PITCH DETECTION RELIABILITY ASSESSMENT FOR FORENSIC APPLICATIONS

Serguei Koval, Veronika Bekasova, Michael Khitrov, Andrey Raev
Speech Technology Center, Sankt- Petersburg, Russia
Tel./Fax:+7(812)3279297; E-mail:master@stc.rus. net

ABSTRACT

For some tasks (e.g. forensic applications) it is vital important to know real pitch. So there is a problem to check-up the correctness of any concrete pitch contour for long speech records for any pitch detection method. Besides it would be useful to see the degree of signal periodicity without strong decision voice/noise.

The new homomorphic method of signal periodicity degree detection with analysis frame length in proportion to the time-lag is described. The powerfull working approach to visual analysis of speech signal periodicity is proposed. This *Voicograms* method of speech periodicity representation ensures the practical correctness of pitch estimation and allows to find periodicity degree for poor quality signal.

ITRODUCTION

Pitch detection algorithms (PDA) and tools have a long history, which goes back even beyond vocoders time [1, 16]. Nevertheless up to now the task of "in-field" pitch estimation must be counted among the most difficult problems of speech analysis.

In spite of existence of many methods of pitch detection and voice/noise decision (e.g.[1-11, 16-18]) real user often has only very poor means for exact extraction of fundamental frequency. The leading specialists in the field of speech analysis agree that: "there is no pitch detection algorithms which operates without errors. There is no reference algorithm, even with instrumental support, that goes completely without manual inspection or control" [16].

Dealing with real speech signal the usual assumption of quasiperiodicity is often rather far from reality [17]. Pitch is only metaphorical, general concept, describing the speech signal periodicity only conditionally. Every voiced speech pitch-period is generally not equal to its neighbours. It has more or less different amplitude, length, form, spectrum. Every concrete pitch extraction algorithm sets admissible boundaries for these differences and rules of decision: how much may one pitch differ from another to remain pitch at all, and what is reckoned to be a period of repetition for non-equal periods. Unfortunately these rules are yet rather far from practice needs for some tasks. So often user obtains numerous mistakes, using any software or hardware for pitch detection, especially for telephone-quality or noise-corrupted speech. For many users this situation is not so essential. E.g. "normal" phonetician can accept 5% mistakes in pitch contour. He has no exigencies to use noisy or poor quality phone speech. But for some tasks it is vital important to know real pitch. For example, forensic expert in the field of speaker identification by voice has very bad speech signal records as usual, have no possibilities to repeat the speech record and must not be mistaken using pitch curves for decision [12, 15]. To drastic variation in temporal structure of the speech signal, even between subsequent pitch periods, narrowband formants at low pitch harmonics, wide voice range (up to 4 octaves), the glottal waveform intrinsic occasional irregularities, voice fry are springing up additional problem when the signal is band limited, noise corrupted or channel performance distorted. E.g. phase distorted after usual analogue recorder speech signal badly fits for pitch detection especially by time-domain methods.

For responsible forensic applications (such as speaker identification by voice records and lie detection by voice) it is necessary to evaluate the performance of the measuring device or software for every concrete working signal. Responsibility of the expert decision (which can some times determine the personal guilt or innocence), demands the on-line evaluation of correctness of current pitch detection results. Often the 1-2% of data can essentially influence the expert decision. It is more important to exclude any possibility of systematic mistake. E.g. typical "Gross" pitch determination errors (in terms of [3]) are selection of subharmonics, or irregular values.

There are some PDA with "good reputation" but all of them make mistakes dealing with noise or tape-recorder, or channel-corrupted speech.

It is a main problem for any algorithm, when it detects an estimate that seems to be wrong to detect reliably whether that is due to a measurement failure or to a momentary irregularity of the signal.

MOTIVATION

These considerations say that the effective method of pitch detection correctness should be an important part of speech signal forensic analysis. This method should be on-line or "near on-line" and allow fast processing with big data arrays (some tens minutes of speech signal. Besides we think many speech users and researchers would be glad to have suitable and easy-to-use means to see the degree of speech signal periodicity without binary decision voice/noise.

