ENCODING HIGHER ORDER AMBISONICS WITH AAC

Supervisor: Prof. Rajesh M. Hegde
TA: Gyanajyoti Routray

GROUP NO. - 8

AKASH SINGH (14048)
SUNIL KUMAR CHAWLA (14721)
VIKAS VERMA (14801)
VIMAL KUMAR MEENA (14804)

OBJECTIVE

- To reduce the bit rate needed for transmitting and storing HOA
- Conversion of AAC compressed B-format signal to D-format signal
- Wave Field Simulations to compare the errors as a function of distance for different no. of channels and different frequencies

METHODOLOGY

- Ambisonic Encoding/Decoding (A-format to B-format to D-format signal conversion)
- Advanced Audio Coding Encoding/Decoding
- Wavefield Analysis by comparing an original sound field with a processed sound field over a reproduction area

AMBISONIC ENCODING

A-Format, is produced by the Soundfield microphone itself and consists of the four signals from the microphone capsules - left-front, left-back, right-front and right-back.

We have assumed a single virtual source positioned at 30° in clockwise direction with respect to listener and is placed at infinite distance, and with no room reflections or reverberation included.

AMBISONIC ENCODING

B-Format encodes the directional information of a given three-dimensional sound-field (A-format signal) to four channels called W, X, Y, Z:

$$W=\frac{1}{k}\sum_{i=1}^{k} s_i[\frac{1}{\sqrt{(2)}}]$$ omnidirectional information

$$X=\frac{1}{k}\sum_{i=1}^{k} s_i[\cos\phi_i\cos\theta_i]$$ x-directional information

$$Y=\frac{1}{k}\sum_{i=1}^{k} s_i[\sin\phi_i\cos\theta_i]$$ y-directional information

$$Z=\frac{1}{k}\sum_{i=1}^{k} s_i[\sin\theta_i]$$ z-directional information

where s_i are our mono audio signals we want to encode at the according positions Φ_i (horizontal angle (azimuth) phi), and Θ_i (vertical angle (elevation) theta).

AMBISONIC DECODING

Ambisonic encoded signals carry the directional information of an entire sound-field. Thus, an Ambisonic decoder is designed to reproduce signals that could be fed to the loudspeaker.

If we choose our environment where four loudspeakers are placed with 90° separation between the adjacent pairs of speakers and $\phi 1$ = 45°

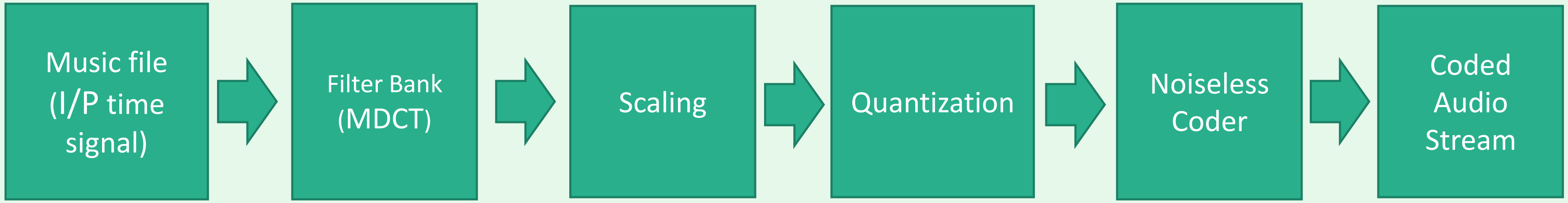
ϕi+1= $\phi 1 + \pi/2$ where 1 ≤ i ≤ 3

$$Speaker\ Weights\ [SW] = \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} & 1/\sqrt{2} & 1/\sqrt{2} \\ \cos\phi 1 & \cos\phi 2 & \cos\phi 3 & \cos\phi 4 \\ \sin\phi 1 & \sin\phi 2 & \sin\phi 3 & \sin\phi 4 \end{bmatrix}$$

$$B\ to\ D\ converter = [SW]\ T \times [W\ X\ Y]$$

AAC ENCODING

- Advanced audio coding (AAC) is a technique used for compressing and encoding audio signals
- It minimizes the amount of data required to store and transmit the signal
- Signal components that are irrelevant are discarded and redundancies in the coded audio signal are wiped out



AAC ENCODING

Modified Discrete Cosine Transform (MDCT)

- The B-format signal is converted from time-domain to frequency-domain using forward modified discrete cosine transform (MDCT).
- Preprocessing of the signal for MDCT:
  The signal is windowed into frames of 2048 samples with 50% overlapping.

$$y[n]=x[n].w[n]$$

- Now, we convert each frame of the signal to the frequency domain using MDCT.

