Fundamental frequency estimation

ISMIR Graduate School, October 4th-9th, 2004

Contents:

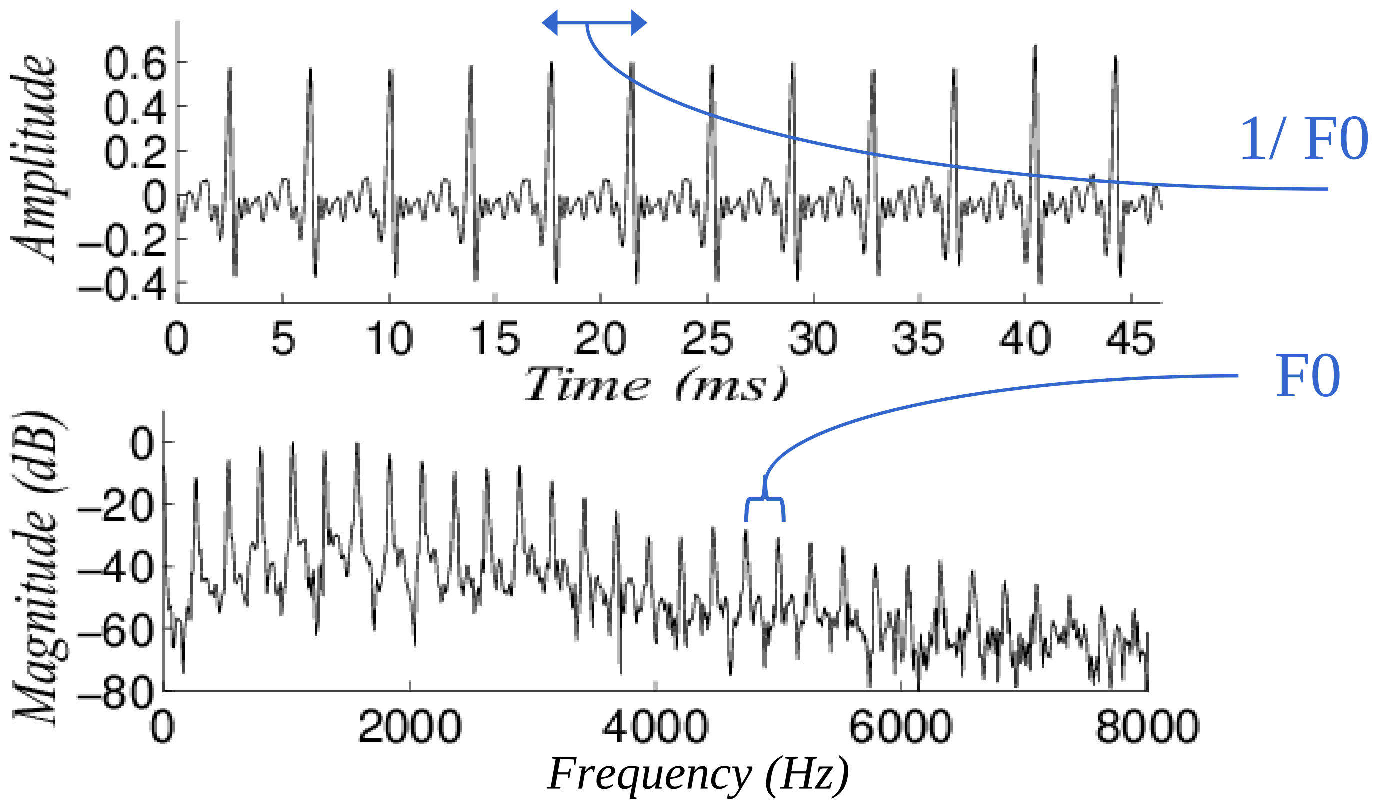
- Musical sounds
- F0 esimation methods
  - Time-domain periodicity estimation
  - Spectral pattern matching
  - Frequency domain periodicity
  - Auditorily motivated methods
- Case study: singing transcription system

1 Introduction

- *Pitch* is a perceptual attribute of sounds, defined as the frequency of a sine wave that is matched to the target sound in a psychophysical experiment
- *Fundamental frequency (F0)* is the corresponding physical term
  - defined only for periodic or nearly periodic sounds
  - F0 is the inverse of the period
  - in ambiguous situations, the period corresponding to perceived pitch is chosen
- Usually pitch ≈ F0, both are measured in Hertz units

2 Musical sounds

- Most Western instruments produce *harmonic sounds*
  - Figure: trumpet sound (260Hz) in time and frequency domains
  - period in time-domain: 1/F0   period in frequency-domain: F0



Musical sounds

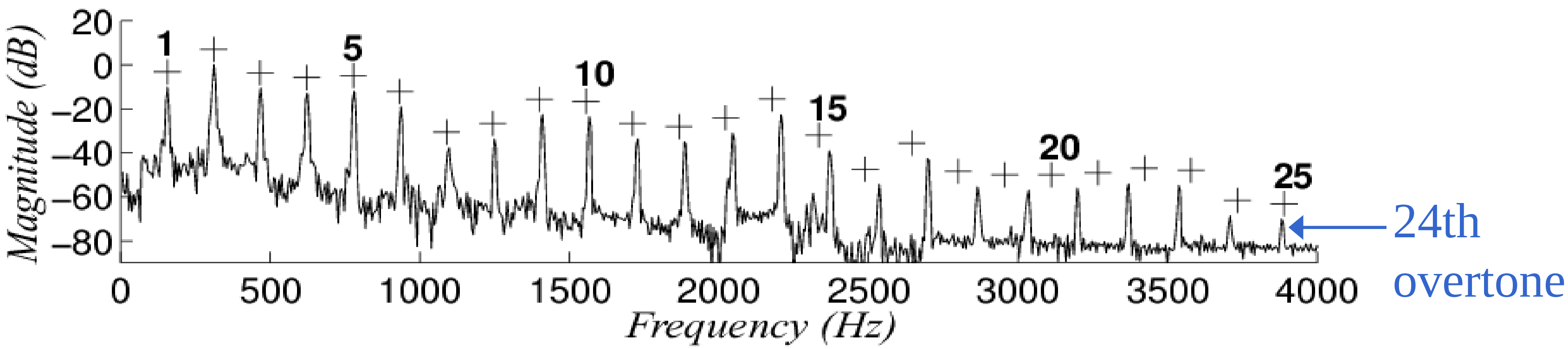
- Western musical instruments which do or do not produce harmonic sounds

Table 1: Western musical instruments which do or do not produce harmonic sounds.

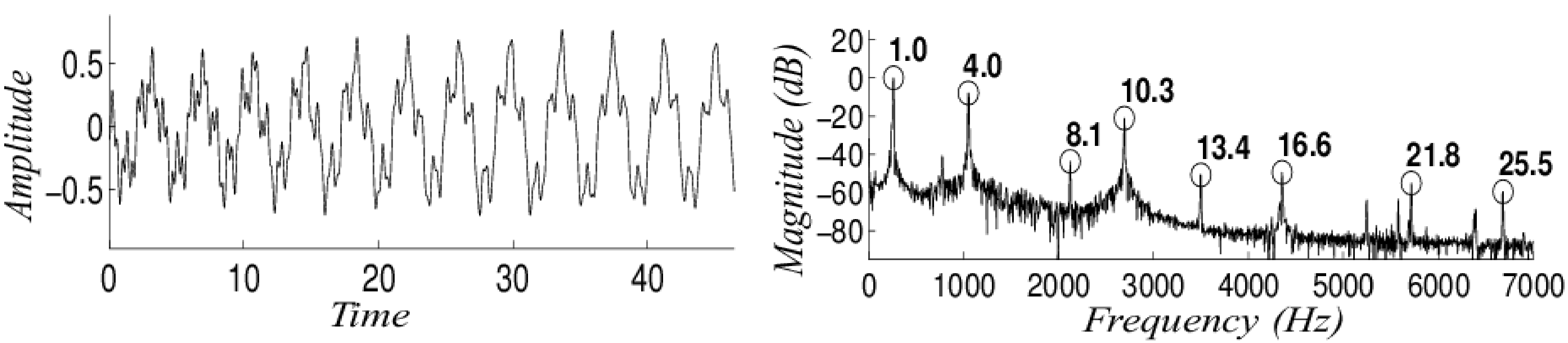
Produced sounds	Instrument family	Instruments involved
Harmonic	String instruments	Piano, guitars, bowed strings (violin etc.)
	Reed instruments	Clarinets, saxophones, oboe, bassoon
	Brass instruments	Trumpet, trombone, tuba, english/french horn
	Flutes	Flute, bass flute, piccolo, organ
	Pipe organs	Flue pipes and reed pipes
	Human voice (singing)	Voiced phonemes
Not harmonic	Mallet percussions	Marimba, xylophone, vibraphone, glockenspiel
	Drums	Kettle drums, tom-toms, snare drums, cymbals

Musical sounds

- *Imperfect harmonicity*: spectrum of a piano sound (156Hz)
  - ideal multiples of the F0 are indicated by ”+” marks and numbered



