Monophonic Fundamental Frequency Estimation

Bertrand SCHERRER

McGill University

March 8, 2007

Outline

- Historically linked to speech processing (Bell Labs)
- Very active field of research : hundreds of methods.

$\Rightarrow$ Will not present **ALL** the methods.

- Historically linked to speech processing (Bell Labs)
- Very active field of research : hundreds of methods.

$\Rightarrow$ Will not present **ALL** the methods.

- Present the **model-level typology** proposed by Anssi Klapuri in his PhD Thesis.
  - Spectral Location Algorithms
  - Spectral Interval Algorithms
  - Auditory Model Algorithms

- Present the **model-level typology** proposed by Anssi Klapuri in his PhD Thesis.
  - Spectral Location Algorithms
  - Spectral Interval Algorithms
  - Auditory Model Algorithms

- Present the **model-level typology** proposed by Anssi Klapuri in his PhD Thesis.
  - Spectral Location Algorithms
  - Spectral Interval Algorithms
  - Auditory Model Algorithms

Autocorrelation Function

The **auto-correlation function** (ACF):

$$r(\tau) = \lim_{N\to\infty} \frac{1}{2.N + 1} \sum_{n=-N}^{N} x(n).x(n+\tau) \tag{1}$$

$\Rightarrow$ maximum when τ corresponds to the fundamental period.

Autocorrelation Function

The **auto-correlation function** (ACF):

$$r(\tau) = \lim_{N \to \infty} \frac{1}{2.N + 1} \sum_{n=-N}^{N} x(n).x(n + \tau) \tag{1}$$

$\Rightarrow$ maximum when τ corresponds to the fundamental period.

Autocorrelation Function

The **auto-correlation function** (ACF) - revisited:

$$r(\tau) = FFT(|FFT(x[n])|^2) \tag{2}$$

Cepstrum

The **cepstrum-based** method:

$$r(\tau) = FFT(\log(|FFT(x[n])|)) \tag{3}$$

Cepstrum

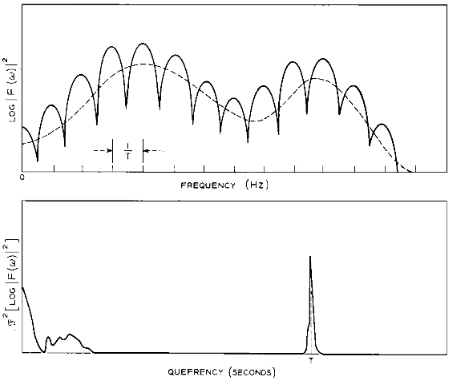

Fig.3 [Noll 67]

- A continuum of techniques in between
- Emphasize frequency partials at the harmonic locations in the spectrum

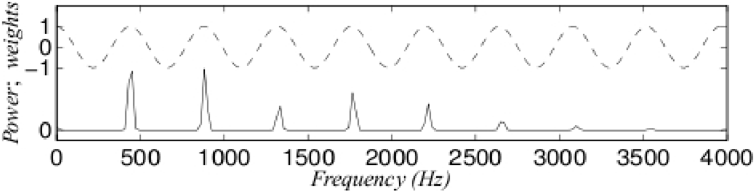

Fig.10 [Klapuri 04]

Filter-bank Approach

In [Brown 92]:

- Logarithmically spaced spectral components $\Rightarrow$ constant spacing between partials of a harmonic sound regardless of F0.
- Fundamental estimation $\Leftarrow$ correlating the filterbank output with spectral patterns

Filter-bank Approach

In [Brown 92]:

- Logarithmically spaced spectral components ⇒ constant spacing between partials of a harmonic sound regardless of F0.
- Fundamental estimation ⇐ correlating the filterbank output with spectral patterns

Filter-bank Approach

In [Brown 92]:

- Logarithmically spaced spectral components $\Rightarrow$ constant spacing between partials of a harmonic sound regardless of F0.
- Fundamental estimation $\Leftarrow$ correlating the filterbank output with spectral patterns