MDCT:

$$X_k = \sum_{n=0}^{2N-1} x_n \cos\left[\frac{\pi}{N}\left(n+\frac{1}{2}+\frac{N}{2}\right)\left(k+\frac{1}{2}\right)\right]$$

$k=0,1,\ldots, N-1$

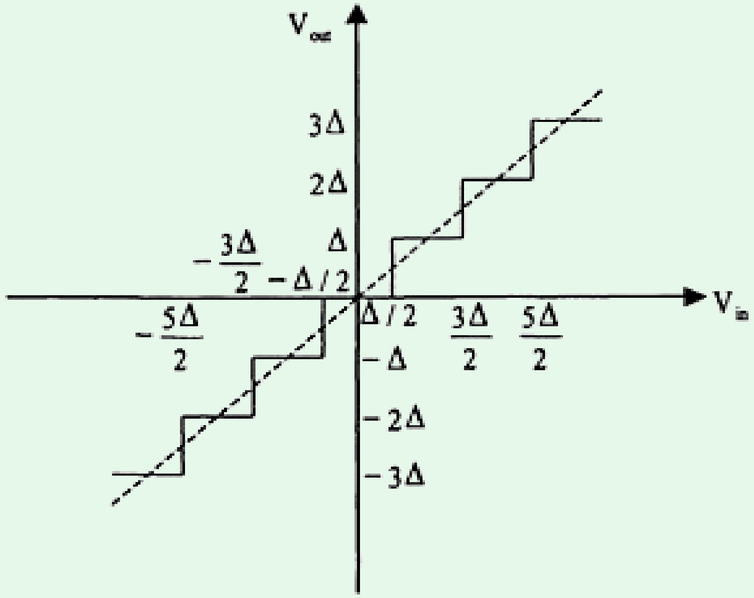
$x_n = h_n a_n$ is the windowed input signal; a_n is the input signal with 2N samples and h_n is the window function.

AAC ENCODING

Quantization

Signal X (n,ω) is in a finite range (fmin , fmax) then it is divided into L equal intervals of length Q (quantization step size)