- Vibraphone sound(260Hz): *nearly periodic, clearly pitched, but not harmonic*



3 Taxonomy of F0 estimation methods

Categorization of F0 estimation methods:

1. Time-domain periodicity estimation
2. Spectral pattern matching
   - look for frequency partials at harmonic spectral locations
3. Frequency-domain periodicity
   - observe spectral intervals (frequency intervals) between partials
4. Auditorily motivated methods
   - periodicity of the time-domain amplitude envelope at subbands

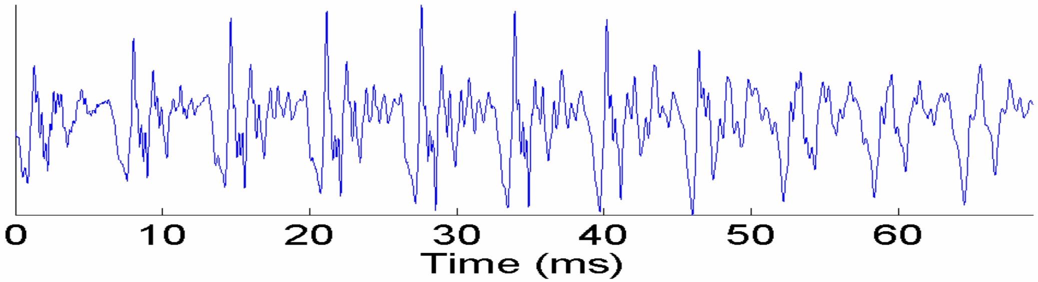
- Processing domain (time/frequency) is not as important as the *information* that is utilized in the calculations

3.1 Time-domain periodicity

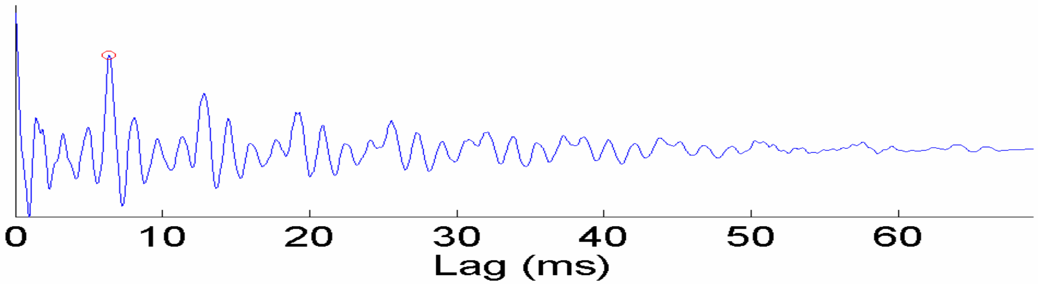
- Autocorrelation function (ACF) based algorithms
  - Among the most frequently used F0 estimators. Usually the maximum value in ACF is taken as 1/F0 period
  - Short-time ACF $r(\tau)$ for a discrete time domain signal $x(n)$:

$$r(\tau) = \frac{1}{N} \sum_{n=0}^{N-n-1} x(n) x(n+\tau)$$

Signal $x(n)$:
(vowel [ae])



ACF:



Time-domain periodicity

- An example of a recent and quite nice ACF-oriented F0 estimator:
  - A. de Cheveigné and H. Kawahara:
    “YIN, a fundamental frequency estimator for speech and music,”
    J. Acoust. Soc. Am., 111 (4), April 2002.
  - a new way of elegantly normalizing ACF + some other ideas

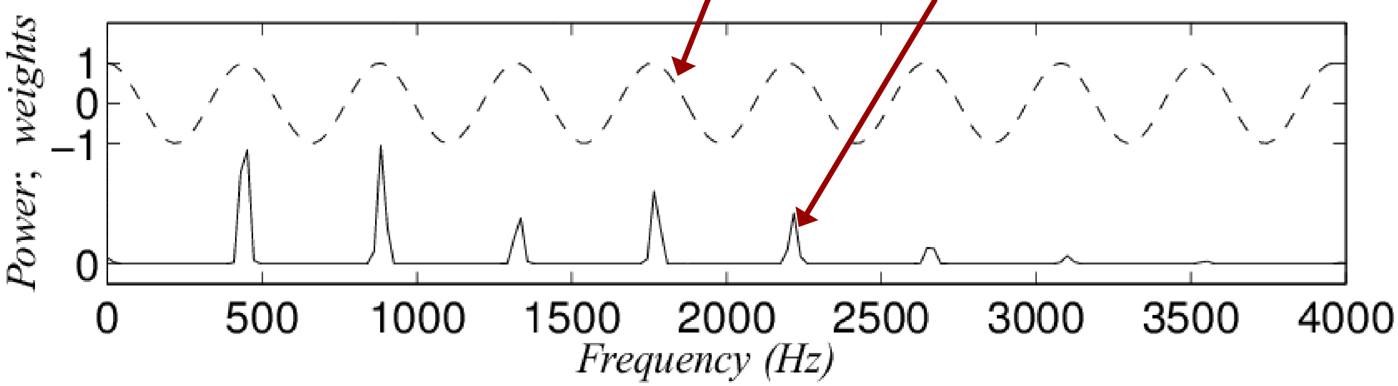
Time-domain periodicity

- Time-domain autocorrelation function (ACF) via the DFT

$$r(\tau) = \mathrm{IDFT}\left\{ \left| \mathrm{DFT}\left(x(n)\right) \right|^2 \right\}$$

- Short-time ACF *emphasizes partials at harmonic locations of the magnitude spectrum*

$$r(\tau) = \frac{1}{K} \sum_{k=0}^{K-1} \left[\cos\left(\frac{2\pi\tau k}{K} \right) \left| X(k) \right|^2 \right]$$



Periodicity in time domain

- Cepstrum pitch detection is closely analogous to ACF

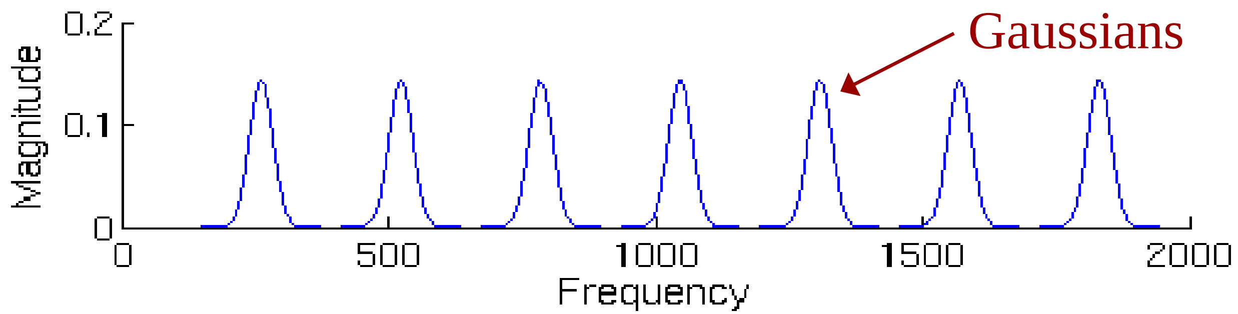
$$c(\tau) = \mathrm{IDFT}\{\log|\mathrm{DFT}(x(n))|\}$$

→ Simply replace $(\ )^2$ in ACF with log( ) in cepstrum

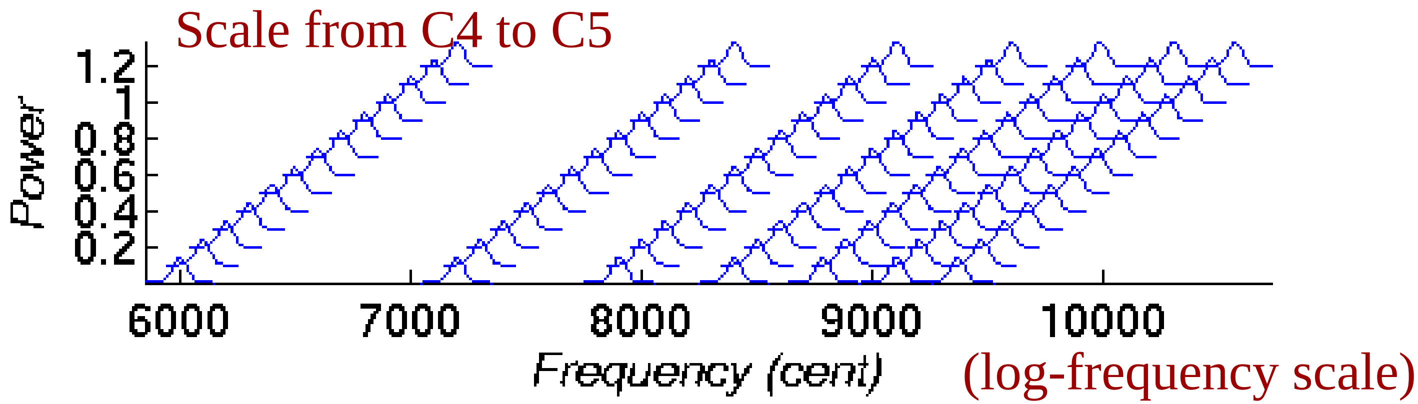
- Difference is quantitative
  - log( ) gives dynamic compression to spectrum
    - \+ flattens the spectra of exotic sounds → robustness for formants etc.
    - – rises the noise level
  - $(\ )^2$ emphasizes spectral peaks in relation to noise
    - \+ noise robustness
    - – further strengthens spectral pecularities of exotic sounds

