# MEASURING INTONATION AT LOW SIGNAL-TO-NOISE RATIOS

V.Pikturna

Technological university, Kaunas, Lithuania

**ABSTRACT**

The method is proposed for evaluating intonation curve from the highly corrupted speech signal. During local processing the adaptive threshold is applied to the short-time FFT-spectrum, pitch harmonics are identified and pitch frequency determined. During glob processing, the intonation curve is smoothed and approximated by the low-order polynomial.

## 1. INTRODUCTION

Evaluating intonation when signal is corrupted with noise is a problem of great difficulty, especially in speech communication systems where only the past of the signal's properties can be taken into consideration. There are however applications where measuring in real time is not necessary, e.g. teaching of deaf persons to speak, studying foreign languages, speech rehabilitation after operations etc. In these cases, uttering must be followed by an intonation curve on the screen for visual comparison to a reference one. This situation is less complicated because shaping of the intonation contour is possible, and both past and future values can be taken into account at every point of it.

When measuring intonation from spectral data, identifying of pitch harmonics simplifies calculating of pitch frequency (PF). The method is trended towards looking for periodicity in the corrupted spectra of speech, so it can find a "pitch" in the spectra of noise too [2]. Therefore the great attention is payed to recognition of noisy frames. The essential features of the method proposed are:
(1) employing of the adaptive threshold (ATH);
(2) identifying of pitch harmonics by their amplitudes, shapes and symmetry;
(3) usage of a multistage procedure for the voiced/unvoiced decision.
The block diagram of the algorithm is presented in Fig.1.

## 2. IDENTIFYING OF HARMONICS

### 2.1. Evaluating of the Short-Time Spectrum

We suppose at least three pitch harmonics to be necessary for taking decision about the PF. If the highest PF for a femail speaker is 450 Hz then the frequency region under consideration must be at least 1350 Hz (1430 Hz in our hardware). The signal is weighted by the Hamming window and zeroes are added to obtain the FFT spectrum ( in the logarithmic scale )



at 64 spectral points. The spectral resolution is 22.3 Hz, the measuring accuracy is improved by parabolic interpolation of spectral peaks.

### 2.2. Adaptive Threshold

A horizontal threshold has a principle disadvantage related to the formant structure of the spectrum: it can either not reach harmonics in the region between formants or cross the spectral components related to background noise. The ATH is obviously necessary changing its shape when the spectral properties of the speech signal change. We propose for this purpose the spectrum of the linear prediction (LP) model. As the narrow frequency band is considered, the low-order LP models can be used. Fig.2 illustrates the effect of thresholding for different sounds and signal-to-noise ratios (SNR), when the ATH is of the type:

$$H(z) = 20 \lg \left| 1 + a_1 z^{-1} + a_2 z^{-2} + a_3 z^{-3} \right|^{-2}$$

$a(i)$ being the LP coefficients, $z = \exp(-jw)$, $w$ being the current frequency. The value of shifting downwards the ATH depends on the SNR and is discussed in [2].

### 2.3. Examination of spectral peaks

The three parameters of every spectral peak exceeding the ATH are examined: amplitude, sharpness and symmetry. The amplitudes are calculated directly from the spectrum (see e.g.[4]) while sharpness and symmetry are evaluated by the parabolic approximation of a spectral peak: the coefficient $a$ of a parabola and the approximation error correspondingly. The ranges of values for these parameters are defined in advance, using statistics of natural speech [2]. A spectral peak is considered a pitch harmonic provided all the three parameters are within the ranges defined.

## 3. CALCULATING OF THE PITCH FREQUENCY

The data for calculating PF are $F(k)$, the frequencies and $A(k)$, the levels of maxima of spectral peaks. Obviously, $k$ is not always a number of a pitch harmonic. We have chosen a method of evaluating PF most close to the visual one: we consider the average distance among harmonics to be the PF. The evaluating is carried out in 2 steps:
(1) the initial value of PF is calculated as the average distance among three harmonics: one of the maximum $A(k)$ all over the spectrum and two closest to it (one from the left and another from the right). The possibility of lacking one (or two) harmonics among these 3 ones is accounted. Such an approach allows to find a correct value of the PF even of high corrupted signal. We find this approach more reliable than those concerning spectral peaks starting from the very first on the left (e.g.[1]). If no equidistance among the three harmonics can be found, the same procedure is repeated with the other three ones in the neighbourhood (on the left and, if necessary, on the right).
(2) the distances between all harmonics approximately equal to the initial value are averaged.

## 4. RECOGNITION OF UNVOICED FRAMES

### 4.1. Spectral energy

The unvoiced sounds are of little low-frequency energy.

We have empirically fixed the level of -(10...15) dB for a horizontal threshold which must not be exceeded to identify the corresponding frame as voiced (Fig.1, V/UV1). This scheme works reliably at high SNR only.

### 4.2. Flattness of the spectrum

The slope of spectra of the white noise computed from short frames is much less than that of voiced sounds [2]. The dynamic range $\Delta$ of the ATH shows to be the very efficient measure of the spectral flattness. We formulate the following feature: a frame is unvoiced if

$\Delta$<10 dB when SNR>10 dB,
$\Delta$< 7 dB when SNR<10 dB
(Fig.1, V/UV2).

### 4.3. Number and disposition of harmonics

If the processing of spectrum results in finding less than 3 spectral peaks, the frame is labeled unvoiced (Fig.1, V/UV3).

If examining of three peaks in the region of spectral energy maximum does not result in finding equidistancies, the frame is labeled unvoiced (Fig.1, V/UV4).