$$Q = (fmax - fmin)/L$$

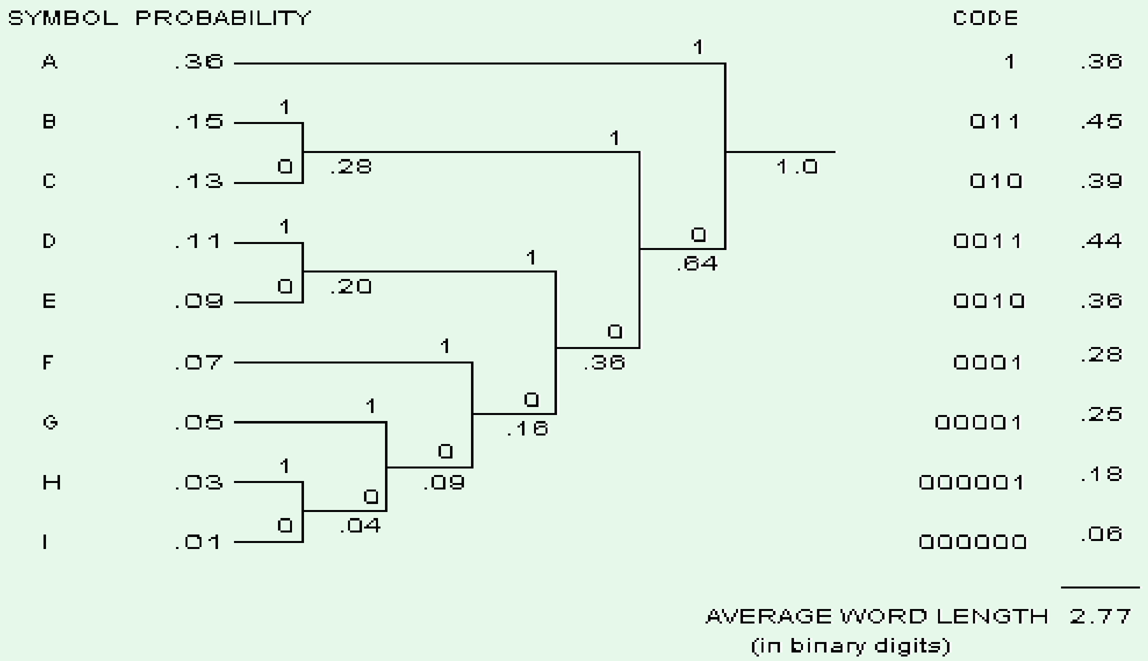


Mid Tread Quantizer

AAC ENCODING

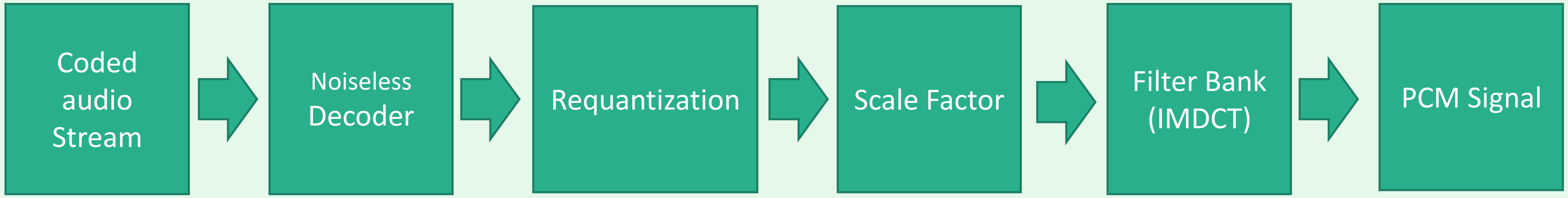
Noiseless coder

- For the Noiseless coder we have used the Huffman coding algorithm.
- Huffman coding is a lossless compression.
- In this method, more common symbols are generally represented using fewer bits than less common symbols.
- We get the quantized signal from quantizer then we apply Huffman coding as stated below:



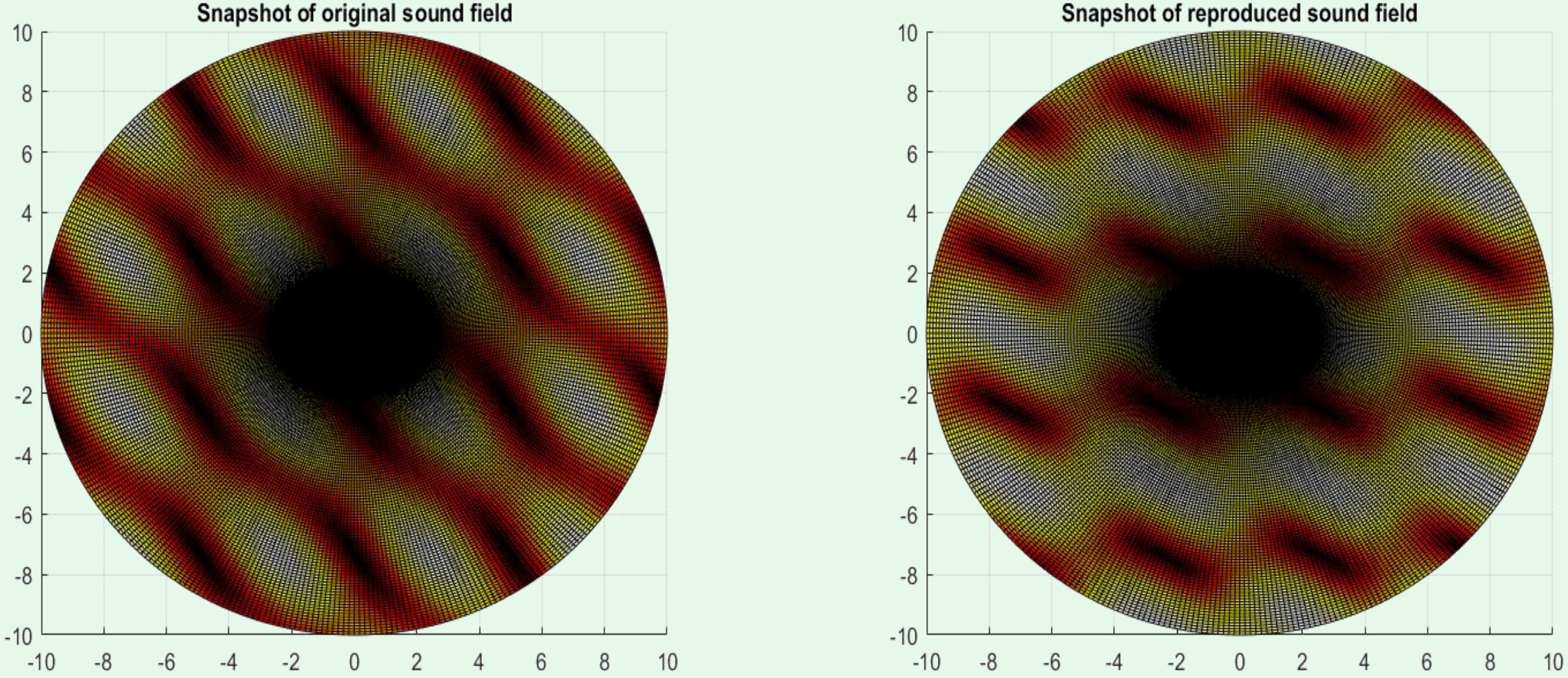
AAC DECODING

Encoded AAC is decoded back to its original form at the reciever's side



WAVE-FIELD ANALYSIS

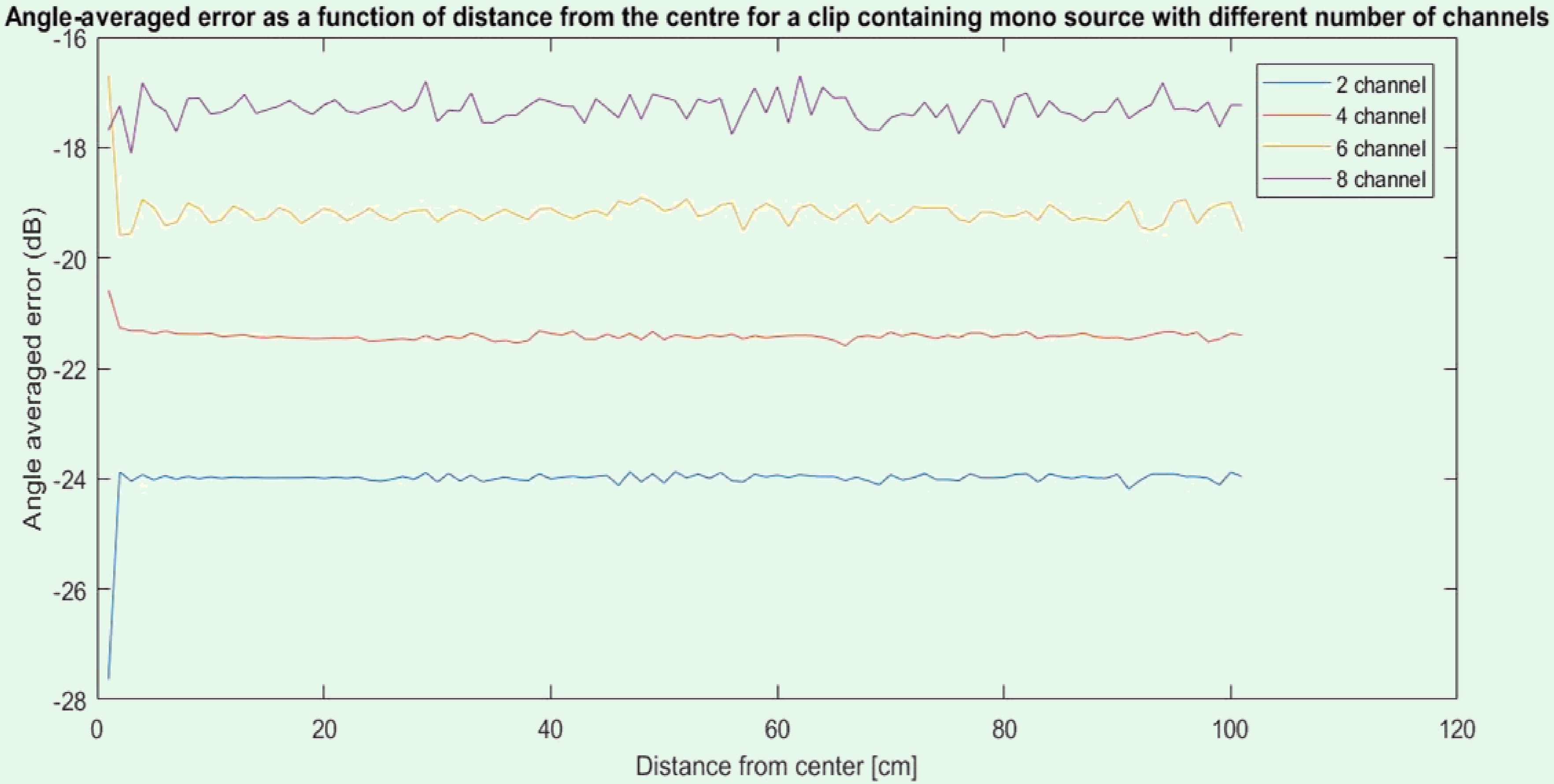
A final wave field snapshot was generated by plotting the instantaneous wave field across the reproduction area (a circle of radius 10 cm).



