Pitch Detection

Notes on Speech and Audio Processing

Chia-Ping Chen

Department of Computer Science and Engineering
National Sun Yat-Sen University
Kaohsiung, Taiwan ROC

Introduction

- Speech and music production can be seen as excitations driving a time-varying filter. The pitch is the frequency of such excitations.
- The excitation signal is quite complicated. There are different things to model.
  - buzz (voicing sounds): periodic opening of glottis and the shape of glottis pressure pulse.
  - hiss (noise-like sounds): turbulent flow is created at some point of constriction.
  - how voicing and turbulence combine (voiced fricatives)
  - the stop consonants

Pitch Perception vs. Detection

- The modeling of the excitation signals is more complex than the modeling of the linear filters.
- Pitch is the frequency of a pure tone that is matched to a more complex signal. This is a subjective definition.
- A pitch detector decides the fundamental frequency of an incoming signal. This is an objective definition. Pitch detection and fundamental frequency estimation are used interchangeably.

Simple Methods

- Based on the idea that the fundamental frequency component must exist at some level.
- Use a slope filter, as shown in Figure 30.1. The amplitude at fundamental frequency is amplified more than other harmonics.
- Or use the peaks in the signal to estimate pitch period, as shown in Figure 30.3.
- However, the fundamental frequency component may be missing due to band-limited communication channels. Furthermore, noises may be present to mask the signals.

Difficulty in Pitch Detection

- Large dynamic range: $60 - 800$ Hz
- Pitch variation in time
- Vocal tract variation in time
- Voiced-unvoiced transition
- Telephone speech (convolutional noise)
- Background noises

Signal Processing

- Low-pass filtering: compare Figure 30.5 and 30.6.
- Spectral correlation: The Fourier transform of harmonics of equal amplitude and zero phase is a pulse train.
- Inverse filtering: The inverse of the vocal tract filter can be used to restore the excitation from the signal.
- Comb filtering: A grid search over all possible pitch periods by subtracting delayed signals.
- Cepstral pitch detection: The high-time peaks correspond to pitch periods.
- High resolution spectrum.

Pattern Recognition Methods

- Combines multiple sources of information, beneficial even when there are dependencies among them. For example, one uses
  - a low-pass filter to smooth speech wave,
  - a processor extracting function values from signal, each leading to a pitch period estimator,
  - overall combination of the results via statistical or other criteria.
- The pitch period estimation is often post-processed to fix errors. A common method is the median smoothing, as shown in Figure 30.16.