3.2 Pattern matching in frequency domain

- Examples:
  - Doval & Rodet, ICASSP 1991:
    Maximum likelihood spectral pattern matching

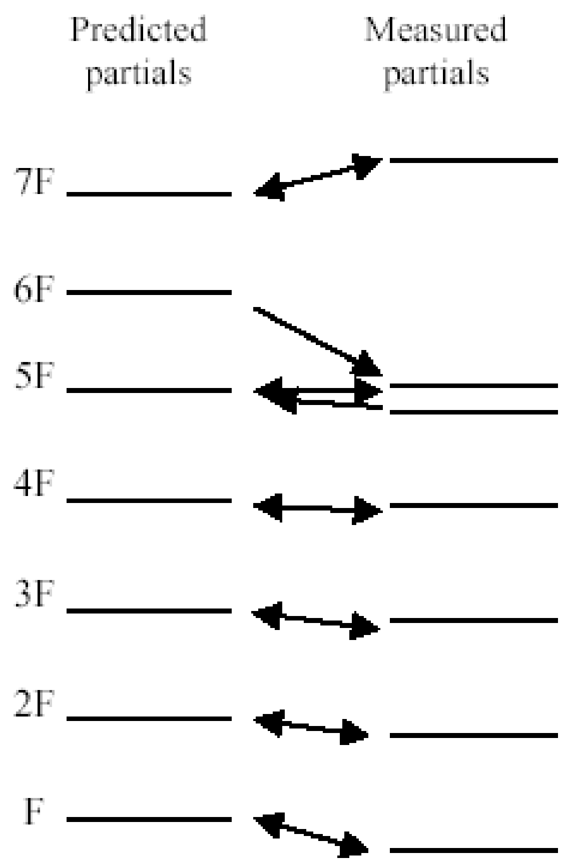


  - Brown, J. C., JASA, 1992:
    Constant-Q transform + correlation with an ideal harmonic pattern



Pattern matching in frequency domain

– Maher, R. C., Beauchamp, J. W., JASA, 1994:
  Two-way mismatch between the observed and predicted partials



Pattern matching in frequency domain

- Both ACF and harmonic pattern matching methods look for partials at *harmonic locations of the spectrum*
- Drawback of these algorithms:
  - perfect harmonicity cannot be assumed
  - higher partials of many real musical instruments are not exactly at harmonic spectral positions
- Let's call these ”*spectral-location type F0 estimators*”

3.3 *Spectral-interval* type F0 estimators

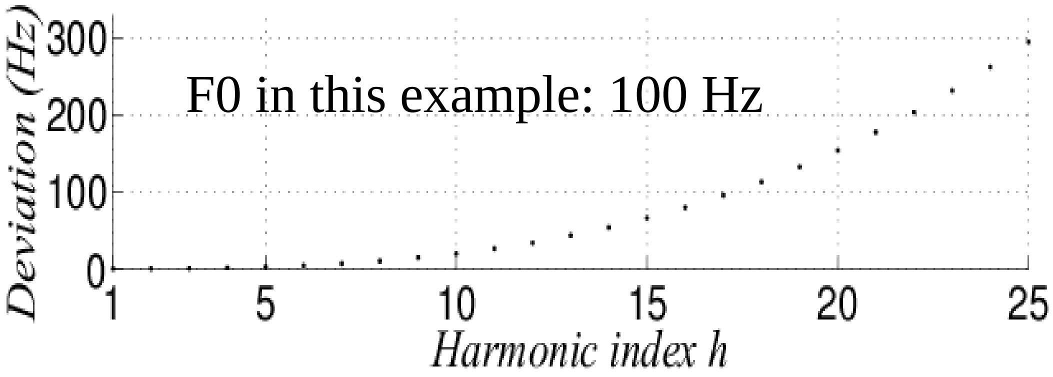
- Idea: harmonics sounds have a *periodic magnitude spectrum* → period is the F0
- Spectrum autocorrelation method in its simplest form:

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

→ Any two spectral components with a *frequency interval* m support the corresponding F0 candidate mf_s/K

  – spectrum can be arbitrarily shifted without affecting the output

- Works better for sounds that exhibit inharmonicity
  – intervals do not remain constant but are more stable than locations
  – **Figure**: deviation of the partial frequencies from the ideal (piano string)

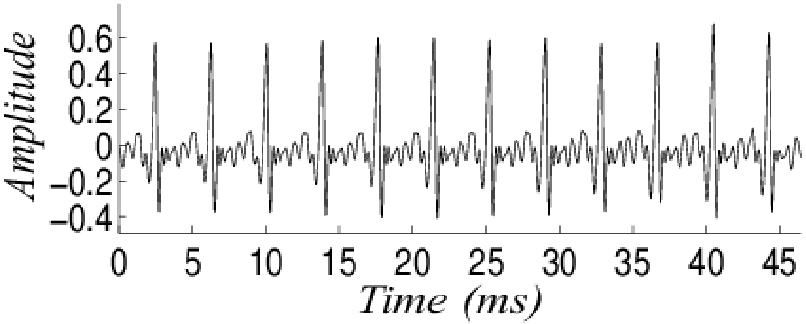


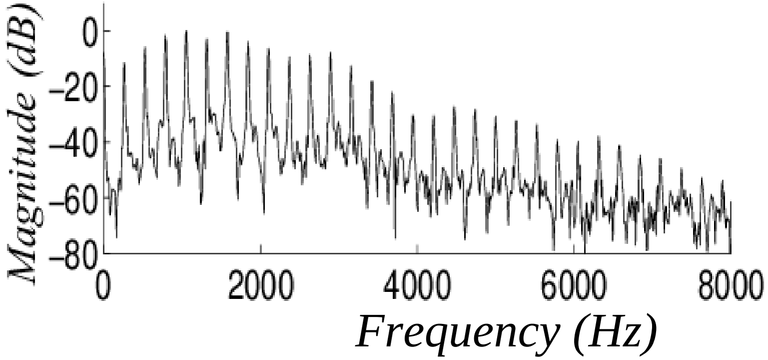
Spectral-location vs. spectral-interval methods

- There are interesting differences between the spectral-location and spectral-interval based approaches

- Period estimates in the time / frequency domains complement each other
  - time domain periodicity: prone to errors in F0 halving
  - frequency domain periodicity: prone to errors in F0 doubling

  → *two complementary types of information*



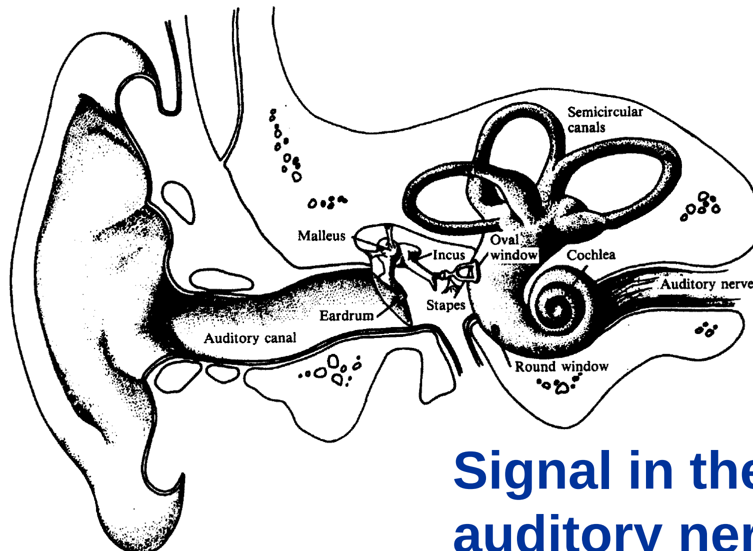


3.4 Auditorily motivated F0 estimators

Physiological parts of hearing:

1. Cochlear frequency analysis
   - typically modeled as a bank of overlapping, linear, bandpass filters

   → ”auditory channels”
2. Hair cells
   - transform mechanical movement into neural impulses
   - typically modeled as cascade of
     a. half-wave rectification
     b. compression
     c. lowpass filtering