Among modern numerous systems and methods of automatic speech and speaker recognition there are very little ones using pitch as working speech signal feature. One of the main reasons of this is the absence of good pitch extractors for noisy speech. The lack of good method of pitch detectors correctness check-up causes the slow progress in this direction of research.

Usual means of correctness control of picked out pitch curve are manual visual waveform analysis, phase portrait analysis, narrow-band spectral analysis, local smoothness analysis of cyclic parameter and usage of laryngograph [17]. Last method is impossible for usual work with speech records, other methods have no realisation for on-line work or work with large amount of speech material to process. Perceptual analysis of pitch often gives data quite far from objective results [16,18].

For "heavy" signals there is the hard attitude of adepts [e.g.16-18] that visual displays (e.g. spectrograms, phase portraits, next cycle parameter contours) are the most useful for understanding of the physical characteristic of the voice oscillating system.

This report describes the new form of visual easy-to-read signal periodicity degree display and working technology for PDA evaluation. training and correction.

METHOD

The result domain of cyclic pitch period is equivalent to the time domain. To avoid confusion, we will henceforth call it the *lag* domain.

After comparison of many speech signal periodicity characteristic functions we find out the best one. It is proposed to be the real cepstrum-based approach [13,14] with the frame length proportional to the time lag:

Main points of the method are following:

1.We calculate the normalised short-time autocorrelation function $\phi(t,\tau)$ with the time frame length in proportion to the time lag [10]:

$$\phi(t,\tau)=\frac{\int_{t-\frac{\tau}{2}}^{t+\frac{\tau}{2}} x(u-\frac{\tau}{2})x(u+\frac{\tau}{2})\,du}{\sqrt{\int_{t-\frac{\tau}{2}}^{t+\frac{\tau}{2}} x(u-\frac{\tau}{2})^2\,du \int_{t-\frac{\tau}{2}}^{t+\frac{\tau}{2}} x(u+\frac{\tau}{2})^2\,du}},$$

where t - current time, $x(t)$ - speech signal, τ - corresponding time-lag,

2. Add some zeros and calculate the inverse fast Fourier transform.

3. Limit spectral range for the best periodicity analysis.

4. Weight the result with the Nuttall window in frequency area and calculate the ***log*** function.

5. Calculate direct Fourier transform and have new periodicity function $\Phi(t,\tau)$.

6. Normalise the $\Phi(t,\tau)$. for t - axis with the slope 3 dB/octave.

We represent this function graphically by assuming lag to the ordinate, time (frame index) to the abscissa and the value of $\Phi(t,\tau)$. at the corresponding time and lag to the degree of shading with manual choice of the best visual and calculating parameters. We call the resulting picture *Voicogram* by analogy to *Sonogram.*

After very short training common user can see and optimally customise the pictures of periodicity function for any signal on the PC screen. The concrete set of options is signal-dependent. It is the reason why the pitch extraction methods have not so good performance. They are too simple for real noisy speech.

To use the technology above for pitch detection in respocible applications we should have the PDA with some get-at-able options. After pitch contours detection with default options this contour should be drawn over the Voicogram. The obviousity of the correspondence or discrepancy between the detected pitch and signal periodicity function are sufficient basis for PDA options correction.

In our working system Speech Interactive Software SIS 4.5 besides the Voicogram-based pitch correction expert can use the hearing control to listen to the speech samples, which were considered in used pitch detection method to be «noise' or «silence» ones. It is possible e.g. listen to speech intervals with the zero-crossing function more than used current threshold frequency, with energy lower than used current energy threshold for pauses etc. Auditory feedback allows to adjust threshold reasonably and extract voiced speech

samples even for very noisy speech and for relatively long speech files.

