A SOFTWARE TOOL FOR SPATIAL LOCALIZATION CUES*

Cristian NEGRESCU, Dumitru STANOMIR, Amelia CIOBANU[1]

This paper focuses on designing a software tool that will facilitate the manipulation of the most relevant spatial localization cues. To achieve this goal, we chose binaural cue coding (BCC) as a suitable application that best emphasizes these cues. Further, methods for extracting binaural cues were explored and then a low cost computational method was implemented. Finally, a description of the designed graphical user interface (GUI) was included.

1. INTRODUCTION

There is a wide variety of applications in both speech and audio processing which make use of spatial localization cues. From this variety we chose binaural cue coding which is a multichannel spatial rendering technique based on one down-mixed audio channel and spatial localization cues [4, 5]. Considering this particular application, we have developed a software instrument which allows us to gain a better control on these cues, a higher level of knowledge on how they interact and how they influence one another. This tool also serves as a mean of validating certain results met in literature (*i.e.* the range of values, the spectral domain in which they are most significant for spatial localization) and provides a good support for future development in sound field manipulation.

The paper includes a chapter which outlines the basic ideas regarding BCC, followed by a chapter where the procedures for BCC parameters extraction, sum signal construction and multichannel audio signal synthesis, derived from acoustical considerations, are described. The last part of the paper presents the designed software tool along with conclusions.

2. BCC-BASIC IDEAS

In order to be able to synthesize a source image we need a set of parameters that contain the spatial localization cues and a broadband channel obtained as a result of down-mixing the channels involved. This set of parameters is either extracted from the original multichannel signal or loaded from a table that stores a different set of parameters for each image source that needs to be (re)created.

The original BCC application aims at compressing the multichannel audio signal to be stored/transmitted. In this case, the global data rate is significantly

* Presented at SISOM 2007 and Homagial Session of the Commision of Acoustics, Bucharest, 29–31 May 2007.

Electronics, Communication and Information Technology Faculty, University "Politehnica" of Bucharest, P.O. Box 35–145, Bucharest, 050461, Romania, E-mail: negrescu@elcom.pub.ro

reduced given the fact that spatial localization cues can be transmitted with only a few kb/s [1].

The multichannel audio signal is split and follows two different paths. Along the first path, the multichannel audio signal is down-mixed to one wideband channel called the sum signal and compressed by a suitable encoder. In this way the data rate is substantially reduced. During the second processing path, the multichannel audio signal is passed through the BCC analyzer where the localization cues are extracted and represented in a compact manner by a limited number of parameters related to the binaural cues. This side information can be transmitted with a very low data rate. The BCC concept can be/is quite efficient because lower data rates are required for the transmission of the sum signal and the side information, in comparison with the transmission of all the input channels. The BCC analyzer provides a set of parameters that contain the localizations cues. These parameters are transformed into "interchannel cues" and sent to the receiver where they are processed locally. The interchannel cues are then transferred to the BCC synthesizer and the sum signal is transformed back into a multichannel signal.

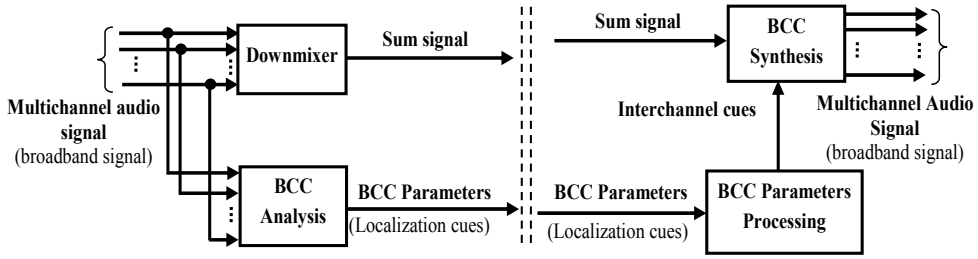

Fig. 1 – BCC in an application for a multichannel audio compression – the transmitter and the receiver.