Signal in the auditory nerve

Brain

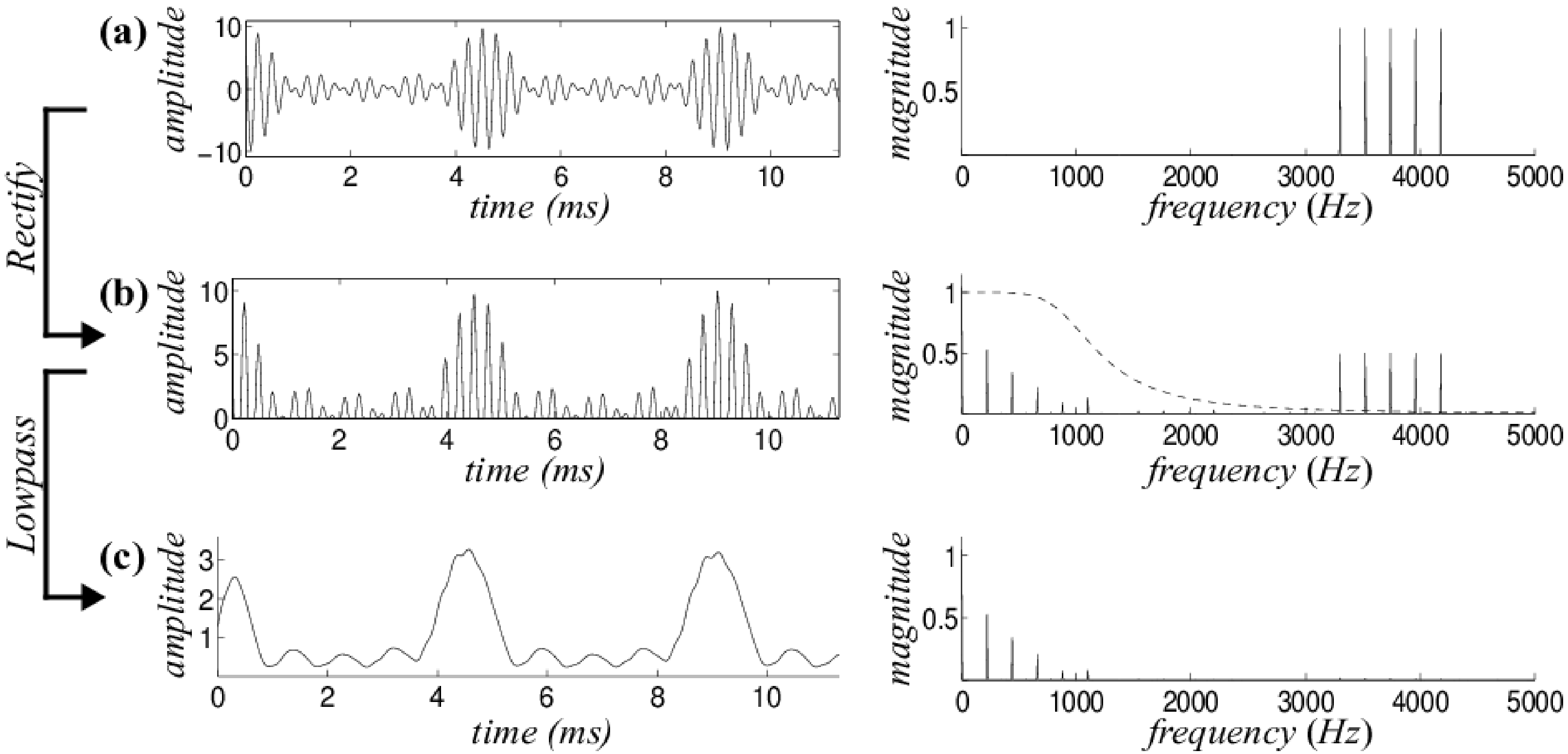
In brain

3. Periodicity analysis within channels
4. Combinarion across channels
   - inter-channel phase differences do not affect

not directly observable
→ more controversial

Periodicity of the time-domain amplitude envelope

- *Idea*: any signal with more than one frequency component exhibits periodic fluctuation, *beating*, in its time-domain amplitude envelope
  - rate of the beating depends on the frequency difference between partials
  - for a harmonic sound, the frequency interval F0 dominates

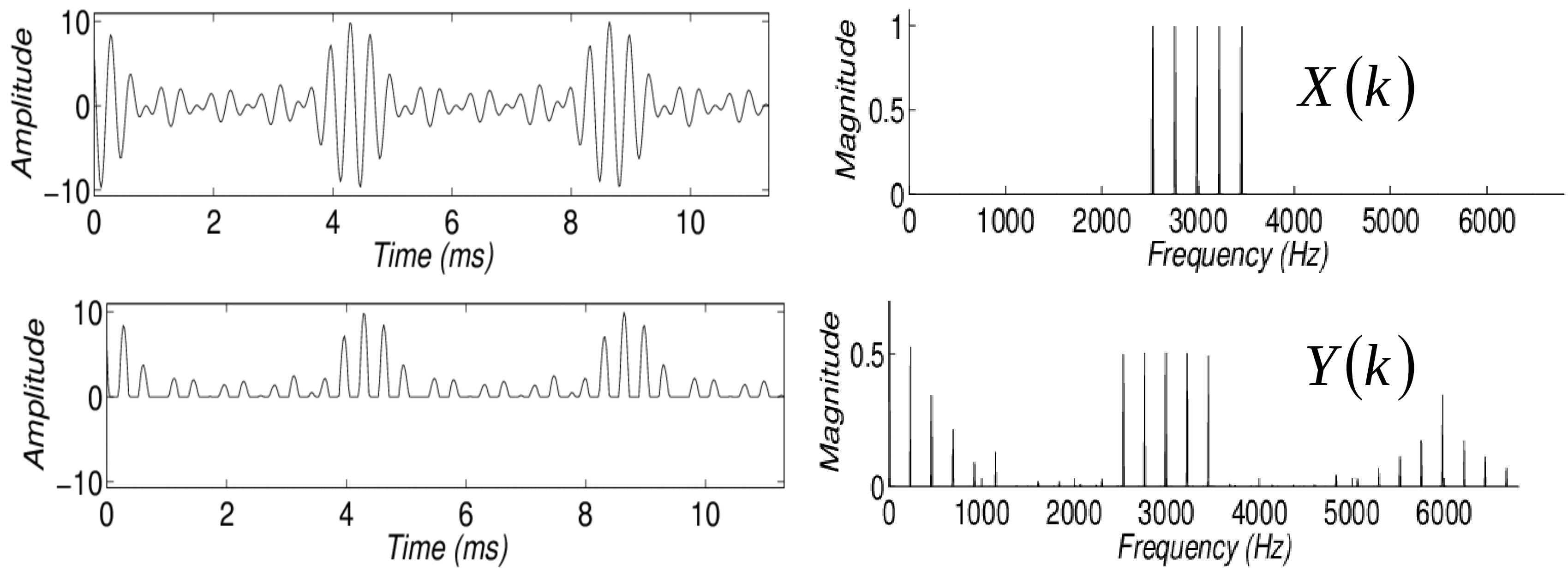


Periodicity of the time-domain amplitude envelope

- When there are several harmonic partials at a subband, the beating at the F0 rate is clearly visible in the amplitude envelope
- Allows F0 estimation at distinct frequency bands, combining results at the end
  - gives flexibility against e.g. band-limited noise

Properties of half-wave rectification

- Half-wave rectification at the output of a bandpass filter
  → *synthesis of spectral-location and spectral-interval information!*

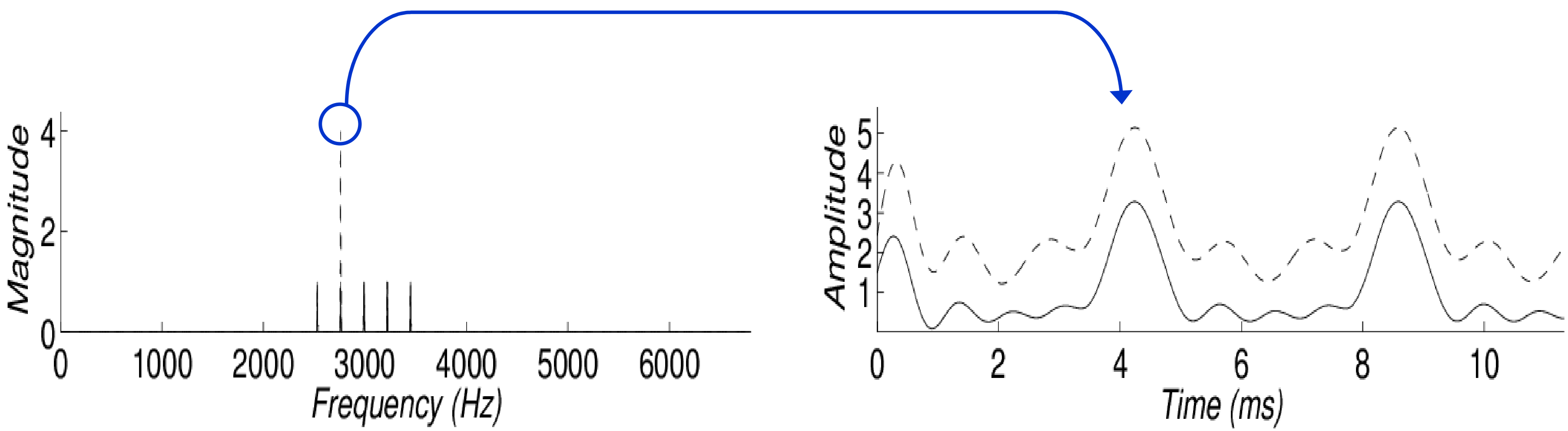


$$\underbrace{\hat{Y}(k)=\frac{\sigma_x}{\sqrt{8\pi}}\delta(k)}_{\text{dc level}}+\underbrace{\frac{1}{2}X(k)}_{\text{input spectrum}}+\underbrace{\frac{1}{\sigma_x\sqrt{8\pi}}\sum_{j=-K/2+k}^{K/2-k}X(j)X(k-j)}_{\text{autoconvolution of the spectrum}}$$