MLE Approach

In [Doval91,93]: maximizing the likelihood of a fundamental frequency candidate given the observation of the sound partials

Two-way Mismatch Method

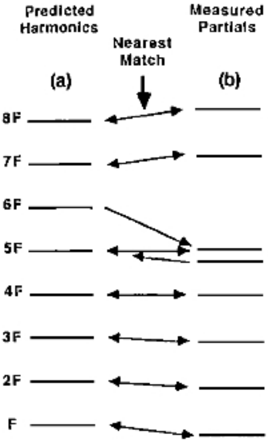

Fig.1, Maher94

Two-Way Mismatch Procedure

- **Step 1**: average of the frequency differences between each observed partial and its nearest neighbour among the predicted harmonic frequencies.

Two-way Mismatch Method

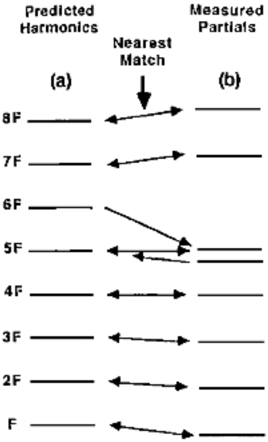

Fig.1 Maher94

Two-Way Mismatch Procedure

- **Step 2**: average the frequency differences between each predicted harmonic frequency and its nearest neighbour among the observed partials.

Spectrum Autocorrelation Method

$$\tilde{r}(m) = \frac{2}{K} \sum_{k=0}^{\frac{K}{2}-m-1} |X(k)||X(k+m)| \tag{4}$$

Basic Observation

- Any signal with more than one frequency component exhibits periodic fluctuations, beating, in its time-domain amplitude envelope.
- The fundamental period can then clearly be identified as the duration between the two highest beatings.

Basic Observation

- Any signal with more than one frequency component exhibits periodic fluctuations, beating, in its time-domain amplitude envelope.
- The fundamental period can then clearly be identified as the duration between the two highest beatings.

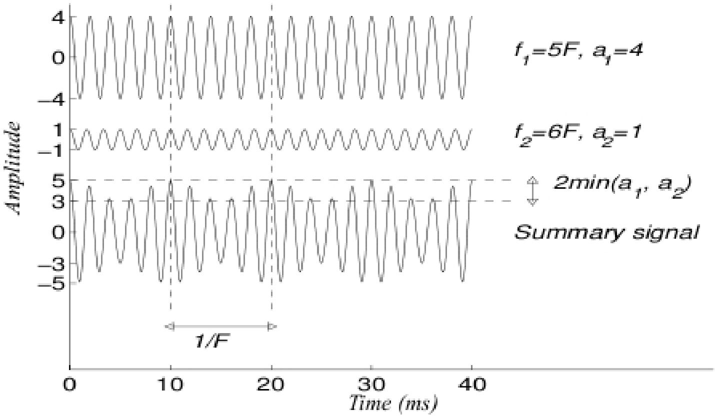

Fig.12 [Klapuri04]

Interest

Compromise between *spectral location* and *spectral interval* methods.

Effect of Half-Wave Rectification

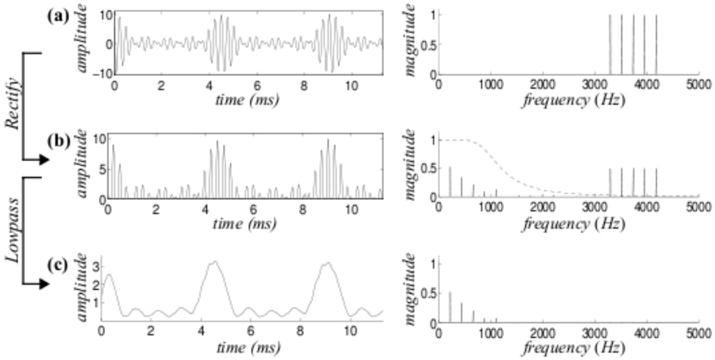

Fig.11 [Klapuri04]

