_n(e^{j\omega}) \) and \( H_h(e^{j\omega}) \) are linear phase whose phase terms are given as \( e^{j\omega N/2} \), we have the frequency response relationship among the three filters as:

\[ \left| H_i(e^{-j\omega}) \right| + \left| H_n(e^{-j\omega}) \right| + \left| H_h(e^{-j\omega}) \right| = 1 \quad (3) \]

Let \( H(z) \) be the low pass filler (LPF) and \( H_h(z) \) be the high pass filter (UPF). Then \( H_{BP}(z) \) will be a band-pass filter (BPFO). The output from low pass filter \( (H_i(z)) \) is calculated as:

\[ Y_1(\alpha) = \sum_{i=0}^{N-1} I_i X(\alpha i - i) \quad (4A) \]

and the output from high pass filter \( (H_h(z)) \) is calculated as:

\[ y_h(n) = \sum_{i=0}^{N-1} I_i X(\alpha i - i) \quad (4B) \]

with \( h_f(n) \) and \( h_h(n) \) designating the respective impulse responses. Then the other output can be calculated just from:

\[ y_{ra}(n) = x(n - N/2) - Y_1(\alpha) - y_h(n) \quad (5) \]
Both equation (3) and equation (5) illustrate the benefit of using the SC linear phase FIR filters. Implementing a low pass filter and a high pass filter and just subtracting their outputs from the input signal gives a band pass filter output. This means that the major computational cost is for calculating only two filter outputs out of the three.

FIG. 3 illustrates the block diagram of each comb filter pair 208 and 218 used for stereo synthesis. Two comb filters are employed in each of the left and right output channels. Let Co(z) and Ci(z) denote the respective transfer functions for the left and right channels, then:

\[
\begin{align*}
C_0(z) &= \frac{(1 \pm \alpha z^{-D})}{(1 + \alpha)} \\
C_i(z) &= \frac{(1 \mp \alpha z^{-D})}{(1 + \alpha)}
\end{align*}
\]

(6)

where: D is a delay that controls the stride of the notches of the comb; and \( \alpha \) controls the depth of the notches. Typically \( 0 < \alpha \leq 1 \). The magnitude responses are given by:

\[
|C_0(e^{-j\omega})| = \sqrt{1 - \frac{4\alpha}{(1 + \alpha^2)} \sin^2 \frac{\omega D}{2}}
\]

(7A)

\[
|C_i(e^{-j\omega})| = \sqrt{1 - \frac{4\alpha}{(1 + \alpha^2)} \cos^2 \frac{\omega D}{2}}
\]

or

\[
|C_0(e^{+j\omega})| = \sqrt{1 - \frac{4\alpha}{(1 + \alpha^2)} \cos^2 \frac{\omega D}{2}}
\]

(7B)

\[
|C_i(e^{+j\omega})| = \sqrt{1 - \frac{4\alpha}{(1 + \alpha^2)} \sin^2 \frac{\omega D}{2}}
\]
The applicable magnitude response depends on the signs of the multiplier that are applied to the delayed-weighted path. Equations (7A) and (7B) show that both filters have peaks and notches with a constant stride of $2\pi/D$. The peaks of one filter are placed at the notches of the other filter and vice-versa. This dc-correlates the output channels resulting in the sound image becoming ambiguous and thus wider.

In a spatial hearing, a sound coming from left side of a listener arrives at the right ear of the listener later than the left ear. The left side sound is more attenuated at the right ear than at the left ear. The intra-aural time difference (ITD) and intra-aural intensity difference (HD) provide sound localization cues that make use of these spatial hearing mechanisms.

Referring back to FIG. 2, different weights and delays are applied to the left and right channels of the comb filter output. For $w > 1$ and $\tau > 0$, the listener will perceive the high pass filtered sound is coming from left side, because the right channel signal is attenuated and delayed. Similarly, the low pass filtered sound will seem to come from right side. This arrangement simulates many live orchestras and some rock bands, in which the low instruments tend to be located towards the right and the high instruments on the left. This produces wider sound image for the entire stereo output than by just employing the comb filters.

