

Lecture 2 : Sounds / Audio & Music

Basics

- Sound: a pressure wave that moves through a compressible medium.
 - Amplitude: Using voltage levels.
 - Frequency
- Audio: 20Hz - 20kHz
- Speech: the sounds human can utter

Digitize

- Time: Sampling for frequency response
 - Discrete time domain
- Amplitude: Magnitude range
 - Discrete sample values

Sampling Rate

Nyquist Sampling Theorem

Sampling Rate: $2 * \text{maximum frequency response}$

Example:

1. 8kHz for speech (Voice: 300-3400Hz)
2. 44.1kHz for CD-audio (Hearing: 22050Hz)

Determination of Quantization Value

Discrete: bits of data

8 bits for telephone

Quantization * sample rate = required bit per second (Data rate)

E.g. telephone: $8 * 8000 = 64\text{kb/s}$

Encoding

PCM (Pulse Code Modulation)

PCM describes how a digital signal can be formed from a series of pulses.

Uncompressed samples for reference.

Number of channels (tracks)

- Mono: 1 channel
- Stereo: 2 channels
- Professional: up to 32 channels
- Interleaving channel samples

Music

- **Music** representation: a means of specifying the **information** needed to produce a piece of music.
- **Operational** representation: the exact **structures**, **timings** and **sounds** to be produced.
- **Symbolic** representation: the form of the music and allows for interpretation.

Speech

- 600 - 6000 Hz
- Periodic Behavior: during certain time interval.
- Frequency Bands: the spectrum shows characteristic maxima, mostly 3-5 frequency bands.

Speech Analysis

- Speech recognition
 - Speech preprocessing
 - Feature extraction
 - Matching and decision
 - 2 Systems:
 - Verification System (1-to-1 matching)
 - Identification System (1-to-many matching)

Speech Features

- Features:
 - Frequency-Ban: Filter band system -> produce spectrum information for comparison
 - Spectrogram: time-varying spectral representation showing the spectral density
 - Energy distribution (Pitch lines)
 - Visual comparison
 - Formant (共振峰) Frequencies (Using resonances; different **position**)
 - Coarticulation (协同发音)
 - Some traces of the **old** sound will retain.
 - Pitch Contours (**fundamental frequency**)
 - Linear Prediction

Performance Evaluation

1. Intra-class variability

- i. Train more voices of an individual.

2. Inter-class similarity

- i. Remove the common features and remain the unique.