In the following, the BCC analyzer and synthesizer for natural rendering will be described. There is a reference approach that uses Cochlear Filter Banks (CFB) [2, 4] for the analysis with a time-frequency response similar to the human inner ear. Also, one can consider another approach [1, 4] based on the Fast Fourier Transform (FFT) [1, 4]. The last one, which was finally implemented in this paper, has the advantage of a low computational complexity. It is important to note that its final performances are not significantly different (compared to those obtained with the CFB). For the FFT approach, the analysis is done block-wise, as opposed to CFB where the analysis is done sample by sample.

In order to build such an encoder we obviously need to deal with the perceptual localization phenomena. The goal set here is to obtain a robust encoder, so a subset of parameters was chosen from a set of localization cues. The selected parameters should be large enough to allow the recreation of a given spatial image, with sufficient accuracy. Therefore, only the perceptual factors, which contribute to

the robustness of the sound image created by the decoder, will be emphasized. Psychoacoustic considerations show that the most important and also the most robust binaural cues, used for localization, are ILD, ITD and IC [3, 6]. The above mentioned cues refer to the listener's ear input signals and their values depend on the playback scenario (loudspeaker or headphones). We only have limited control over the binaural cues since the head related transfer function (HRTF) are not precisely known and can only be roughly estimated. Due to the fact that the encoder's target is to recreate efficiently the spatial image, we can assume that the impact of the HRTF on the spatial image is similar for the playback of the original and synthesized signal. It is clear now the need to introduce [1] the terms "inter-channel level difference" (ICLD), "inter-channel time difference" (ICTD), and "inter-channel correlation" (ICC). Further, we shall use these new terms to implement the BCC analyzer and synthesizer.

3. BCC ANALYSIS AND SYNTHESIS USING FFT

A convenient approach, useful for real time implementation, can be built if a frame-wise processing scheme is used, where the spectral decomposition is done via FFT algorithm. The broadband audio signals are individually processed (temporal segmentation, spectral decomposition, subband partitioning), as shown in Fig. 2. Afterwards, for each pair of two signals x and y, and also for each pair of corresponding spectral partitions the ICLD, ICTD and ICC are computed.

3.1. TEMPORAL SEGMENTATION AND WINDOWING

In all the processing schemes where the spectral decomposition is based on FFT, the analysis is done frame by frame, and consequently, the input signal is multiplied by a window function with a compact support. We consider that a Hann type window, with 50% time advance is a reasonable solution. The window is padded with zeros at both ends (to achieve positive and negative time shifts). In this way we avoid distortion generated by circular time shifts, inherently present when we introduce spectral phase shifts. The lengths of the zero regions are chosen according to ICTD's range values. The size of the symmetrical window is N samples. The nonzero part has W samples length and corresponds to a Hann window. The zero part is Z samples at each end. Adjacent window are shifted by W/2 samples. The window was chosen such that the overlapping windows add up to a constant value of 1. Signal's reconstruction can be done by overlap-add technique. Due to the particularity of the chosen analysis window, the need for a synthesis window is eliminated and thus the BCC synthesizer is simplified. For the analysis-synthesis chain, the reconstruction is perfect, if the spectral components are not modified.

