Pitch Perception

Notes on Speech and Audio Processing

Chia-Ping Chen

Department of Computer Science and Engineering

National Sun Yat-Sen University

Kaohsiung, Taiwan ROC

Introduction

- Human pitch perception is performed by the entire auditory system, of which our knowledge is still fragmentary.

- The modeling of pitch perception is primarily based on psychoacoustic experiments, at times also supported by physiological explorations.

- Explanations can be controversial.

Theory of Pure Tone Perception

- How is a pure tone perceived?

- The basilar membrane responds to different frequencies with peaks at different locations. So different hair cells respond to different frequencies.

- A pure tone would cause the greatest vibration at a specific place on the BM. This ultimately leads to the perception of this tone: the brain knows which fiber(s) is excited and is able to perceive the tone.

Pitch Perception

- While the previous theory (due to Helmholtz) explains the perception of pure tones, it cannot explain the following pitch perception.

- A pitch can be perceived when the energy at the corresponding frequency is completely absent.
    - Experiment by Seebeck: The pitch at frequency $\frac{1}{T}$ can be perceived even the spectrum has no energy there.
    - A modern version: The perceived pitch remains as 200 Hz even when the pure tones at that frequency is removed.

Periodicity Model

- A periodicity model is shown in Figure 16.9. The BM is modeled as a filter bank and the hair-cell auditory-nerve complex is modeled as elementary pitch detectors (EPD).

- Each EPD fires neuron independently. A global unit accumulates the intervals between spikes. The pitch period is determined by choosing the peak of histogram.

Place Model

- The underlying assumption of the place model is the ability of the auditory system to resolve harmonic peaks of the stimulus.

- There are two stages in this model:
  - Stage 1 performs statistical separation of the frequency spacing between spectral peaks.
  - Stage 2 computes the correlation of the spectrum of lags proposed by stage 1. The winning candidate with maximum correlation is the pitch.
  - This is shown in Figure 16.10.