Modulation Spliced Transform Binaural Cue Coding Algorithm-based Encoder

Xiaoping Xu , Jie Rong [a] and Anqi Wang

Electronic Information and Control Engineering,Beijng university of technology, China

Abstract. Perceptual parameters ICLD, ICTD and ICC has a very important practical significance, so the use of these parameters, the researchers propose parameters based on FFT frequency domain binaural cue coding method to achieve a better stereo encoding quality. Compared with the FFT, modulation spliced transform (Modulated Lapped Transform, MLT) has an energy concentration, facilitating the removal of the characteristics of inter-blocking effect. Therefore, this article will be applied to binaural cue coding MLT transform them, to achieve a binaural cue coding method based MLT transform

1 Implementation of the encoder

Left channel signal and right channel signal input, each frame of MLT transform signal independently, the time-domain signal into a frequency domain MLT coefficients; extract relevant information left channel signal and right channel signal in the frequency domain i.e. between spatial parameters, and use the extracted spatial parameters will do the next right-channel signal is mixed to mono MLT coefficient; and finally the mono signal in the frequency domain inverse MLT transform into the time domain, after windowing, and then through the stack then adding the processed output. At the same time, spatial parameter information obtained by coding a frequency domain is formed independently of the secondary side stereo coding information.After the encoded bit stream is formed in two parts, one containing the inter-channel stereo correlation parameters (ICDL, ICTL and ICC), the other part is a single channel signal after mixing.

2 Time - frequency transform

In the sub-frame processing voice and audio signals usually small error algorithm will bring inter-block effect, seriously affecting the quality of the signal processing, so as to avoid this effect, can be introduced into the language MLT transform audio signals processing them, when used for the conversion of audio signals in the frequency domain.Splicing modulation signal conversion can be accurately reconstructed, while both DCT (Discrete Cosine Transform, DCT) with similar energy concentration characteristics. MLT transform sampling precision, good signal reconstruction effect, a linear transformation.When modulation spliced transform, signal processing between frames will be half of the adjacent overlapping. If the input signal of 16kHz, select the frame length is 20ms, the actual processing frame length is 640 sampling points, and each performs a MLT transform the number of samples needed to get to 1280, so every time you need to change the modulation spliced two frames signal. After the MLT transform the current frame is completed, when the next frame is modulated spliced transformation, retention time domain information of

Corresponding author: 376760445@qq.com

the frame as the first half of the time domain of information processing, while 640 samples will be used later as a signal for one frame the second half of temporal information processing frame. By analogy, then transform modulation overlapping coefficient can be obtained when each time-frequency transform.Forward modulating expression of spliced transformation：

$$X(k)=\sum_{n=0}^{2M-1}x(n)P(n,k) \quad (1.1)$$

Where, k = 0,1, ... M-1, M is the frame length, P (n, k) too - domain aliasing filter offset basic functions by cosine modulation of smooth window to get this function, such as a specific expression formula（1.2）：

$$P(n,k)=h(n)\sqrt{\frac{2}{M}}\cos(\frac{(2n+M+1)(2k+1)\pi}{4M}) \quad (1.2)$$

Wherein the window function h (n) of the low-pass half-band filter. IMLT need to use the current frame X (k) and the previous frame Xp (k), in order to restore the original signal becomes x (n), the formula (1.3) shown in the expression for the MLT inverse transform equation.

$$x(n)=\sum_{k=0}^{M-1}(X(k)P(n,k)+X^{p}(k)P(n+M,k)) \quad (1.3)$$

Usually used as formula (1.4) shown in the window function as a window to allow the signal reconstruction operation to achieve the best results.

$$h(n)=\sin(\frac{(2n+1)\pi}{4M}) \quad (1.4)$$

MLT operation for ultra-wideband speech input audio signal according to the following process of implementation:Let x (n) for each input signal MLT transform recent 1280 speech signal samples, each containing 640 samples of each of the previous and current frames.Where: x (0) is the first sampling point, 0≤n≤1279.After conversion output 640 MLT transform coefficients smlt (m), where: 0≤m <640MLT transform is:

$$s_{mlt}(m)=\sum_{n=0}^{1279}\sqrt{\frac{2}{640}}\sin\left(\frac{\pi}{1280}(n+0.5)\right)\cos\left(\frac{\pi}{640}(n-319.5)(m+0.5)\right)x(n)$$

(1.5)

MLT transform decomposed into a window function, overlapping, accumulate and type Ⅳ discrete cosine transform. Window functions overlap, accumulating detailed process is:

$$w(n)=\sin\left(\frac{\pi}{1280}(n+0.5)\right) \quad 0\le n\le 640 \quad (1.6)$$

Thus, we get v (n) be able to use type Ⅳ discrete cosine transform MLT transform operations to achieve the original signal. MLT transform can be rewritten in the form below：

$$s_{mlt}(m)=\sum_{n=0}^{639}\sqrt{\frac{2}{640}}\cos\left(\frac{\pi}{640}(n+0.5)(m+0.5)\right)v(n) \quad (1.7)$$

The current practice, FFT fast algorithm has been applied to the DCT by adding a fast algorithm can effectively reduce the complexity of the DCT.After the MLT transform changes made above, in programming practices can be considerably reduced when the amount of computation to reduce the complexity of the algorithm.