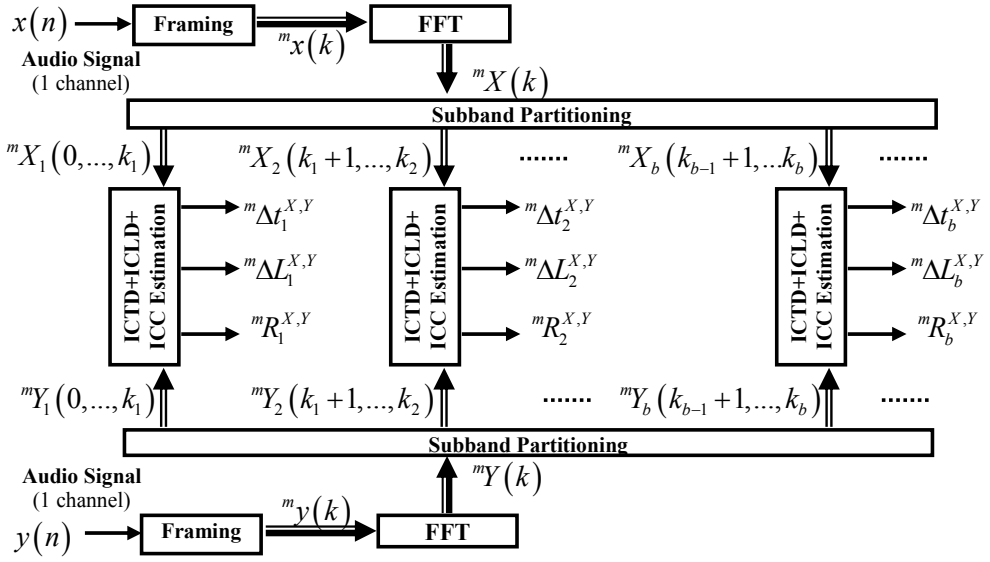

Fig. 2 – Spectral decomposition and the BCC analysis.

3.2. SPECTRAL DECOMPOSITION AND SUBBAND PARTITIONING

As discussed, FFT is applied to each frame. The obtained spectral components are then divided into non-overlapping partitions, each partition representing a critical band, characteristic to human hearing (resembles to frequency representation in the cochlea). The partition bandwidth was chosen equal to 2 ERB (Equivalent Rectangular Bandwidth).

3.3. BCC PARAMETER ESTIMATION

ICLD. First we need to compute, for each input channel, the power/energy within each partition. Then, ICLDs, in dB, are estimated by calculating the power/energy ratios of the corresponding partition of *x* and *y* (Fig. 3).

ICTD. At low frequencies (below about 1.5 kHz), the ICTD for each partition is computed by averaging the phase difference between the two signals. At medium and high frequencies (above 1.5 kHz) the group delay is relevant for spatial localization. To estimate group delay, first is computed the phase difference between the two channels. Then, for each partition, linear regression is used to determine the slope of the phase difference of the spectral coefficients. The group delay is proportional to the slope of the regression line. The corresponding algorithm is presented in Fig. 4.

ICC. This parameter denotes the degree of correlation between the two signals. After selecting the spectra of the two input channels, the magnitude coherence of these channels is computed. This is done, in a recursive manner, for each spectral component (Fig. 5).

For each spectral component, to obtain the correlation degree, the coherence estimates are averaged and normalized,

$$ {}^{m}R_{X,Y}(k)=\sqrt{\frac{{}^{m}\hat{\Phi}_{X,Y}(k)\ {}^{m}\hat{\Phi}^{*}_{Y,X}(k)}{{}^{m}\hat{\Phi}_{X,X}(k)\ {}^{m}\hat{\Phi}_{Y,Y}(k)}}=\frac{\left|{}^{m}\hat{\Phi}_{X,Y}(k)\right|}{\sqrt{{}^{m}\hat{\Phi}_{X,X}(k)\ {}^{m}\hat{\Phi}_{Y,Y}(k)}}\in[0,1]. \quad (3.1) $$