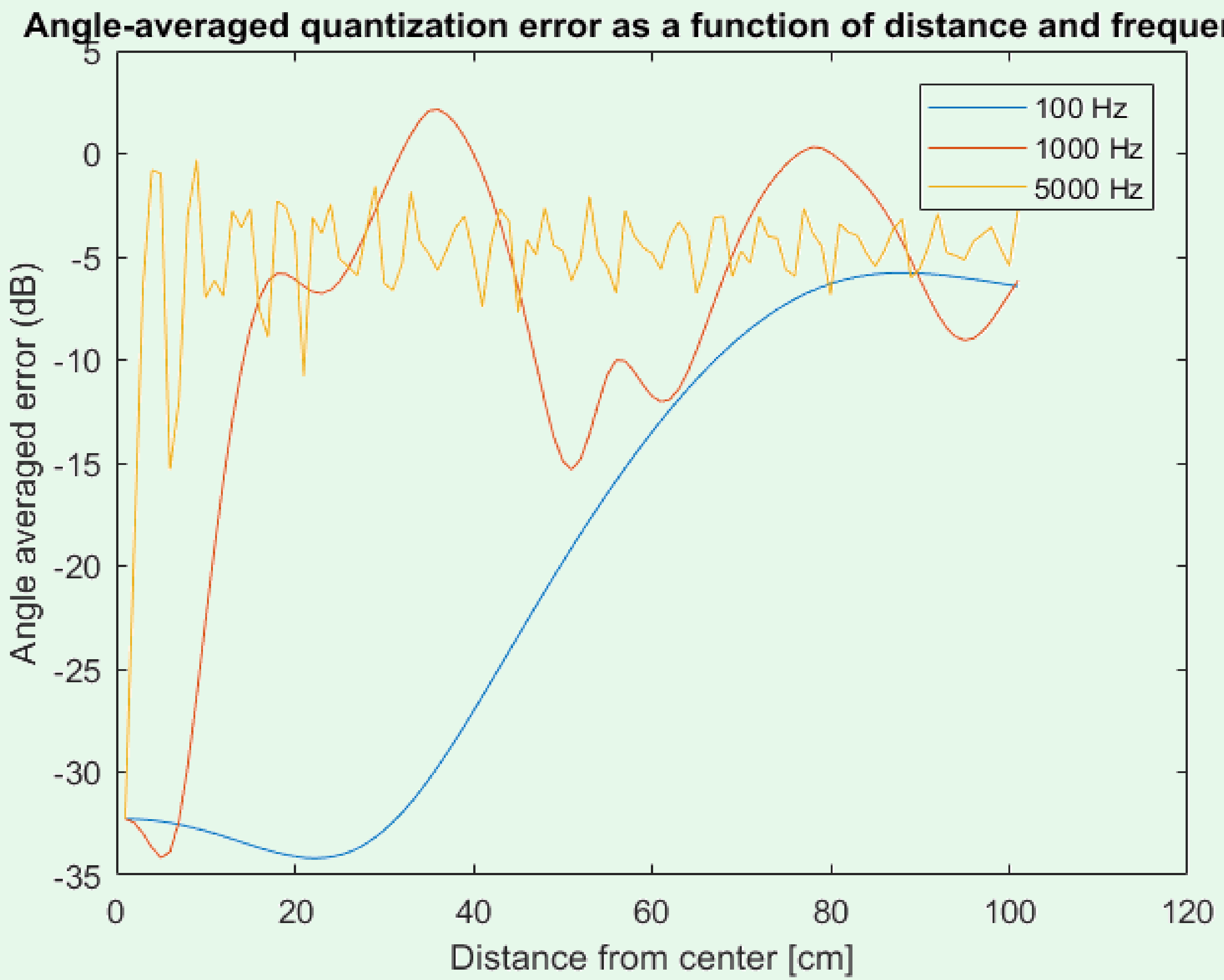
As can be seen from above wavefield snapshot (for 500 Hz), there is a slight change in the reproduced sound field.

