Lecture 9: Scale Factor 25 May 2023

The central objective in audio coding is to represent the signal with a minimum number of bits while achieving transparent signal reproduction, i.e., while generating output audio which cannot be distinguished from the original input, even by a sensitive listener ("golden ears"). This paper gives a review of algorithms for transparent coding of high-fidelity audio (Painter and Spanias, 1997).

1. CD-quality digital audio has essentially replaced analog audio.

High fidelity quality audio signal, sampling frequency 44,100 signal per second (48 kH for UK), typical audio 16 bits depth per sample.

2. Describe Human Psychoacoustic Model

Human Audio System(HAS) refers to frequency perceptual model. It is how the human ears treat the audio sound at various frequency. Sound can be treated as an array of light or wave at various frequency.

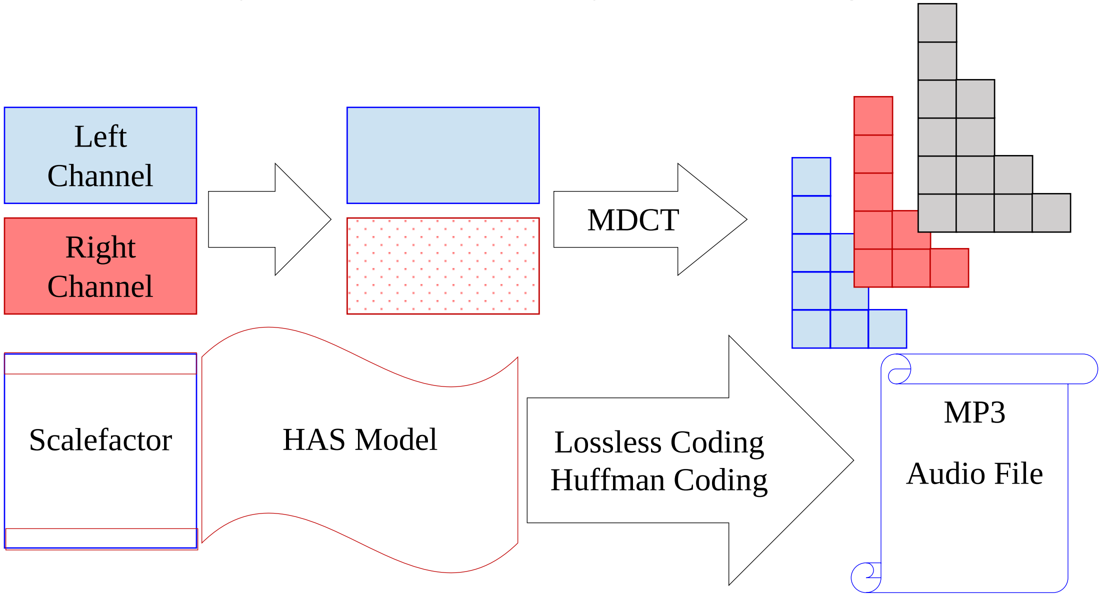
3. Give the frequency range of HAS.

The human ear can nominally hear sounds in the range 20 Hz to 20,000 Hz. The lower the frequency, a sound can travel further.

4. Describe Frequency Masking Concept.

Suppose a listener can normally hear a given acoustical signal under silent condition. When an acoustic signal is playing while stronger sound is being played (a masker), the acoustic signal cannot be heard anymore. The acoustic signal has to be stronger for the listener to hear it.

Describe 5 steps by steps process of Audio Psychoacoustic Coding.



MDCT: Modified Discrete Cosine Transform

We operate under transform domain: Tone sound versus Noise sound.

Give 3 similar process between JPEG Baseline Coding and Audio Psychoacoustic Coding.

*Give 3 different process between JPEG Baseline Coding and Audio Psychoacoustic Coding.

Quantization versus Scalefactor

In quantization process from JPEG Baseline Coding, a quantization table is typically prescribed. Standard luminance and chrominance tables are in the system. Alternative tables may be given in the file header.

However, the Audio Psychoacoustic Coding uses the scalefactor process.

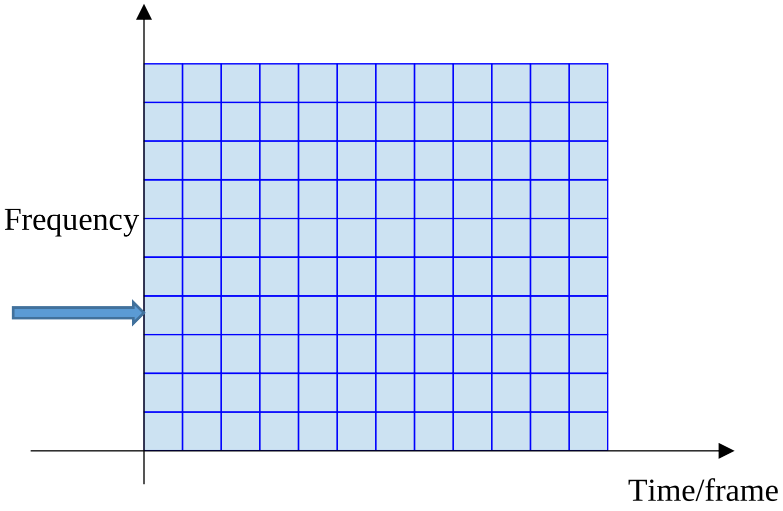
In Perceptual Audio Encoder, let us say, we use 1024(2048) sample frames. We will have 1024 frequency samples out of Fast Fourier Transform (FFT).