3 Subband Division

During the encoding process the audio signal needs to subband division. MLT spectral coefficients subband division process using a uniform way division, in this paper the frequency domain signal lines per frame value N is 640, the 640 spectral coefficients B = 20 is divided into sub-bands, such as sub-band b, where A0 = 0, sub-band border demarcation as shown in Table 1, where b is the sub-band number, kb b represents the initial sub-band frequency points.

Table.1 Sub-band boundary comparison

b	0	1	2	3	4	5	6
k_b	0	32	64	96	128	160	192
b	7	8	9	10	11	12	13
k_b	224	256	288	320	352	384	416
b	14	15	16	17	18	19	20
k_b	448	480	512	544	576	608	639

4 Mixing down

In order to allow the input signal contains the original audio signal information will be taken down mixed multi-channel signal. Between simple superposition channel signal such that the signal can be weak or enlarge the content, and causing signal and the signal power of each channel are not equal, larger or smaller.

Hybrid technology is usually taken down in order to avoid the occurrence of this problem, this technology and signal proper balance, so that the signal power is approximately equal to the signal power of each channel.

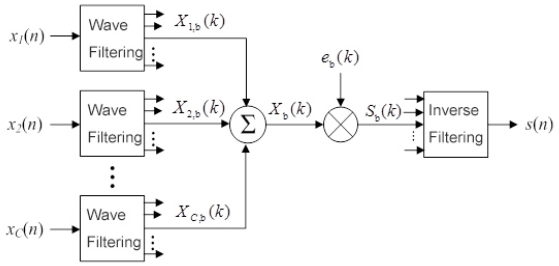

Figure 1 Down multi-channel signal mixing process schematic

As shown in the figure is down mixed multi-channel signal process. Downmix signal forming process, the first input audio signal xc (n) (1≤c≤C) by filtering decomposed into a plurality of sub-band signals Xc, b (k) (Xcb (k) for the MLT coefficients, b is the sub-the subscript, 0≤b≤B), followed by the corresponding sub-band signal for each channel signal obtained by adding Xb (k), and then multiplied by the weighting factor eb (k) generate Sb (k), the last of the B Sb (k) (for the MLT coefficients) of inverse filtering generation and signal s (n), where C is the number of channels, B is divided into sub-band number.The following two-channel signal will be deduced as an example the case of stereo weighting factor eb (k) of the expression:Dual-channel short-term power of the b sub-band within the mixed-signal estimation formula (1.8) below.

$$\begin{aligned} P_b &= \sum_{k=A_{b-1}}^{A_b-1} |S_b(k)|^2 \\ &= \sum_{k=A_{b-1}}^{A_b-1} |e_b(k)(X_{1,b}(k) + X_{2,b}(k))|^2 \\ &= {e_b}^2(k) \sum_{k=A_{b-1}}^{A_b-1} |X_{1,b}(k) + X_{2,b}(k)|^2 \end{aligned} \tag{1.8}$$

Short-term power estimate of the left and right channels and formula (1.9) below.

$$\begin{aligned} P_b' &= P_{1,b} + P_{2,b} \\ &= \sum_{k=A_{b-1}}^{A_b-1} |X_{1,b}(k)|^2 + \sum_{k=A_{b-1}}^{A_b-1} |X_{2,b}(k)|^2 \end{aligned} \tag{1.9}$$

Mixing process need to satisfy Pb = Pb ', can be derived eb (k) of the formula as in formula (1.10) below.

$$e_b(k) = \sqrt{\frac{\sum_{k=A_{b-1}}^{A_b-1} |X_{1,b}(k)|^2 + \sum_{k=A_{b-1}}^{A_b-1} |X_{2,b}(k)|^2}{\sum_{k=A_{b-1}}^{A_b-1} |X_{1,b}(k) + X_{2,b}(k)|^2}} \tag{1.10}$$

Left and right channel signals can be omitted in the subscript sub b, formula (1.11).

$$e_b(k) = \sqrt{\frac{\sum_{k=A_{b-1}}^{A_b-1} |X_1(k)|^2 + \sum_{k=A_{b-1}}^{A_b-1} |X_2(k)|^2}{\sum_{k=A_{b-1}}^{A_b-1} |X_1(k) + X_2(k)|^2}} \tag{1.11}$$

References

1. Ping Ping, Yicheng .Song Digital communications, Electronic Industry Press，2008．562
2. Xuanpeng Li. Based on spatial perception of stereo coding information [J] Southeast University, 2006,(04).