PRINCIPLES OF THE HEARING PERIPHERY FUNCTIONING
IN NEW METHODS OF PITCH DETECTION AND SPEECH ENHANCEMENT

Ludmila Babkina*, Sergey Koval**, Alexander Molchanov*

* Research Institute of Ear, Nose, Throat and Speech disorders, St. Petersburg, Russia
** Speech Technology Center, St. Petersburg, Russia
Tel./fax: +7(812)3279297, E-mail: master@stc.rus.net

ABSTRACT

Spent researches show that one of mechanisms of human auditory system ensuring high noise resistance of vocal speech sounds recognition is an electromechanical envelope feedback, effecting in structures of inner ear in man. Digital modeling of hearing system peripheral section with a similar multichannel envelopes feedback has shown to be useful for pitch determination of vowels in noisy environment. The offered model provides robust pitch detection for signal/noise relation up to \(-12\) - \(-14\) dB. In number of cases such a noiseproof feature is better than for other existing methods and systems.

INTRODUCTION

Up to now the human hearing system is a better instrument for noisy speech recognition than the majority of technical means. One of the important problems lays in the adequate modeling of human periphery frequency-selective performance and in particularly in sharp understanding the role of the feedback in hearing processes.

The problem of determination of the fundamental tone of vocal speech sounds is traditionally one of basic for analysis, coding and automatic recognition of speech. Numerous technical and software developments described in publications are devoted to this problem. Only a small part of them can be mentioned in the given work (see, for example, review [5]).

There is the well-known method of pitch pulses determination using amplitude selection [1] and consisting in detection of speech signal envelope amplitude, determination of maximal values of this envelope, for example, with the help of peak-peaking and central restriction with the subsequent estimation of intervals between points of maxima.

Disadvantage of this method is low noise resistance to amplitude distortions of speech and to pulse hindrances, forming false clicks in speech signal envelope.

There are some other ways of speech fundamental frequency detection based on the spectral analysis of voiced speech samples, determination of maxima positions in low-frequency spectral area and selection among them the one corresponding to the pitch frequency [2].

To spectral methods of pitch detection adjoin other «windows»-based methods, such as correlational [3] and cepstral ones [4]. Thus the presence of a pitch in signal is displayed in existence of a maximum in auto correlation or cepstrum function of a signal.

The position of the maximum in «lag» - time dimension shows the estimation of the real pitch period value for given analyzed frame of speech signal.

Described above spectral, correlational and cepstral methods are realized in a wide range of systems and devices for pitch detection [5], [12]. The common shortcoming of these methods of pitch detection is a significant volume of calculations, which is necessary for their on-line realization, and also low noiseproof feature. Despite the extraordinary progress in development of powerful microprocessors the problem of realization of signal processing comprehensive methods, for example, in hearing aids of new generation is quite complex. High noise resistance of speech vowels recognition by humans gives the basis for hoping, that study of processes of the signal analysis in man's auditory system will allow to find out the ways of improvement of existing approaches to detect pitch and enhance speech perception by patients in noisy environment.

METHOD

As the characteristic of an estimation and quality comparison for different pitch detection methods we used the noise resistance, i.e. method's ability of correct determination of fundamental frequency of vowels sounds in a background noise.

In the last years understanding of principles of auditory periphery functioning have extended greatly. In particular has been proved the existence of an electromechanical feedback in inner ear of mammals, the action of which renders essential influence on formation of frequency-selective properties of auditory system [6].

We shall consider the mechanism of action of one of types of a feedback in a cochlea, i.e. feedback on summatiing potentials. The latter are formed by structures of organ of Corti and have the form of an effecting sound signal envelope. The hypothesis about existence of such an auditory feedback on envelope was formulated by the authors in 1978 [7].

The basic concept element of an auditory feedback on envelope in inner ear is occurrence in oscillatory system of a cochlea of waves with frequencies, equal to frequency of envelope of effecting stimulus. So if the stimulus $U_c$ consists of two tones with frequencies $f_1$ and $f_2$, then in a cochlea due to action of an auditory feedback on envelope there should appear the waves with frequencies $f_2 - f_1$ and $2f_1-f_2$. It is really observed both in otoacoustical emission [8], and in electrophysiological experiments on measurements of microphonic potentials [9], as well as in experiments, devoted to doubling effect of microphonic potenialcs frequency in mammals. [10].
RESULTS

In figure 2 as an example a normalized spectrum of a vowel sound «A » in the point $\gamma$ of the circuit fig. 1 is resulted by $\beta=0$.