## 5. SHAPING OF THE INTONATION CURVE

### 5.1. Jumps to a neighbouring harmonic

To avoid jumps to the 2nd or to the 0.5th harmonic, the past of the intonation curve is used: the current value of the PF is compared to the average of all previous non-zero values of the PF. If it exceeds twice or is twice less than the average mentioned, it is devided (multiplied) by 2. If the declination is greater than 2 times, the PF is set to zero. We find such an approach more effective than one-step-back control.

### 5.2. Smoothing and approximating

The 3-points nonlinear smoother [3] and polynomial approximation are applied to the intonation curve. When approximating by a polynomial, the question arises how long must be the segments under approximation. Approximating of every voiced segment and of the whole curve are two extremities. Fig.3 shows the intonation curve consisting of 5 voiced segments where 3 and 2 segments are approximated by the 3rd and 4th order polynomials.

## 6. RESULTS

The method was tested with 3 speakers (two males and one femail) using a limited speech material. When using knowledge of a human expert, the intonation curve remains at SNR down to 0 dB.

## 7. REFERENCES


[1]ALLIK,J.,MIHKLA,M., ROSS, J.(1984),"Comment on "Measurement of pitch in speech:An implementation of Goldsteins theory of pitch perception", *J.Acoust.Soc.Am.*,75(6),1855-1857.

[2]PIKTURNA,V., RUDŽIONIS,A. (1990),"Pitch measuring from spectra of noisy speech:amplitude thresholding versus identifying of harmonics", *Proc. 3rd Australian Int. Conf. on Speech Science and Technology*, Melbourn, 6 p.

[3]RABINER,L.R.,SAMBUR,M.R., SCHMIDT,C.E. (1975), "Applications of a nonlinear smoothing algorithm to speech processing", *IEEE Trans. Acoust.,Speech and Signal Processing*,23,554-557.

[4]SREENIVAS,T.V.,RAO,P.V.S. (1979),"Pitch extraction from corrupted harmonics of the power spectrum", *J.Acoust.Soc.Am.*,65,223-228.




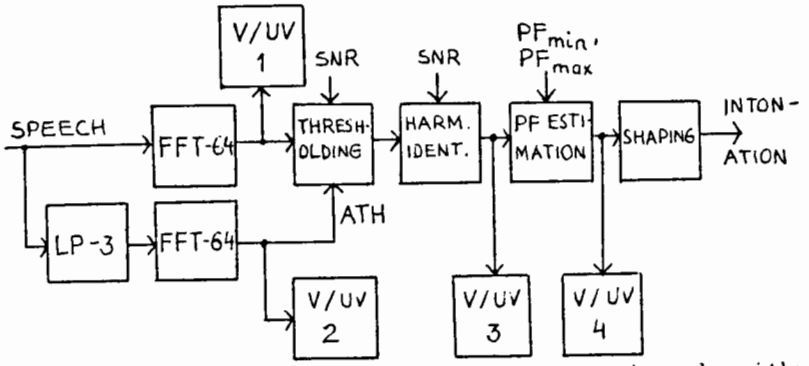


Fig.1. Block diagram of the intonation measuring algorithm

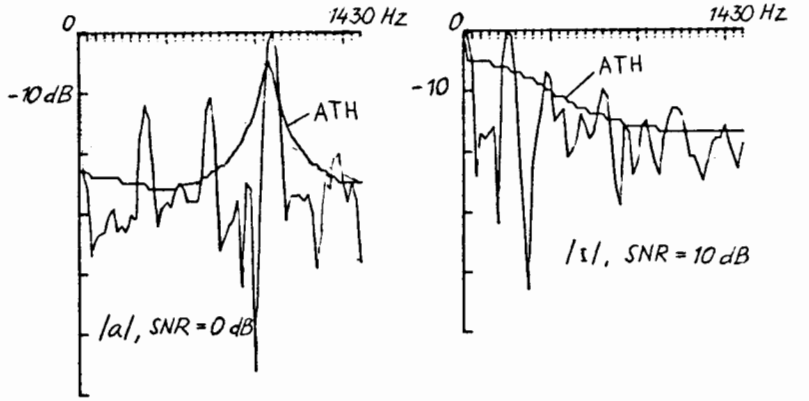


Fig.2. Effect of the adaptive threshold ATH

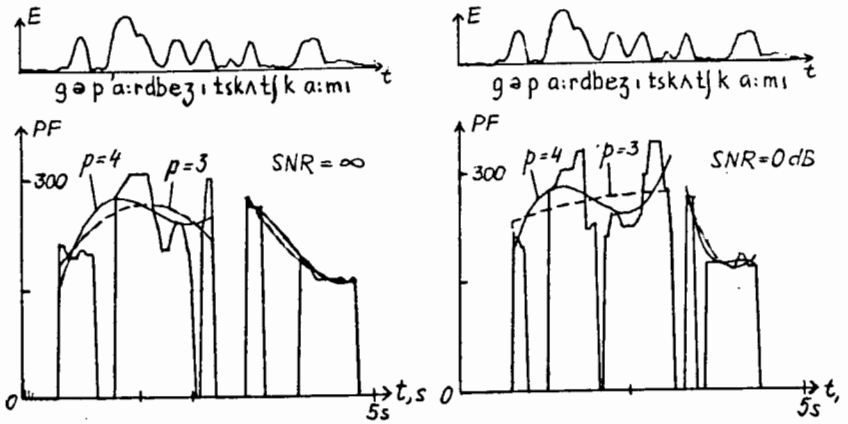


Fig.3. Intonation curves approximated by the 3rd and 4th order polynomials