The following is a description of a design example. In this example, a sampling frequency was chosen 44.1 kHz. The SC FIR filters were designed using MATLAB. This example uses order 32 FIR $H_l(z)$ and $H_h(z)$ selected based on the least square error prototype. The cut off frequency of the low pass filter $H_l(z)$ was chosen as 300Hz and the cut off frequency of the high pass filter $H_h(z)$ was chosen as 3kHz. These selections put the lower formant frequencies of the human voice in their stop bands. The band pass filter $H_m(z)$ was calculated using equation (5). This was confirmed as providing a band pass filler magnitude response. The low and high pass filters were implemented using equation (4).
The comb filters were designed as follows. Comb filters $C_{l,0}$ and $Q_{l,1}$ for the low channel:

\[
C_{l,0} = \frac{1 + 0.7z^D}{1 + 0.7} \\
C_{l,1} = \frac{1 - 0.7z^D}{1 + 0.7} \tag{8A}
\]

Comb filters $C_{h,0}$ and $Ch_{.i}$ for the low channel:

\[
C_{h,0} = \frac{1 - 0.7z^D}{1 + 0.7} \\
C_{h,1} = \frac{1 + 0.7z^D}{1 + 0.7} \tag{8B}
\]

where: $D \approx 8$ milliseconds corresponding to 352 filter taps was selected for the all comb filters. The purpose of flipping the signs of the multiplier for low band and high band was to cancel the notches of each other in the transition region of LPF and HPF. This contributed to further centralizing the human voice, while the sound image for the other instruments was unaffected. In this example only intra-aural-intensity differences (UD) were implemented. The intensity difference $w$ was 1.4.

Brief listening confirmed that this method provides wider sound image than the two previous methods, while the voice band signals were centralized the same as with those methods.

Referring back to FLU. 2, the SC FIR filters produce most of the computational load. This is because the comb filters can be considered as order 1 FIR implementations and 1ID/ITD can be considered as order 0 FIR implementations. The low pass filter and the high pass filter require much longer taps to obtain a desired frequency band separation. The EQ filters, if present, can be designed with first order infinite impulse response (HR) filters, which is of lower computational cost. Thus a make computation comparison between the present method and previous methods can be made by just considering the SC FIR filters that implement exactly the same filter bank structure. The computational cost does not differ
appreciably. The prior methods employ two band separations using a band-pass and a band stop filter, where only one of the two must be actually be implemented because of the SC linear phase FIR property. This means that the method of the present invention is one-filter-heavier than the earlier approach. However, low-pass filters (LPF) and high pass filters (HPF) can be designed with shorter filter taps than band-pass filters (BPF). Indeed order 32 finite impulse response (FIR) filters were used for low pass and high pass filters in the research leading to this invention. These FIRs employ about one-half the taps used in prior methods for the band pass filter (BPF). As a result the computational cost for this invention is essentially the same as previous methods.

This invention is a stereo synthesis method that combines two previous methods, the comb filter method and intra-aural difference method. Through listening tests it has been confirmed that this method provides a wider stereo sound image than previous methods, while the human voice centralization property is retained. The computational cost of the present invention is almost the same as the previous methods.
What is claimed is:

1. A method of synthesizing stereo sound from a monaural sound signal comprising the steps of:
   - low pass filtering the monaural sound signal;
   - producing first and second decorrelated low pass filtered signals;
   - producing respective first and second low pass intra-aural difference signals from said first and second decorrelated low pass filtered signals;
   - band pass filtering the monaural sound signal;
   - high pass filtering the monaural sound signal;
   - producing first and second decorrelated high pass filtered signals;
   - producing respective first and second high pass intra-aural difference signals from said first and second decorrelated high pass filtered signals;
   - summing said first low pass intra-aural difference signal, said band pass signal and said second high pass intra-aural difference signal to produce a first stereo output signal; and
   - summing said second low pass intra-aural difference signal, said band pass signal and said first high pass intra-aural difference signal to produce a second stereo output signal.

