PITCH-MARKING BASED ON THE DFE ALGORITHM

6th ECESS meeting

Hynek Bořil & Petr Pollák

Czech Technical University in Prague, Faculty of Electrical Engineering
CTU FEE K13131, Technická 2, 166 27 Prague, Czech Republic
Phone: +420 224 352 820; e-mail: borilh@gmail.com, pollak@fel.cvut.cz

MOTIVATION

Pitch-marks (PMs)

- Locations of significant instantenous energy peaks in the pitch periods.
- Correspond to the glottal closure instants (GCI).
- Crucial for pitch-synchronous speech analysis and synthesis (LPC, PSOLA).
- Energy peak locations - to reduce distortions during the overlapping.
- Detected from electroglottograph (EGG) or extracted from the speech signal.

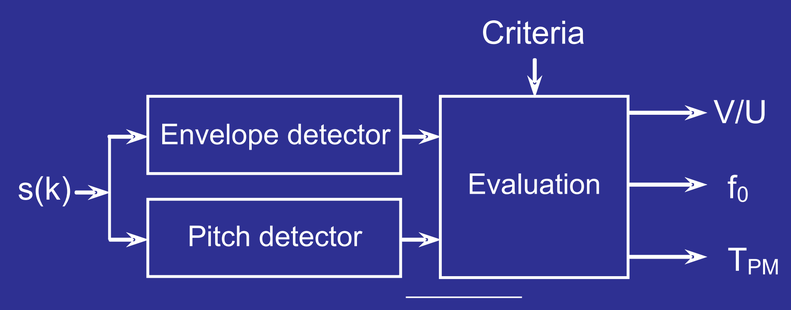
Pitch Extraction

- Autocorrelation - better time/frequency resolution than STFT or LPC.
- Signal segmentation required.
- Phase information loss.
- Multiple variable-variable multiplications => computation costs.

Goals of DFE

- Comparable time/frequency resolution to autocorr.
- Phase information preserved.
- No signal segmentation.
- Reduction of computation costs.
- PM detection as an integral part of the algorithm.

DFE CHAIN

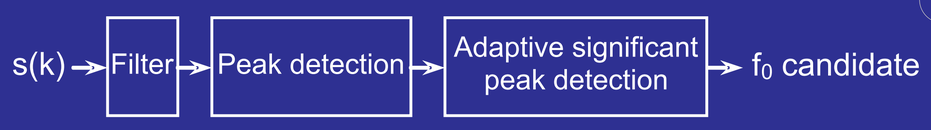


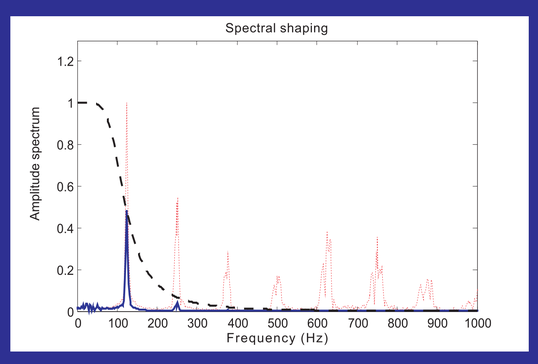
Envelope detector - a short-time moving average of the signal energy.
- realized by low-pass FIR filtering of the squared signal.
- FIR order - compromise - envelope smoothing
- ability to follow fast energy changes on the voiced/unvoiced (V/U) boundaries.

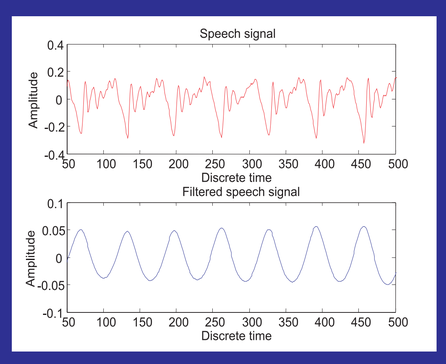
Pitch detector - frequencies are detected from significant peak-to-peak distances.

Evaluation - truth criteria are applied to the data from the envelope and pitch detector .

PITCH DETECTOR



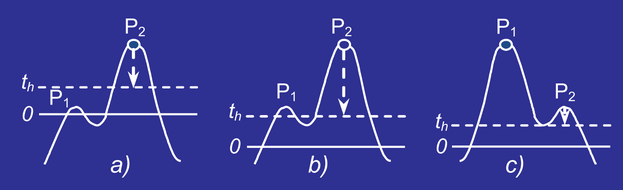




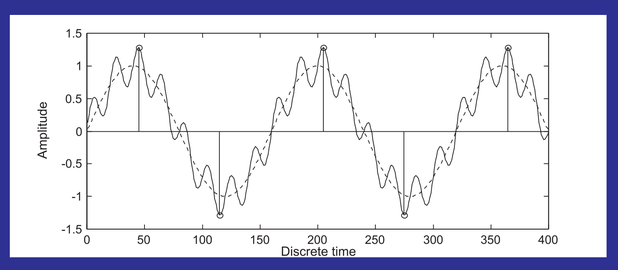
- After spectral shaping, all local extremes are detected.
- Due to the low order of the filter, some "false" peaks and zero-crossings still may remain in the signal.
- To identify locations of singificant extremes, the adaptive significant peak detection based on the neighboring peaks thresholding is performed.