Properties of half-wave rectification

- Spectral "smoothing"
  - flat spectra lead to strong beating
  - a single higher-amplitude partial stands out as independent sound
  - periodicity analysis is not affected by individual interfering partials

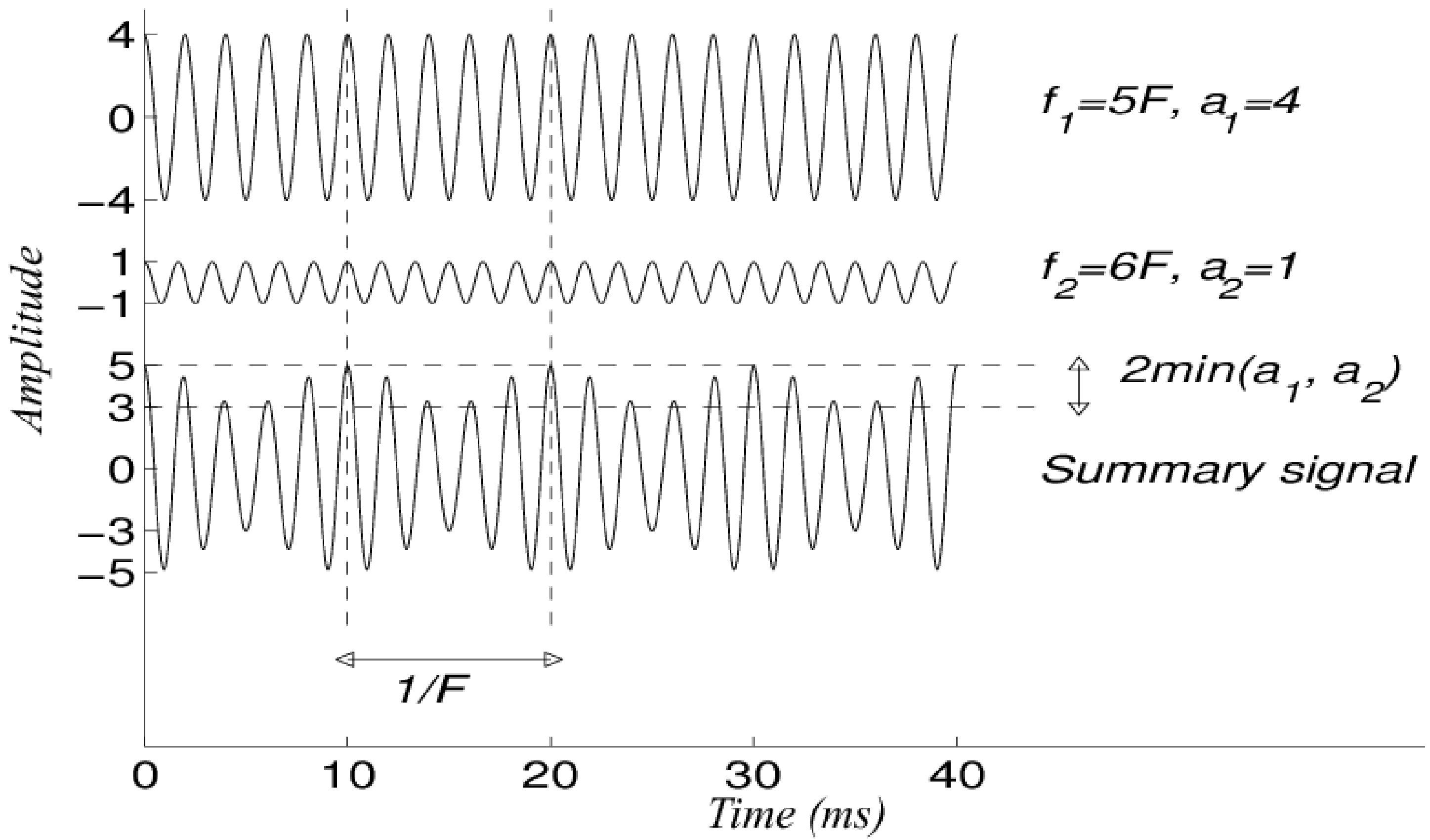
affects only the dc-level of the amplitude envelope



Magnitude spectrum *Corresponding amplitude envelope*

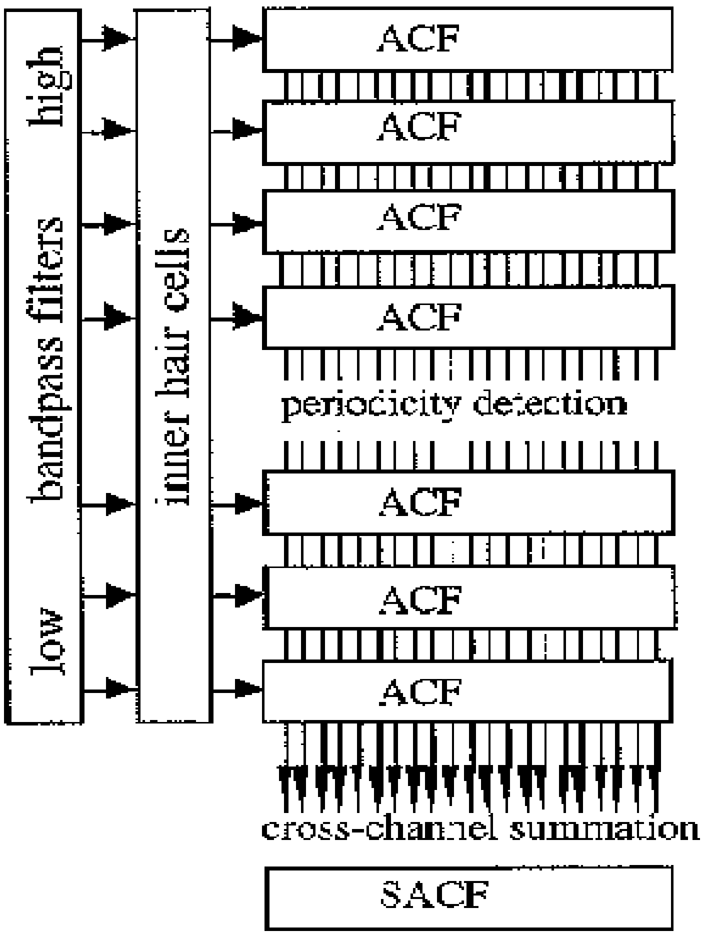
Periodicity of the time-domain amplitude envelope

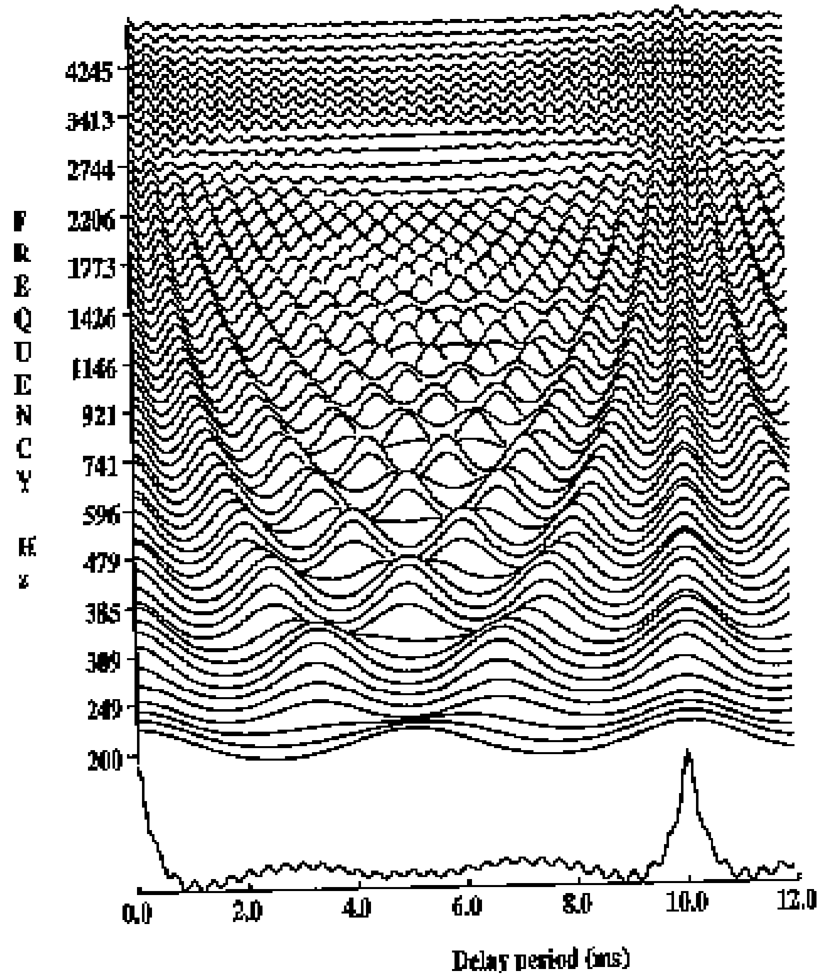
- The magnitude of the beating caused by two sinusoidal partials is determined by the smaller of the two amplitudes