RESULTS

The happy choice of periodicity function and chosen controls of visualisation provide experts with very obvious and clear displays of cyclic oscillations in signal analysed. The direct comparison with manual waveform analysis, dynamic spectrograms analysis and correlograms-based technique [17] shows the indisputable advantages of the new approach. The manual analysis of the waveforms confirms the validity of the new technique.

Fig.1-4 illustrate the usage of this technology to the pitch contour correctness check-up. If user has the pitch curve, as on Fig.1, he (or she) does not know if it is the true pitch or not. The comparison with sonogram or oscillogram can not help to solve this problem for speech of pour quality or for big quantity of speech. According to the method discussed above, one can see the Voicogram for the same signal (Fig.2), put pitch contour on it (Fig.3), and see, that: for marked with circles fragments: 1- mistakes of pitch values, 2- transfer to 2-nd harmonic, 3- transfer to 1/2 harmonic, 4- noise instead of voice, 5- voice instead of noise.

CONCLUSION

The proposed method of check-up of pitch detection correctness is proved by many-years successful forensic, medical and research practice. Discussed technique provides users with real possibility to evaluate, train and correct any pitch detection algorithm or tool. The direct comparison with correlograms-based technique [17,18] and other techniques [18] shows the indisputable advantages of the new approach. We think the usage of *Voicograms* could be common standard for voice periodicity analysis just as *Sonograms* analysis for formants tracking.

REFERENCES

[1] M.Gruetzmacher, W.Lottermoser, "Ueber ein Verfahren zur traegheitsfreien Aufzeichnung von Melodiekurven", Akustische Z., 1937, h.2s.242-248.

[2] H. Fujisaki, "Automatic extraction of fundamental period of speech by auto-correlation analysis and peak detection", JASA, 1960, vol.32 (A), p.1518.

[3]L.R.Rabiner,M.J.Cheng,A.E.Rosenberg,C.A.McGonegal, "A comparative performance study of several pitch detection algorithms", IEEE Tr.on ASSP, 1976, vol. ASSP-24(5),pp.399-413.

[4]S. Seneff,. "Real time harmonic pitch detector", IEEE Tr. on ASSP, 1978, vol. ASSP-26 (4), pp.358-365.

[5]M.Lahat, R.J.Niederjohn,P.A. Krubsack, "A spectral autocorrelation method for measurement of the fundamental frequency of noise-corrupted speech" IEEE Tr.on ASSP, 1987, vol. ASSP-35 (6), pp.741-750.

[6] J. Picone et al. "Robust pitch detection in a noisy telephone environment", Proc. IEEE ICASSP-87, N.Y.: IEEE, 1987, pp. 1442-1445 .

[7]A.M.Sutherland, M.A.Jack, J.Laver, "Improved pitch detection algorithm employing temporal structure investigation of the speech waveform", IEE Proc., 1988, vol.135(2), PTF, pp.169-174.

[8] Y.M.Cheng, D.O.Shaughnessy, "Automatic and reliable estimation of glottal closure instant and period", IEEE Tr.on ASSP, 1989, vol. ASSP-37, pp. 1805-1815.

[9]D.W.Howard, "Peak-picking fundamental period estimation for hearing prostheses", JASA, 1989, vol. 86(3), pp.902-910.

[10]K.Hirose, H.Fujisaki, S.Seto, "A scheme for pitch extraction of speech using autocorrelation function with frame length proportional to the time lag", Proc. ICASSP-92, vol. I, pp.149-152.

[11]B.Mak"A robust speech/non-speech detection algorithm using time and frequency -based features". Proc. ICASSP-92, v.I, pp.269-272.

[12]H.J.Kuenzel."Sprechererkennung. Grudzuege forensisher Sprachverarbeitung". Heidelberg: Kriminalistik Verlag. 1987 .

[13] A.V.Oppenheim,R.W Schafer. "Digital signal processing". New Jersey, Englwood Cliffs: Prentice Hall, 1975.

[14] C.Rowden "Analysis, Speech Processing", London: McGraw Hill, 1992.

[15] H.Hollien "The Acoustics of Crime. The New Science of Forensic Phonetics", N.Y. and London: Plenum Press, 1990.