2. The method of Claim 1, wherein:
   - said steps of producing first and second decorrelated low pass filtered signals and producing first and second decorrelated high pass filtered signals each include filtering an input with respective first and second complementary comb filters, wherein frequency peaks of said first comb filter matches frequency notches of said second comb filter and frequency notches of said first comb filter matches frequency peaks of said second comb filter.

3. The method of Claim 2, wherein:
   - said first comb filter \( C_0 \) is calculated by:
     \[
     C_0 = (I + \alpha z \beta)(I + \alpha)
     \]
said second comb filter $C_i$ is calculated by:

$$C_i = \frac{(1 - \alpha z^D)}{(1 + \alpha)}$$

where: $D$ is a delay factor; and $\alpha$ is a scaling factor.

4. The method of Claim 1, wherein:

said step of producing said first decorrelated low pass filter signal $C_{i,0}$ is calculated by:

$$C_{i,0} = \frac{(1 + \alpha z^D)}{(1 + \alpha)}$$

said step of producing said second decorrelated low pass filter signal $C_{i,1}$ is calculated by:

$$C_{i,1} = \frac{(1 - \alpha z^D)}{(1 + \alpha)}$$

said step of producing said first decorrelated high pass filter signal $Q_{i,0}$ is calculated by:

$$C_{h,0} = (1 - \alpha z^D) (1 + \alpha)$$

and

said step of producing said second decorrelated high pass filter signal $C_{h,1}$ is calculated by:

$$C_{h,1} = (1 + 0.7 z^D) / (O + 0.7)$$

where: $D$ is a delay factor; and $\alpha$ is a scaling factor.
5. The method of Claim 3 or 4, wherein:
   - the delay $D$ is 8 mS; and
   - the scaling factor $\alpha$ is within the range $0 < \alpha \leq 1$.

6. The method of Claim 1, wherein:
   - said steps of producing first and second intra-aural difference low pass filtered signals and producing first and second intra-aural difference high pass filtered signals each include providing a differential gain on said first and second decorrelated signals.

7. The method of Claim 1, wherein:
   - said step of producing first and second intra-aural difference low pass filtered signals comprises amplifying said first decorrelated low pass signal with a first gain to produce said first intra-aural difference low pass filtered signal and amplifying said second intra-aural difference low pass filtered signal with a second gain lower than said first gain;
   - said step of producing first and second intra-aural difference high pass filtered signals comprises amplifying said first decorrelated high pass signal with said second gain to produce said first intra-aural difference high pass filtered signal and amplifying said second intra-aural difference high pass filtered signal with said first;
   - said step of summing to produce said first stereo output signal produces a left stereo signal; and
   - said step of summing to produce said second stereo output signal produces a right stereo signal.

8. The method of Claim 1, wherein:
   - said steps of producing first and second intra-aural difference low pass filtered signals and producing first and second intra-aural difference high pass filtered signals each include delaying one of said decorrelated signals.

9. The method of Claim 1, wherein:
   - said step of producing first and second intra-aural difference low pass filtered signals comprises delaying second intra-aural difference low pass filtered signal;
   - said step of producing first and second intra-aural difference high pass filtered signals comprises delaying said first decorrelated high pass signal;
said step of summing to produce said first stereo output signal produces a left stereo signal; and

said step of summing to produce said second stereo output signal produces a right stereo signal.

10. The method of Claim 1, wherein:
said steps of low pass filtering the monaural sound signal, band pass filtering the monaural sound signal and high pass filtering the monaural sound signal comprises using strict complementary (SC) linear phase finite impulse response (FIR) filters.

11. The method of Claim 10, wherein:
said step of low pass filtering is calculated as:

\[ y_l(n) = \sum_{i=0}^{L} h_l(i)x(n - i) \]

said step of high pass filtering is calculated as:

\[ y_h(n) = \sum_{i=0}^{R} h_h(i)x(n - i) \]

and

said step of band pass filtering is calculated as:

\[ y_B(n) = x(n - N/2) - y_l(n) - y_h(n) \]

where: N is a number of filter taps; \( h_l(i) \) is the low pass filter impulse response; \( h_h(i) \) is the low pass filter impulse response; and \( i \) is an index variable.