”Unitary model” of pitch perception

- Left: algorithm,
- right: bandwise ACFs, summary ACF



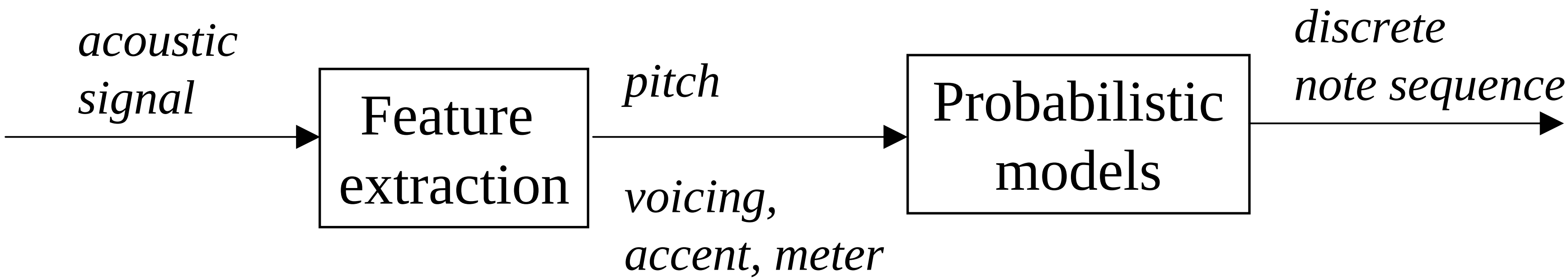


”Unitary model” of pitch perception

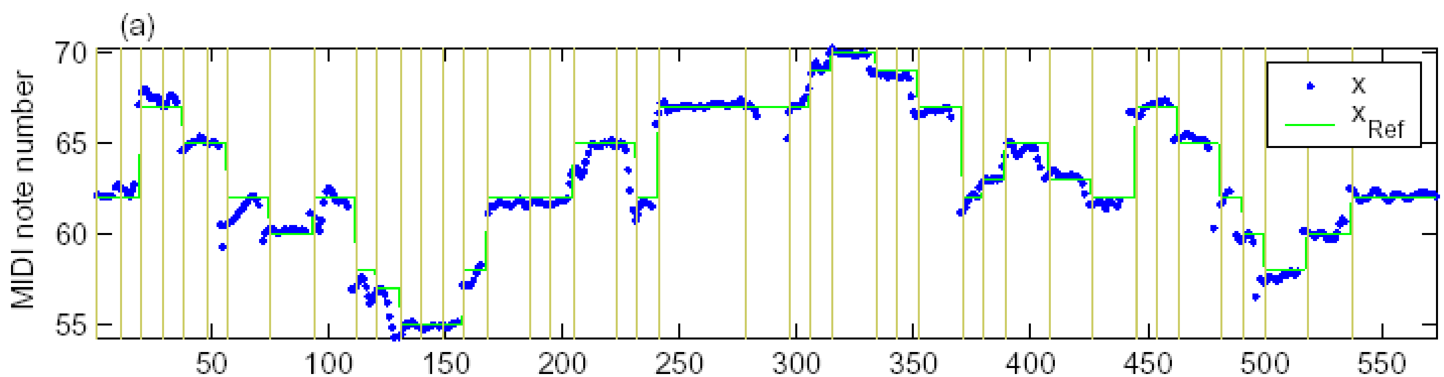
- Note: it is a *model of perception*, not primarily designed for practical F0 estimation
  → modifications are needed to improve accuracy and computational efficiency
- A computationally efficient implementation of the unitary model can be found in
  – Tolonen, Karjalainen,
  ”A computationally efficient multipitch analysis model,”
  IEEE Trans. Speech and Audio Processing, 8(6), Nov. 2000.
  – Uses only two subbands instead of 40-120 bands in the original model

4 Case study: Singing transcription system

- Ryynänen, Klapuri, ”Modeling of note events for singing transcription,” *SAPA Workshop*, 2004.

