HASR SPEAKER RECOGNITION METHOD BASED ON THE STATISTICAL ANALYSIS OF F0

Porticus Inc.

Bernardas Salna (bernardas@centras.lt)

Juozas Salna (juozas.salna@gmail.com)

Germano Di Mambro (germano@porticusinc.com)

Statistical analysis of F0 of software package SIVE was used for HASR1 test by Porticus team. Because HASR1 test contains records in different recording conditions, we can use only one F0 statistical analysis from our SIVE package.

F0 is relatively robust against poor recording quality and differences in transmission channel. In first stage of investigation we hand segmented test and model records and normalized by amplitude and after that calculated F0 for each test and model segment.

The pitch estimation procedure consists of two separate programs:

1. Pitch and its derivatives calculation program;
2. Statistical comparison of the pitch and its derivatives program.

Pitch and its derivatives calculation

Speech signal S(n) divide in frame 30-40 ms length. For each frame calculate

Forth order LPC (linear prediction) parameters are calculated and residue signal e(n) is obtained . Next the FFT is applied to the signal e(n) and we get spectrum of a residue signal (see Fig.1).

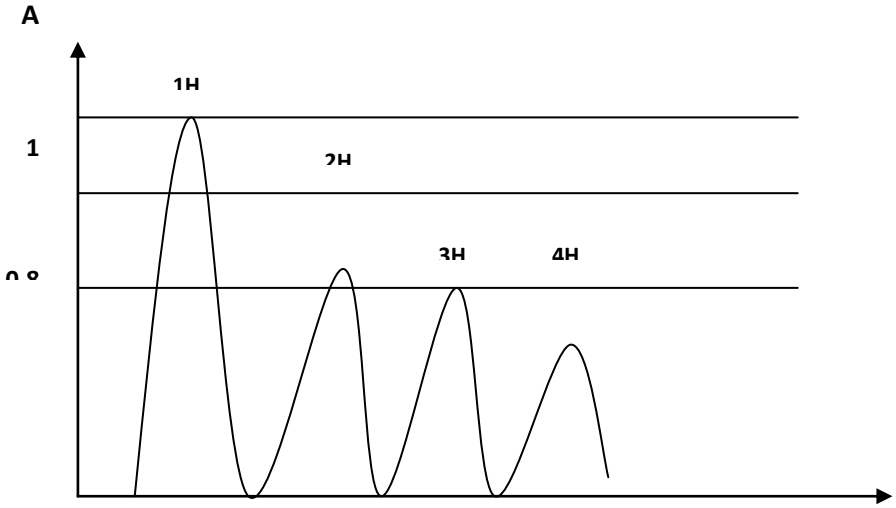

Fig.1. Spectrum of a residue signal.

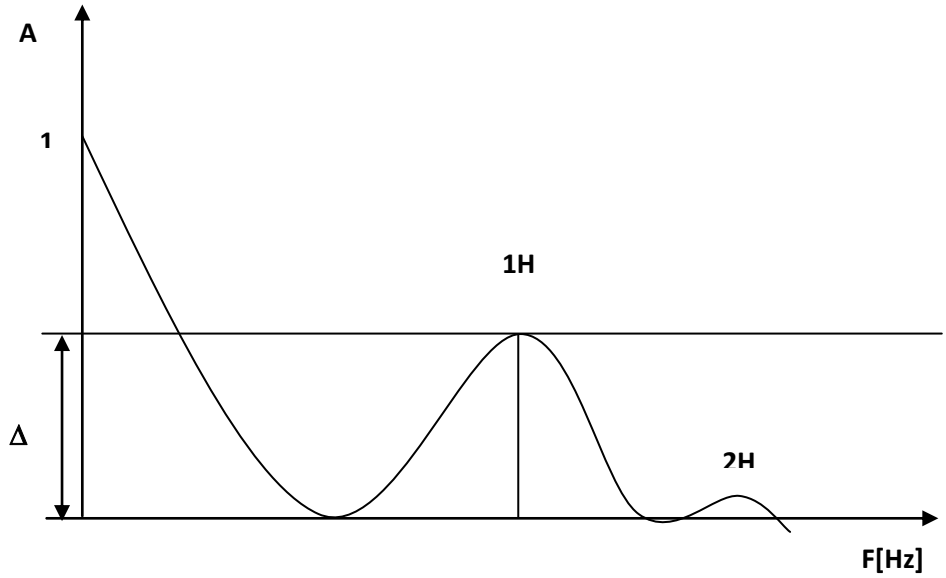

Fig.2 . Autocorrelation function of the residue signal spectrum

Further autocorrelation function of the spectrum is calculated and after from spectrum of residue signal. As we see from Fig.2, autocorrelation function has maximum at the frequency of the first harmonic and 1H is pitch value.

From Spectrum of residue signal calculate Pitch derivative parameters: number of the pitch maximal harmonics (MH), voice clearness (VC) and timbre (T).

MH is pitch harmonic which have maximal amplitude, it is found from the residue signal spectrum in the range 60-3000 Hz. In Fig.1. maximal harmonic is the first harmonic.

Voice clearness (VC) is calculated in the following way:

$$VC = \frac{PGT * MH}{CR_{0.5}},$$

where $CR_{0.5}$ is line 0.5 crossing number of the normalized residue spectrum ,

MH – pitch maximal harmonic in Hz

The timbre is calculated in the following way :

$$T = \frac{PGT}{2CR_{0.8}},$$

where $CR_{0.8}$ is line 0.8 crossing number of the normalized residue spectrum.

Statistical comparison of the pitch and its derivative parameters

Statistical comparison of the pitch and its derivative parameters is devoted to calculate, display in the graphic form and to compare results of the pitch calculation. Before calculation of the statistics, smoothing of the pitch trajectory is performed.

After smoothing, histogram of the pitch values is calculated. The histogram displays distribution of the pitch frequency in the range of $0 \div 500$ Hz. Also pitch minimum, maximum variance and variation coefficient are calculated. The same statistical calculations are performed with the derivative parameters of the pitch: number of the pitch maximal harmonic, voice clearness and timbre. Statistics of all these parameters are calculated for investigative as well as for comparative pitch files. While comparing two pitch files, additionally correlation coefficient between corresponding histograms and common coincidence coefficient (which consists of correlation coefficients with corresponding weights) are calculated. Common coincidence coefficient is calculated according to the formula:

$$K = \frac{3K_{PGT} + K_{MH} + K_{BS} + K_T}{6}. \qquad (5)$$

Identification procedure is performed in the following way. From comparative voice record speech signal (words and phrases uttered by this person) of the known person is selected manually. This voice signal is written into separate file, namely comparative file. Then phrases of the unknown person are selected from the investigative speech record and written into investigative speech file. For investigative and comparative speech files pitch and derivative parameters are calculated. Then statistical comparison of these results is performed. Having statistical results we evaluate these results: mean, minimum, maximum, variance, variation coefficient of the pitch, correlation coefficient between histograms. If common coincidence coefficient is:

Greater than 0.85 it is possible to say with high probability that it is the same person.

0.85-0.75 probability that it is the same person,

0.75-0.65 result is uncertain.

0.65-0.45 most probably speech corresponds to different person.

If less than 0.45, then it is possible to say with high probability that speech records correspond to different persons.

In HASR case, if this coefficient is more than 0.85 (TRUE) , it is the same persons voice, if less than 0.85 (FALSE) – voices is different person.