The inter-channel correlation for a given spectral partition (subband) can be obtained using the expression:

$$ \hat{ICC}_{X_b,Y_b}(0,m)\overset{\text{Not}}{=}{}^{m}R_{b}^{X,Y}=\sqrt{\frac{\sum_{k=k_{b-1}+1}^{k_b}{}^{m}R_{X,Y}^{2}(k)\ {}^{m}\hat{\Phi}_{X,X}(k)\ {}^{m}\hat{\Phi}_{Y,Y}(k)}{\sum_{k=k_{b-1}+1}^{k_b}{}^{m}\hat{\Phi}_{X,X}(k)\ {}^{m}\hat{\Phi}_{Y,Y}(k)}}= $$

$$ =\sqrt{\frac{\sum_{k=k_{b-1}+1}^{k_b}\left|{}^{m}\hat{\Phi}_{X,Y}(k)\right|^{2}}{\sum_{k=k_{b-1}+1}^{k_b}{}^{m}\hat{\Phi}_{X,X}(k)\ {}^{m}\hat{\Phi}_{Y,Y}(k)}}\in[0,1]. \quad (3.2) $$

The corresponding algorithm is presented in Fig. 6.

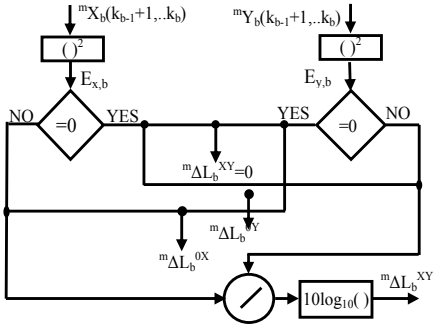

Fig. 3 – Algorithm for ICLD estimation.

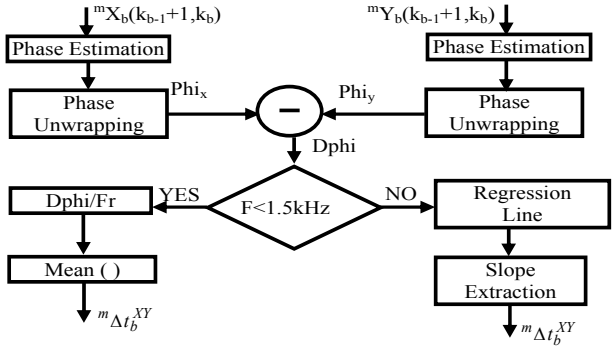

Fig. 4 – Algorithm for ICTD estimation.

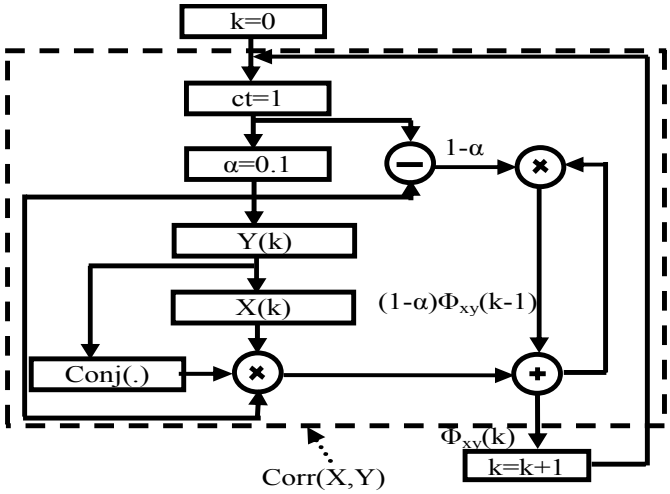

Fig. 5 – Spectral coherence (for one spectral component – *k*).

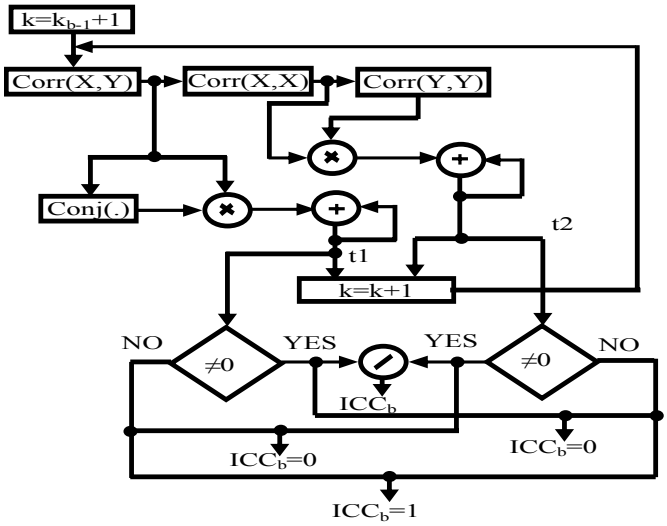

Fig. 6 – ICC in one subband ($k \rightarrow (k_{b-1}+1, k_b)$).