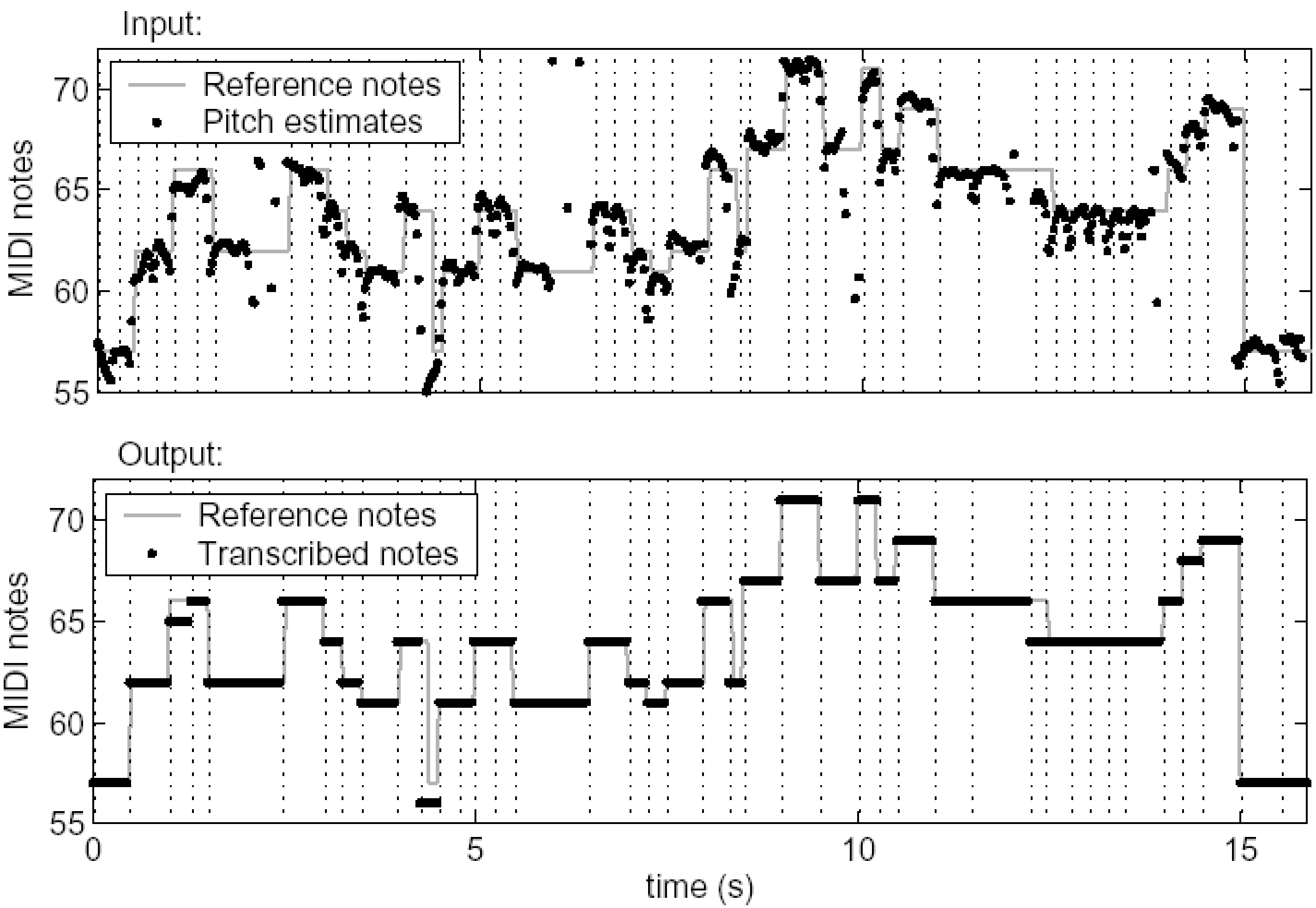


- Estimated pitch track has to be post-processed to get notes

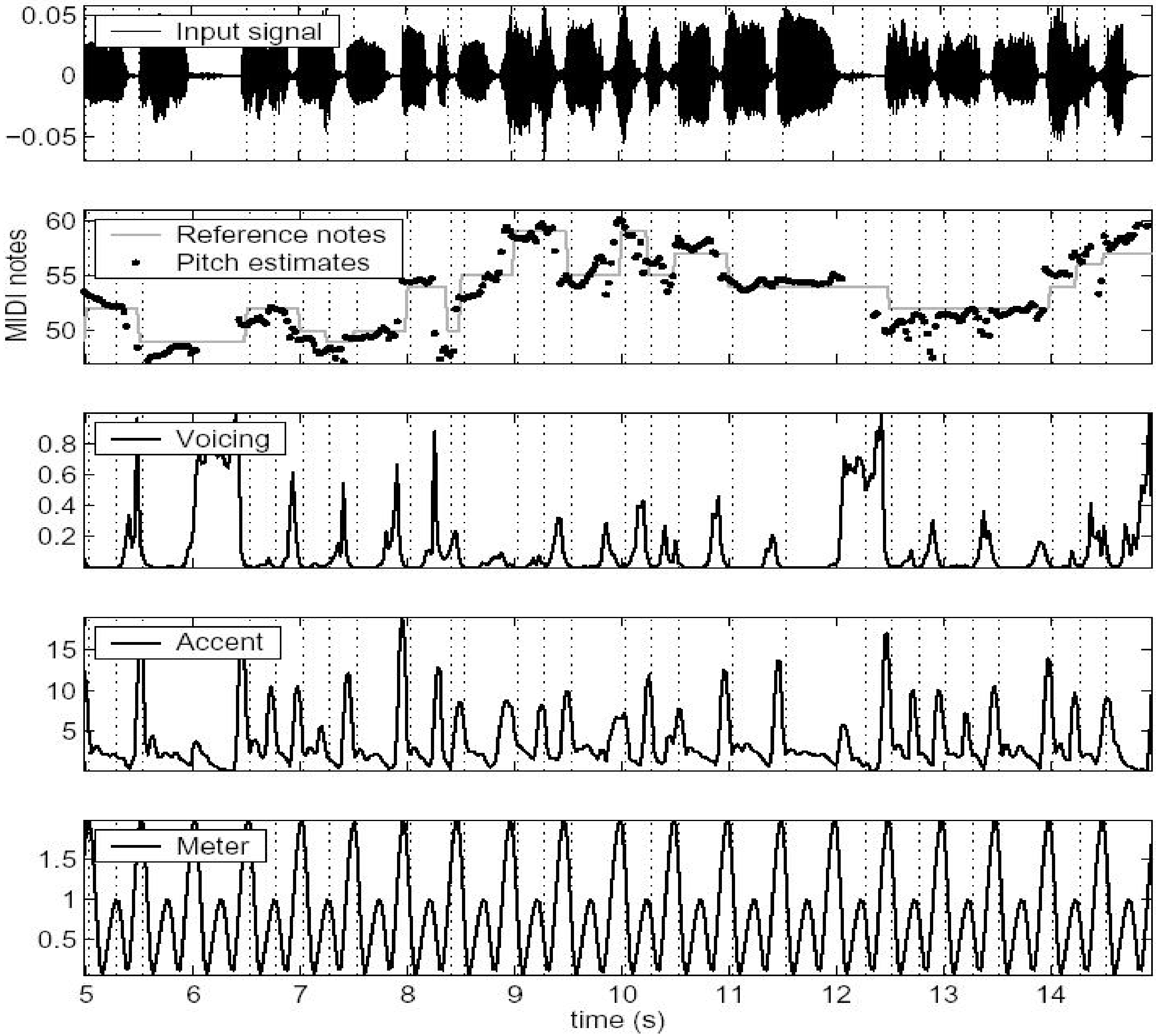


Singing transcription system

- 1: Pitch estimation
- 2: Note segmentation and labeling



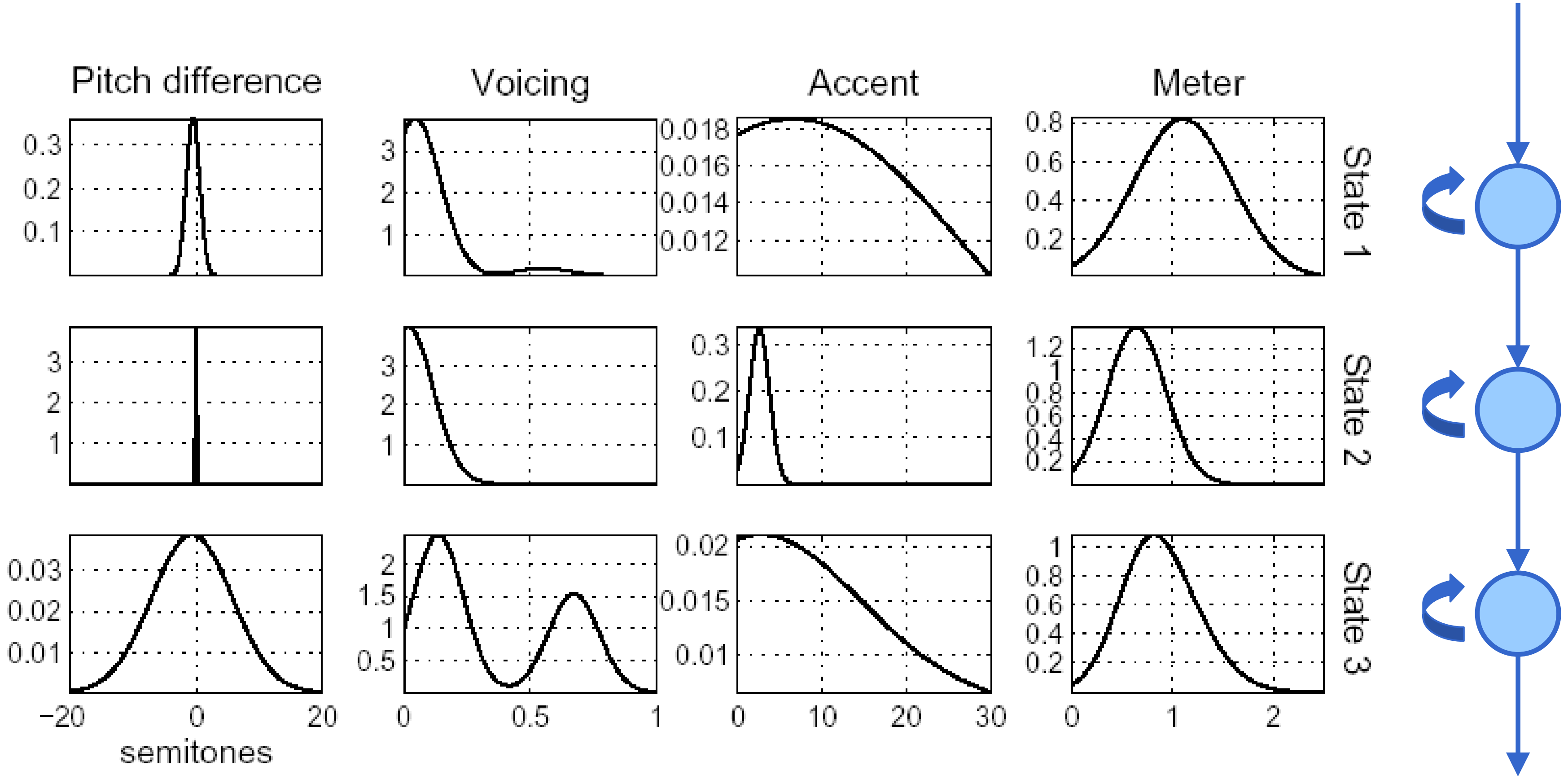
Features


Input signal
0.05
0
−0.05
Reference notes
Pitch estimates
MIDI notes
60
55
50
Voicing
0.8
0.6
0.4
0.2
Accent
15
10
5
Meter
1.5
1
0.5
5
6
7
8
9
10
11
12
13
14
time (s)

Singing transcription system

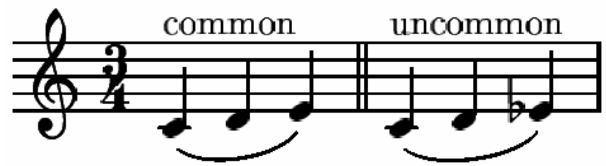
Note model HMM – feature distributions

- Note model performs temporal segmentation and pitch labelling (allows more deviation in beginning and end)