RESULTS

Error as function of distance

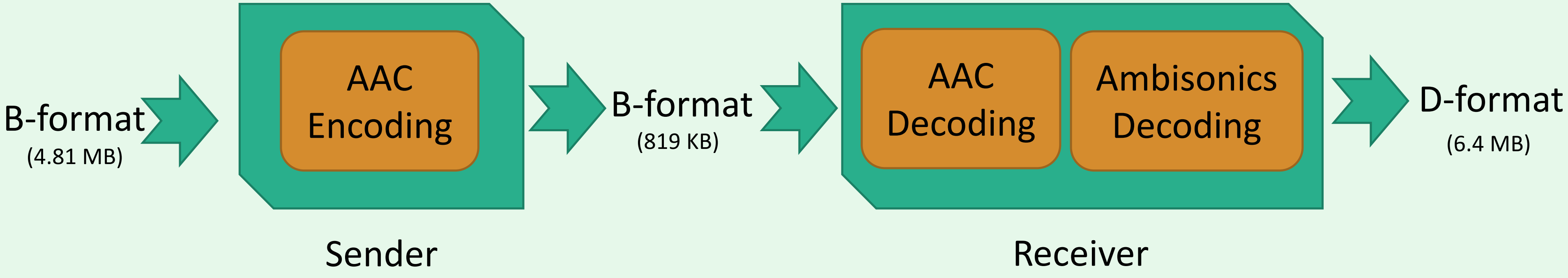


The angle averaged error for different no. of channels is plotted and it is found that more the no. of channels the less is the error.

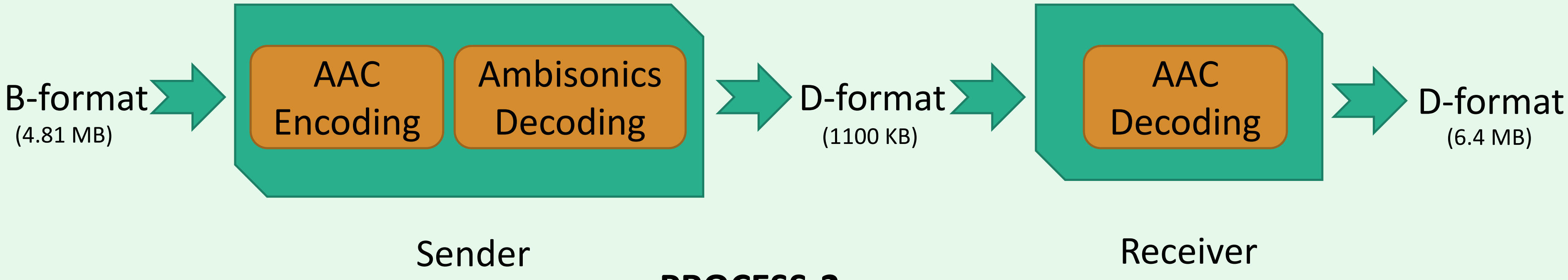


The radial extent of the area with the lowest error depends on the frequency.The higher the frequency is, the smaller is the area with a lower error. The perfect reconstruction radius for a frequency of 1000 Hz is 30 cm.

Comparing two-system model process as follows:



PROCESS-1



PROCESS-2

The size of the file needed during transmission is less in Process-I than Process-II. And as the no. of channels increases, the size of transmitting file may increases further in Process-II. Thus, for transmission Process-I is efficient.

THANK YOU!