[16]W.J.Hess "Pitch Detection of Speech Signal - with Special Emphasis on Time-Domain Method", in Proc. WAVA, 1995, Denver: NCVS, pp. HES1-34.

[17]I.R.Titze "Summary statement of Workshop on Acoustic Voice Analysis", in Proc. WAVA, 1995, Denver: NCVS, pp TITZ2-36.

[18]D.Talkin "Cross correlation and Dynamic Programming for Estimation of Fundamental Frequency", in Proc. WAVA, 1995, Denver: NCVS, pp Talk 1-8

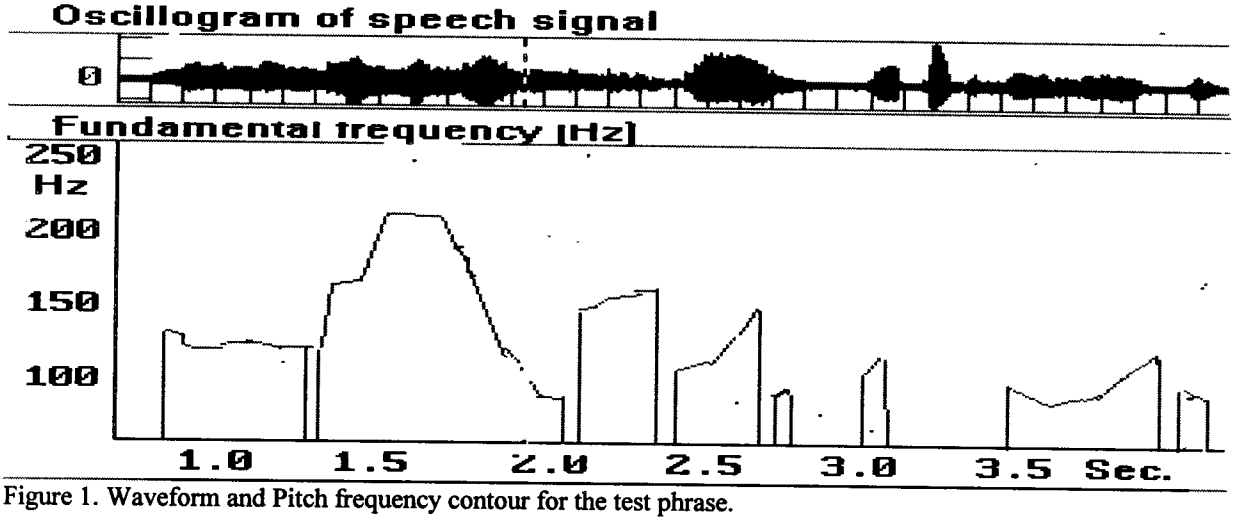

Figure 1. Waveform and Pitch frequency contour for the test phrase.