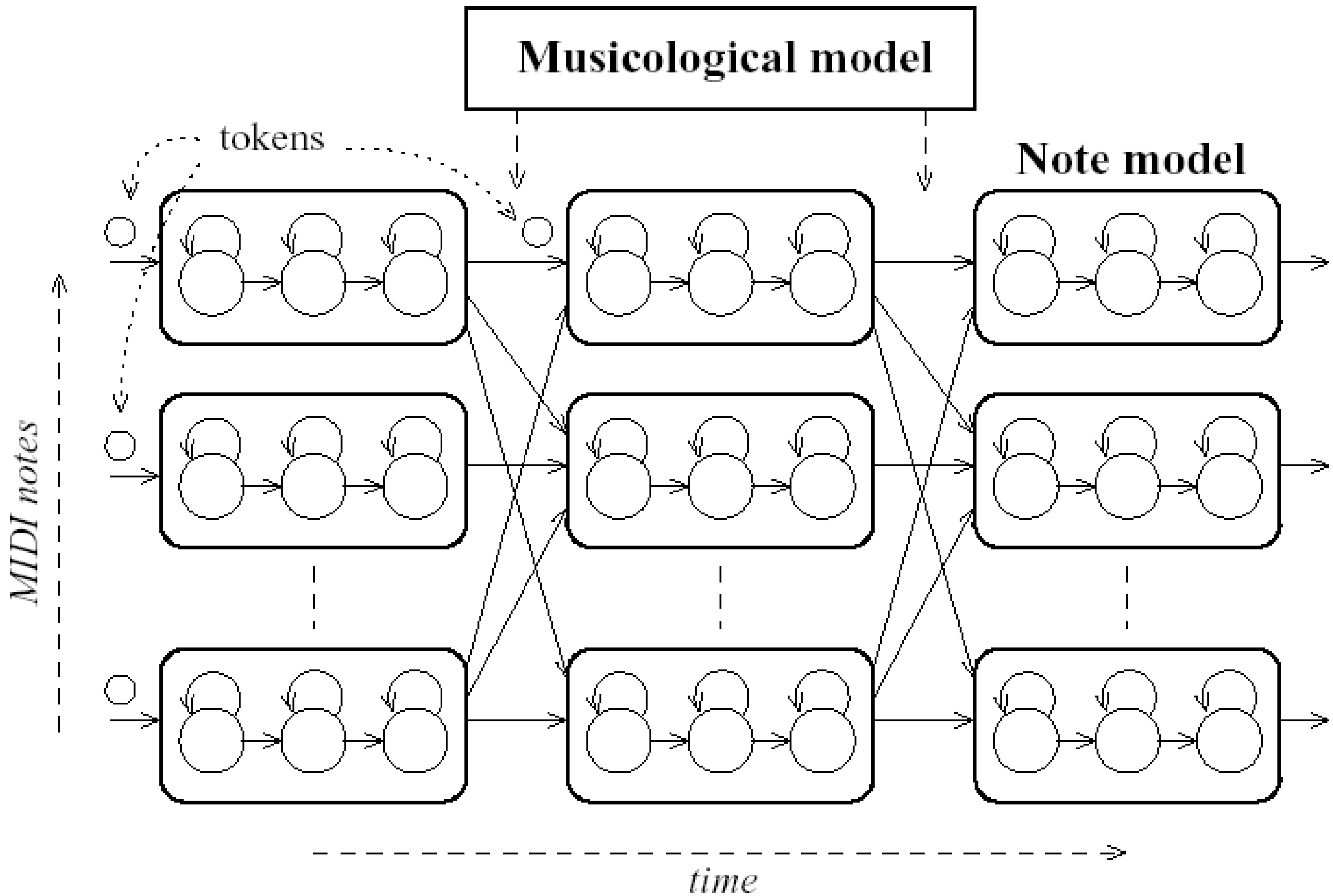


Singing transcription system

Musicological model

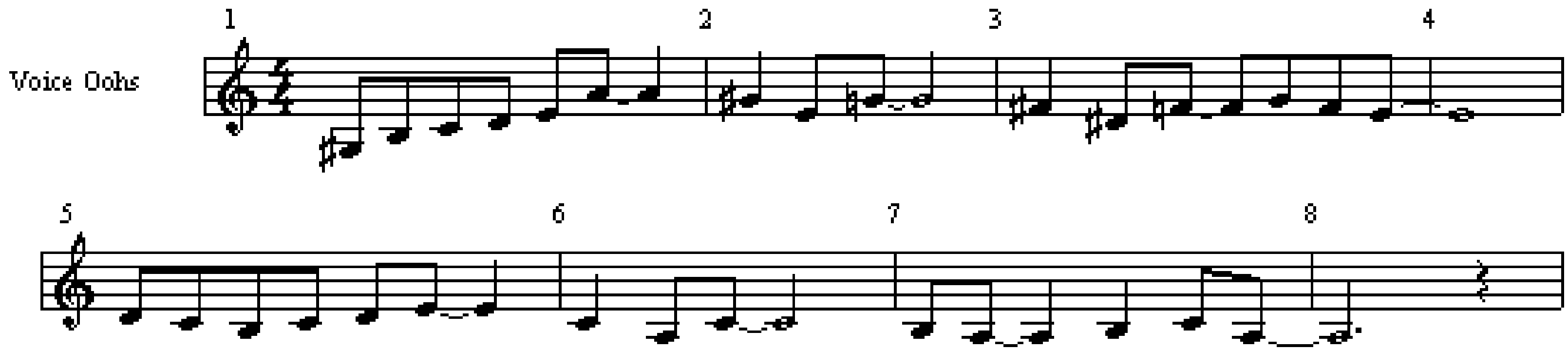


- Melodic continuity (*N*-gram), *musical key* constraints

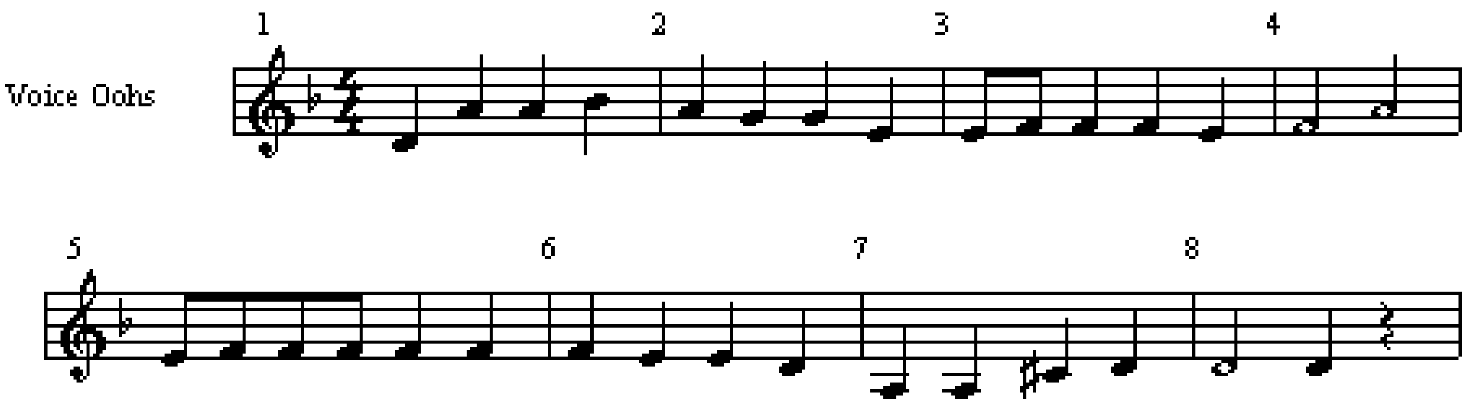


Results

- ”Brother can you spare me a dime”



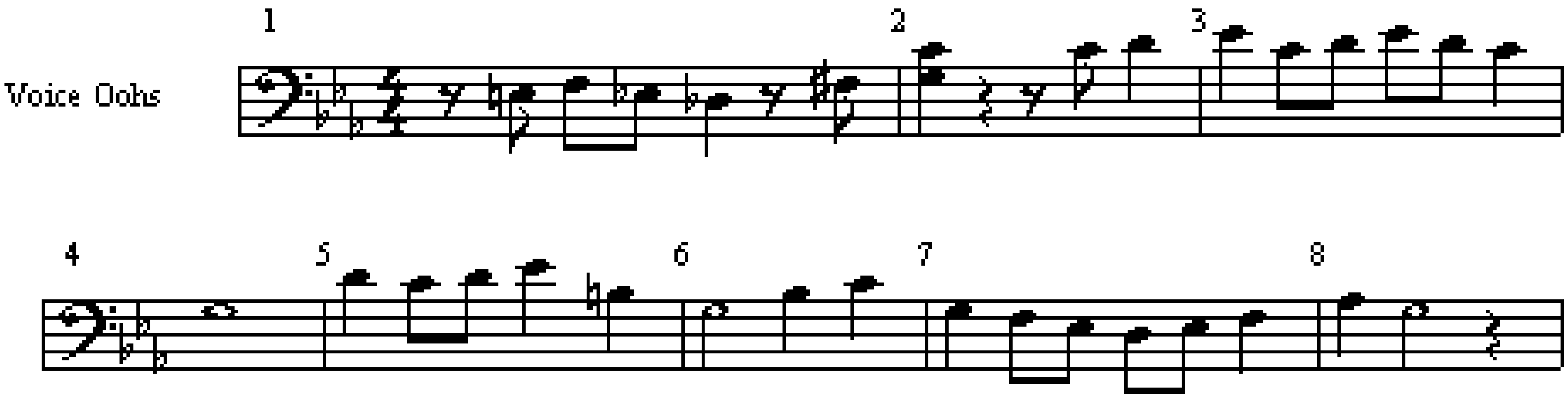
- ”Pieni tytön tylleröinen”



Singing transcription system

Results

- ”Puhelinlangat laulaa”



- More: http://www.cs.tut.fi/sgn/arg/matti/demos/monomel.html