First, incoming audio samples, s(n), are normalized according to the FFT length, N, at the same time s(n) will go through MDCT.

Let us encode 32 samples(typically 8, 12, 16 samples) or (6, 12, 18 samples)

STEP 1: Spectral Analysis and Sound Pressure Level(SPL) NORMALIZATION

With 16-bit sample resolution, SPL estimates for very low amplitude input tones are lower bounded by -15 dB SPL.



STEP 2: IDENTIFICATION OF TONAL AND NOISE SIGNALS

Local maxima is a significant signal which will be a candidate of a masker.

Local maxima in the sample PSD which exceed neighboring components within a certain bark distance by at least 7 dB are classified as tonal.

STEP 4 CALCULATION OF INDIVIDUAL MASKING THRESHOLDS

We need to assign the quite threshold level at each frequency critical band.

All the lower signals than the global quite threshold level will be deleted.

For each row at certain frequency, we will take a **scale factor** as the max value of the row.

STEP 5 ENCODE INDIVIDUAL FREQUENCY SIGNALS

Let us encode 32 samples(typically 8, 12, 16 samples) or (6, 12, 18 samples)

Longer block 18 samples is meant for tonal melodical sound.

Shorter block 6 samples will be used to encode a transitional or noise sound.

Take the maximum value of the signal within a block as a scale factor. Take a predetermined bit allocation and encode the signal as a ratio of the scalefactor.

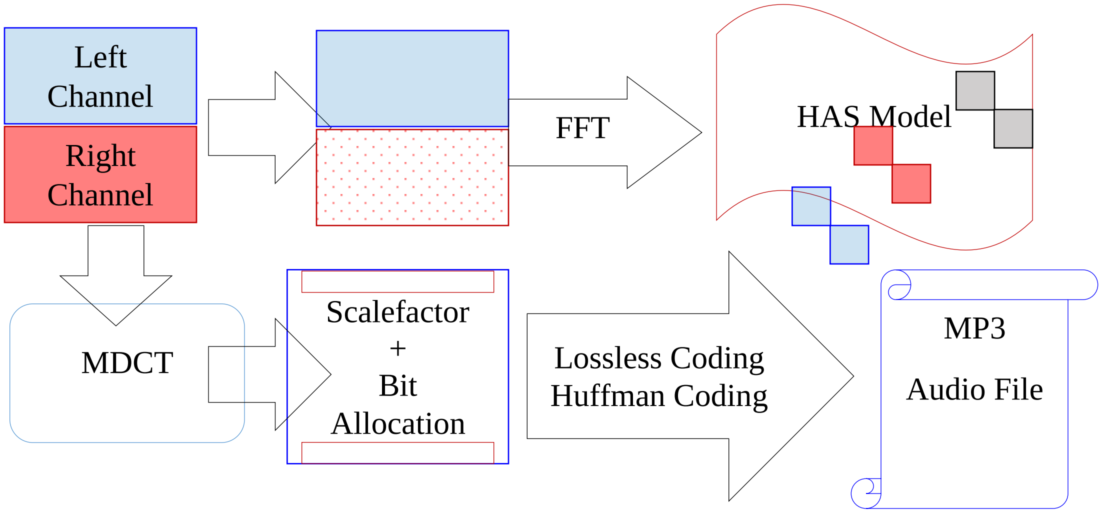
Describe **THREE(3)** technical process in Audio Psychoacoustic Coding which compress an audio mono input.

Give a basic strategy of embedding a watermark or stegano message under Audio Psychoacoustic Coding.

A basic textbook technique is to use LSB in image watermarking. In audio a basic technique is to add an echo.

The study of perceptual entropy (PE) suggests that transparent coding is possible in the neighborhood of 2 bits per sample [45] for most for high-fidelity audio sources (~88 kbps given 44.1 kHz sampling * 16 bits per sample).

Pre-masking: absolute audibility thresholds for masked sounds are artificially increased prior to, during, and following the occurrence of a masking signal. Premasking tends to last only about 5-10 ms. Postmasking will extend further and tends to last only about 50-200 ms.


Left Channel
Right Channel
FFT
HAS Model
MDCT
Scalefactor + Bit Allocation
Lossless Coding Huffman Coding
MP3 Audio File