- P_{last} - last significant peak detected before P_1.
- ZC(X, Y) = 1 if there is at least one zero-crossing between peaks X and Y, else 0.
- Then P_1 is significant peak related to the maximum only if not:

$$P_1 < 0 \quad \cup \quad ZC(P_{last}, P_1) = 0 \quad \cup \quad P_1 < P_2 \cdot th \quad \cup \quad (P_1 < P_2 \quad \cap \quad ZC(P_1, P_2) = 0)$$



Example of significant peak properties



Example - significant peak detection in additive noise

Pitch-mark Extraction

- Neighboring significant peaks bound a pitch period.
- Within the pitch period, the pitch-mark is determined as

$$k_{PM} = \arg\max_k(s(k)) \quad \text{or} \quad k_{PM} = \arg\min_k(s(k)), \quad k_{P1} \le k \le k_{P2}$$

- The min/max function is chosen to follow peaks that exceed neighboring local extremes more significantly.
- Consistency of the pitch-marks, i.e. PM distance vs. pitch period length, is evaluated.
- If it is not possible to detect PM positively as a signal extreme, actual PM position is determined from the previous PM and actual pitch period length.

CRITERIA

Energy

- Actual level of energy *E(k)* is evaluated by the envelope detector.
- No frequency estimations for signal level lower than the threshold *Eth*.

Frequency Range

- No frequency out of the specified range 60 - 600 Hz can be a valid estimation.

M-order Majority Criterion

- More than half of *M* consecutively detected freqs must lie in the same frequency band of a chosen width ⇒ **voiced/unvoiced classification.**

MAJORITY CRITERION

Definition

- {*fm*} - sequence of *M* consecutively detected frequencies.
- Let *fk* ∈{*fm*}.
- $count_{fk}(\{fm\})$ - number of *f* that

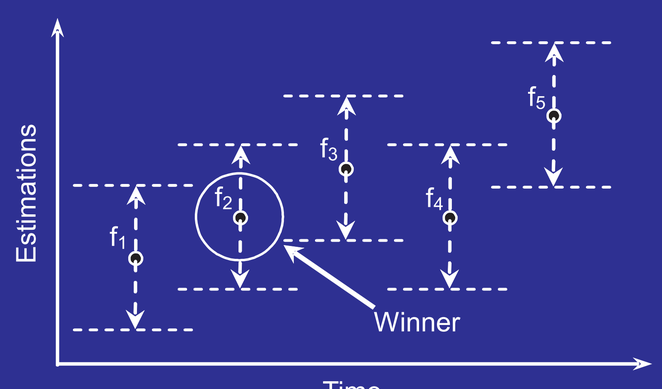
$$f \in \{f_m\} \quad \cap \quad f \in \left(\frac{f_k}{\sqrt[24]{2}}; f_k \cdot \sqrt[24]{2} \right) \qquad (1)$$

- The interval in (1) equals to the frequency bandwith of 1 halftone, centered to *fk*.

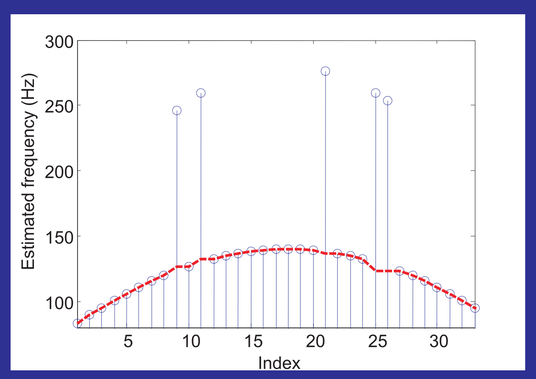
$$p = \max_k(count_{fk}(\{f_m\})), \; q = \arg\max_k(count_{fk}(\{f_m\})), \; k = 1,...,M \qquad (2)$$

$$\text{If} \quad p > \left\lfloor \frac{M}{2} \right\rfloor \quad \Rightarrow \quad f_{est} = f_q. \qquad (3)$$

- If more than one *fk* satisfies (2) and (3) => *fest* = *fmin(k)*.
- **If majority criterion is satisfied => actual signal is evaluated as voiced.**



Criterion - Principle



Solution of the frequency doubling problem

TESTS

Pitch-tracking Evaluation

- **#f** - number of detected frequencies that were compared to the referential channel frequencies.
- **Average difference**

$$\overline{\Delta} = \frac{1}{N} \sum_{n=1}^{N} \Delta_n, \quad \Delta_n = 1200 \cdot \log_2 \frac{f_2}{f_1} \quad (\%).$$

- **VE - voiced error** - *Tref* and *T* are total voiced times in the referential and evaluated channel.

$$VE = \left| \frac{T_{ref} - T}{T_{ref}} \cdot 100 \right| \quad (\%)$$

- **OE** - **octave errors** - number of differences equal or greater than one octave.

- **Standard deviation** - octave errors excluded

$$\sigma = \sqrt{\frac{1}{N} \sum_{n=1}^{N} \left(\Delta_n - \overline{\Delta}^* \right)^2}, \quad \Delta_n < 1200 \quad (\%)$$

SNR/SNR$_{ref}$ (dB)	#f	$\overline{\Delta}$ (%)	OE (%)	σ (%)	VE (%)
D/P 28.1/28.1	188734	39.69	1.11	64.31	N/A
D/D 17.9/28.1	147545	33.14	0.25	60.06	0.47
P/P 17.9/28.1	76957	80.44	3.16	66.50	N/A
D/D 9.6/28.1	94516	103.47	4.98	102.66	21.53
P/P 9.6/28.1	72742	133.64	5.48	92.12	N/A
D/D 4.9/28.1	5100	246.43	15.01	141.48	92.24
P/P 4.9/28.1	48096	1157.76	51.36	206.76	N/A

- D - DFE channel.
- P - Praat autocorrelation channel.

- Tests were performed on the Czech Speecon database.
- Performance was compared to the Praat modified autocorrelation algorithm (www.praat.org).