The Unitary Model

In [Meddis91]:

1. signal is passed through a bank of bandpass filters $\Rightarrow$ frequency selectivity of inner ear.
2. each channel is compressed, half-wave rectified and low pass filtered $\Rightarrow$ behaviour of the signal in the auditory nerve.
3. periodicity estimation in each channel using short-time ACF.
4. ACFs summed over all the channels, at each frame t :

$$s_t(\tau) = \sum_c r_c(\tau) \tag{5}$$

The value of the lag corresponding to the highest value of $s_t(\tau)$ is taken to be the fundamentla frequency on frame t.

The Unitary Model

In [Meddis91]:

1. signal is passed through a bank of bandpass filters $\Rightarrow$ frequency selectivity of inner ear.
2. each channel is compressed, half-wave rectified and low pass filtered $\Rightarrow$ behaviour of the signal in the auditory nerve.
3. periodicity estimation in each channel using short-time ACF.
4. ACFs summed over all the channels, at each frame t :

$$s_t(\tau) = \sum_c r_c(\tau) \tag{5}$$

The value of the lag corresponding to the highest value of $s_t(\tau)$ is taken to be the fundamentla frequency on frame t.

- First two steps are widely accepted as general by the research community
- 3rd and 4th steps are still subject to debate and research
- Global processing chain very successful in reproducing phenomena in human hearing $\Rightarrow$ still a reference today.

- First two steps are widely accepted as general by the research community
- 3rd and 4th steps are still subject to debate and research
- Global processing chain very successful in reproducing phenomena in human hearing $\Rightarrow$ still a reference today.

- Choice of F0 estimation technique $\Leftarrow$ prior knowledge on the signal (if any):
  - For example: Harmonicity of the signal $\Rightarrow$ Spectral Location vs Spectral Interval

Harmonic Signal

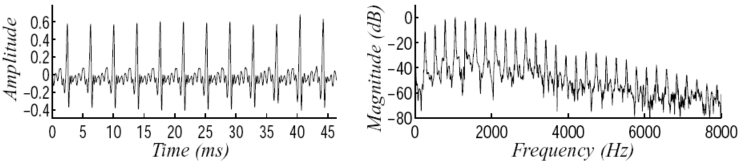

Fig.6 [Klapuri04]

Inharmonic Signal

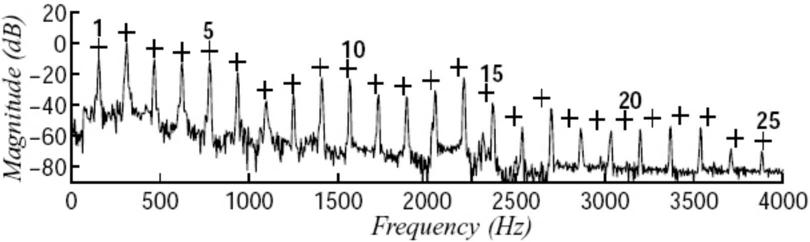

Fig.7 [Klapuri04]

Non Harmonic Signal

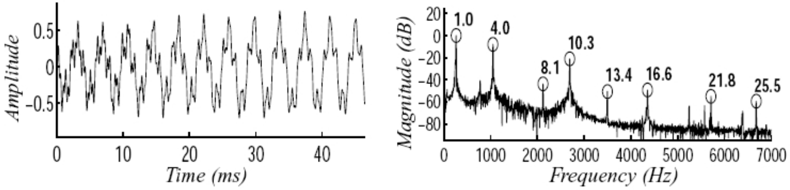

Fig.9 [Klapuri04]

- Use of the instantaneous energy computed from the Teager Energy Function [Abu-Shikhah99]
- ...

Questions ?