3.4. THE SUM SIGNAL

When the audio input signals are independent, the sum signal can be generated by simple addition, but when the input signals are no longer independent, we have to consider certain measures in order to avoid cancellation and/or undesired amplification of certain spectral components. To reduce this negative effect of the summation, the power of the sum signal must be equal to the sum of the input signal powers, in each partition. Further improvement can be achieved by implementing mechanisms for phase modification, which will diminish the negative effect of cancellation.

3.5. THE BCC SYNTHESIS

The start point for the audio signals synthesis is the sum signal. First, the sum signal is segmented temporally and converted to a spectral representation. The spectrum is divided into non-overlapping partitions. The partition bandwidth is 2ERB. Every spectral component of the sum signal is modified by using the localization cues provided by the BCC analyzer. The spectral coefficients for each output signal are obtained by multiplying each spectral coefficient of the sum signal with two factors: one that determines the level modification for each spectral coefficient(a positive real number that controls ICLD and ICC) and one that determines phase modification for each spectral coefficient (a complex number that controls ICTD). The spectra of the two output signals are converted back to the time domain.

4. THE SOFTWARE IMLPEMENTATION

To validate the extraction of the spatial localization cues considered in the application presented in the previous chapters, to verify the ability of building a multichannel audio signal, in which the interchannel cues are controlled in agreement with the extracted parameters, and also to outline the possibility of synthesizing a source image having within reach a set of localization parameters and a broadband channel, we have implemented a multichannel audio coding software application. The software was implemented under Matlab environment, which has a high level of flexibility regarding research, debugging and tuning activities, and was completed with a suggestive and friendly graphical user interface (GUI). A snapshot of this GUI is presented in Fig. 7.

As can be seen, the GUI window is divided into three main areas: two areas dedicated to analysis and synthesis and one area reserved for graphical displays. The *Browse* buttons allow the user to select the name and the path for the audio multichannel input file used for analysis or for the broadband channel used in the synthesis process. At the user's request, the selected signals can be visualized and/or listened (*Play* button).

The software can be used as an encoder/decoder or as an advanced multichannel audio analysis tool or as a synthesizer of source images. For the detailed analysis, the user has the possibility to adjust the analysis parameters according to his needs. Unfortunately, these parameters are not independent, that is why it is important to set a list of priorities. For example, if one desires a good spectral resolution, then the chosen length for the analysis window (N) should be large enough, but not too large because then the signal will not satisfy the stationary conditions. Another important parameter is Z (see §3.1), which dictates the range values for interchannel delays (positive and negative delays). A larger value for Z implies a smaller value for W ($N = 2 \cdot Z + W$), and thus the arithmetic complexity will increase and so will the global data rate for transmission. The presented software takes into account these aspects and allows the user to set either (N, W) or (N, Z) or (W, Z), computing the third parameter. After setting the analysis parameters, the analysis button is available. When the analysis process is complete, the localization parameters can be visualized. To increase the versatility for these

parameters, the software allows two visualizing modes: frame by frame visualization or whole temporal sequence. If one chooses the frame oriented visualization (*Frame Visualization*) (this is the visualization mode presented in Fig. 8), in the first two graphic areas the current frame of the channels involved are plotted, while in the next three graphic areas ICLD, ICTD and ICC are plotted as a function of frequency/2ERB.