In figure 3 a normalized spectrum of the same sound «A » in the point $\gamma$ is resulted by $\beta=0.9$

In figure 4 a normalized spectrum of vowel sound «A» in the point $\mu$ is resulted by signal/noise relation (SNR) in the input of the model equal to -13.8 dB.

For comparison of noise resistance of the considered pitch frequency detection circuit there have been carried out estimations of SNR, at which appeared errors for given circuit and also for circuits of pitch detection based on peak-picking, autocorrelation and cepstral methods. Last three methods were taken from developed by Speech Technology Center Speech Interactive Software system SIS 4.0 [12].

For an estimation of pitch detection noise resistance by the method of using a principle of auditory periphery action was used a three-step deciding rule.

1-st step - determination of the channel, where the energy of the response in reply to an effecting signal in an output of the LPF has appeared to be the greatest one.

2-nd step - determination of existence of periodic structure in envelope of the processed signal by a spectrum of its response in the channel found during the 1-st step.

The decision is made with the help of an estimation of excess of the relation of the spectral maximum to average spectral value of given threshold. This threshold is defined as mean of the same relation for clear noise signal in the input of the model.

3-d step consist in determination of meaning of vowel pitch frequency according to frequency position of the greatest component in a considered spectrum.

During every measurement session there was carried out a determination of fundamental tone frequency not less than in 10 consecutive samples of stationary vowel in pronunciation of man and woman. Duration of each sample was 100 msec. The failure was fixed in the event if at least in 20 % of measurements the received pitch frequency values differed more than in 10% from those, determined for the same sound by absence of noise.

Value of SNR at which a failure was observed was fixed as extreme allowable for the given method of vowel sounds pitch detection.

The average of extreme allowable SNR for all vowels was used as evaluation describing noise resistance of an examined method.

The results of measurements are in table 1.

<table>
<thead>
<tr>
<th>Pitch Detection Method</th>
<th>Maximal admissible S/N Ratio, dB</th>
</tr>
</thead>
<tbody>
<tr>
<td>Peak-Picking</td>
<td>+ 5</td>
</tr>
<tr>
<td>Cepstrum-based</td>
<td>+ 2</td>
</tr>
<tr>
<td>Autocorrelation-based</td>
<td>- 5</td>
</tr>
<tr>
<td>Model of auditory periphery</td>
<td>- 12</td>
</tr>
</tbody>
</table>
From data, displayed in table 1 it follows, that the considered model of processes of the analysis and processing of signals in peripheral section of hearing allows to determine vocal sounds fundamental frequency by SNR in 7dB lower, than the best of the described "technical" methods, which specifies the good perspectives of a described approach.

CONCLUSION

The existence of electromechanical conversion in inner ear structures results in the appearance of the feedback mechanisms in the cochlea both for microphonic potentials level and for summating potentials level. Because of this feedback effects combintive oscillations arise in the hydrodynamic system of cochlea. They are generated by components of input acoustical stimulus in every particular frequency range.

By several conditions due to the feedback connections through summating potentials in cochlea some filtering structures are realized, which are performing the optimal pitch harmonization enhancement and extraction. This mechanism improves the hearing selection performance for vocal speech patterns especially in noisy acoustical environment.

Detailed description of the functioning of the inner ear feedback mechanisms gives the possibility to analyze and calculate the effects of COMB-frequency-selection which has not been taken into consideration before.

Modeling of the feedback process through summating potentials allows to understand one of the reasons of the good extraction of vocal sounds by humans with normal hearing even in very noisy environment. In distinction to previous works we give the main attention not to tuning curves of hearing periphery, but to the role of the feedback through the multichannel signal envelopes in the forming of the frequency-selective performance of the human ear.

Experimental practice and calculated results show the SNR on the output of the proposed signal processing model is significantly higher than this one on the input for vowel-like stimuli. For this model the direct comparison has shown that it functions effectively when input signal SNR is in 6-8 dB less than the minimal admissible SNR for commonly used "technical" systems.

Usage of this model in hearing aids gives significant improvement of their working performance for some kinds of persons with hearing losses of the neurosensoric type [13].

REFERENCES

1. Filip M. Envelope periodicity detection, - JASA, 1969, 45, pp. 719-732


Figure 1. The chart of the proposed model of the auditory system periphery using multichannel envelope feedback.
Figure 2 Spectrum of the 'A' vowel in the point y by $\beta = 0$

Figure 3 Spectrum of the 'A' vowel in the point y by $\beta = 0.9$

Figure 4 Spectrum of the 'A' vowel in the point Z by $\beta = 0.9$