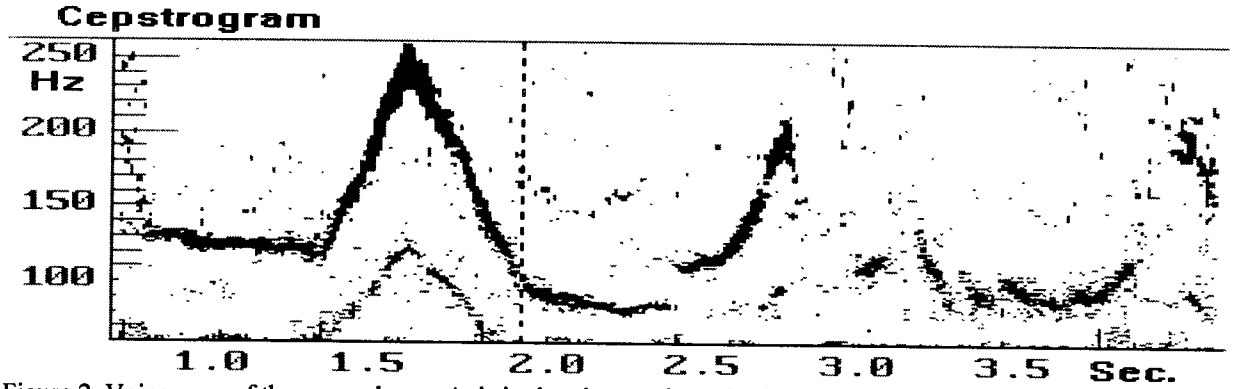

Figure 2 Voicogram of the same phrase. Axis in the picture: time - horizontally, Period duration – time lag - vertically, Shadow degree reflects the degree of the periodicity.

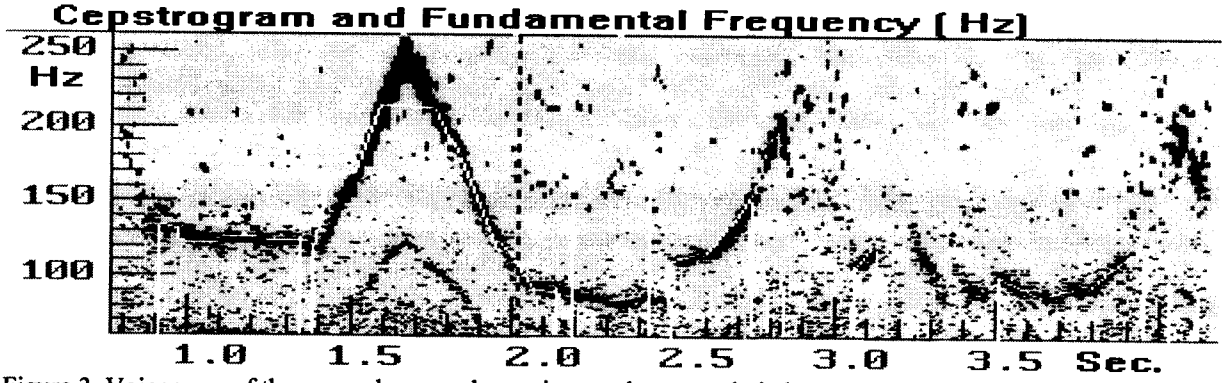

Figure 3 Voicogram of the same phrase and superimposed extracted pitch contour.

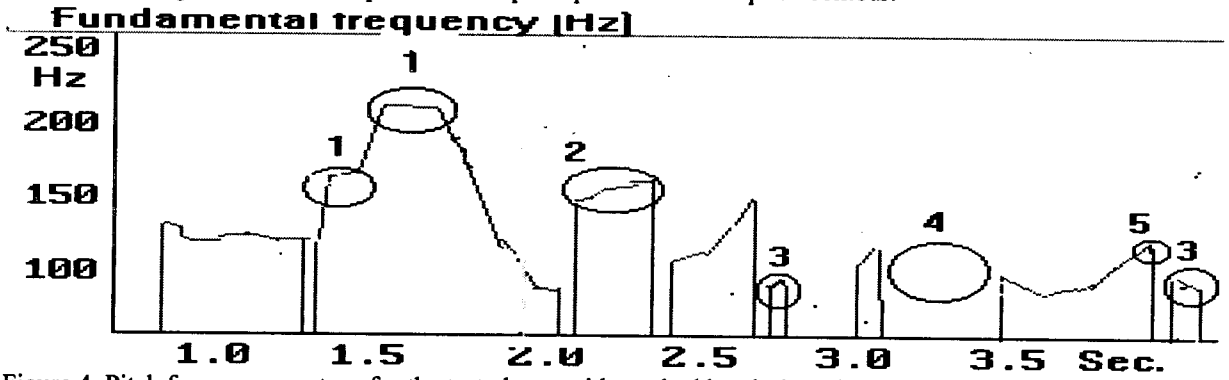

Figure 4. Pitch frequency contour for the test phrase with marked by circles pitch detection mistakes of different kind. See description in text.