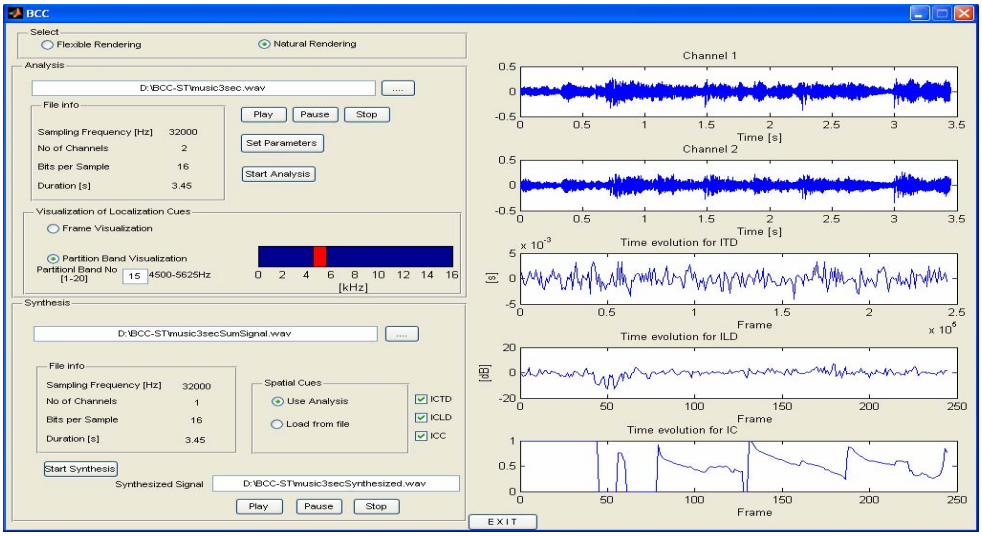

Fig. 7 – BCC-software tool-critical band visualization mode.

The current frame can be chosen manually (*Choose frame*), or if the zero value is selected (0 for *continuous*), the visualization is performed automatically, frame by frame. The second visualization mode is partition band oriented (*Partition Band Visualization*) and it is presented in Fig. 7. In the first two graphic areas the whole temporal sequence of the channels involved is presented, while in the next three graphic areas, the time evolution of each localization parameter is plotted. The partition band for which we wish to see the parameters' evolution is established manually (*Partition Band No*). In order to have a better sense regarding the spectral placement of the selected band, a small graphic is plotted next to the *Partition Band No* box, providing this information The red bar represents the position in frequency domain for the visualized parameters.

In the synthesis process, after selecting the broadband channel, depending on the application, the user can specify which set of localization cues should be utilized: the one extracted in the analysis stage (*Use Analysis*) or the one existing in a text file (*Load from file*). Also, the user can decide which cues from the selected set contribute to synthesis, by simply ticking the *ICTD, ICLD* and *ICC* boxes. In this way, one can estimate to a certain extent how strong and robust is each of these parameters in spatial localization of sound sources. When the synthesis process is complete, the file name and path name of the synthesized signal are specified and the file can be listened.

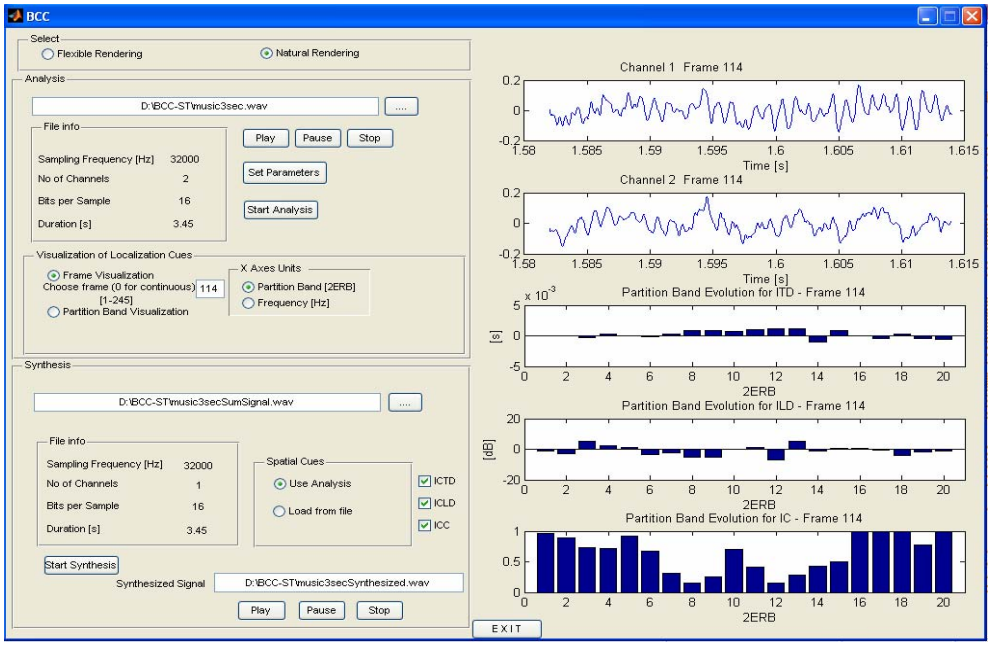

Fig. 8 – BCC-software tool – Frame Visualization mode.

For the presented version, the BCC encoder has the following characteristics:

– Sampling frequency: 32kHz
– No. of bits per sample: 8, 16, 20, 24, 32
– No. of channels: 2, 3, 4, 5, 6
– File format: .wav
– Analysis/Synthesis method: FFT
– Analysis/Synthesis frame length: $N = 1024$
– Frame advance: $W = 896$
– Guard interval: $Z = 64$
– No. of spectral partitions: $B = 20$
– Partition width: $2\,ERB$
– Averaging parameter for correlation: $\alpha = 0.1$
– Partition up to the ICTD estimates the time delay between signals chronograms: 10
– Partition from which ICTD estimates the time delay between the temporal signals envelopes: 11
– ICTD correction for ICLD estimation: no
– ICTD correction for low correlated signals: yes
– Method for building the sum channel: spectral adding
– Compensation of undesired attenuation for out of phase signals in the sum channel: no
– Compensation of the undesired amplification for in phase signals in the sum channel: yes
– Algorithm for channels decorrelation: yes
– Values range for decorrelation sequence –5dB...+5dB
– Values range for ICTD estimation and synthesis: –4ms...+4ms
– Values range for ICLD estimation and synthesis: full range (depending on No. bits/sample)
– Values range for ICC: 0...1

REFERENCES

1. BAUGMARTE, F., FALLER, C**.**, *Binaural Cue Coding – Part I: Psychoacoustic Fundamentals and Design Principles*, IEEE Trans. Speech Audio Processing**, 11**, pp. 509–519 (2003).
2. BAUGMARTE, F., *Improved audio coding using a psychoacoustic model based on a Cochlear Filter Bank*, IEEE Trans. Speech Audio Processing, **10**, pp. 495–503 (2002).
3. BLAUERT, J., *Spatial Hearing. The Psychophysics of Human Sound Localization,* Cambridge, MA: MIT Press, 1983.
4. FALLER, C., BAUGMARTE, F., *Binaural Cue Coding – Part II: Schemes and Applications*, IEEE Trans. Speech Audio Processing, **11**, pp. 520–531 (2003).
5. FALLER, C., MERIMAA, J., *MSource Localization in Complex Listening Situations: Selection of Binaural Cues base don Interaural Coherence*, J. Acoust. Soc. Amer., **116**, pp. 3075–3089 (2004).
6. KUHN, G., *Model of the Interaural Time Differences in the Azimuthal Plane*, J. Acoust. Soc. Amer., **62**, pp. 157–167 (1977).