INVESTIGATION OF THE VOICE SOURCE MODELS

A.M.Smetanin

Department of Technical Cybernetics Mechanical Institute
Ustinov, USSR, 426069

ABSTRACT

An experimental research of conformity of the forming voice source impulse models to real process is carried out. The technique of the research is based on the analysis of power spectra of the actual and synthesized voice signal with using the linear prediction. The synthesis model of the vocalized speech signal accounting the influence of the voice source upon the voice canal is proposed.

INTRODUCTION

Before now the accepted notions about the speech forming process of the vocalized speech sounds proposed the excitation of the voice canal by quasiperiodic impulses of air stream such as the smooth unimodal time function. Proceeding from these notes it is obvious that the excitation of the voice canal must start near the moment of glottis opening. However, at present it is known that the start of the excitation of the voice canal coincides with the vocal chords closing moment [I]. The causes of this obvious contradiction weren't analised in literature about speech forming. In this paper the experimental researches were carried out in order to analyse real processes of the excitation of the voice canal by the voice source and build the model of synthesis of vocalized voice wave based on received results that is adequate to real speech signal forming. The carried out researches concerned only the interaction of the voice source with the voice canal and didn't discuss those processes as the loss in the canal, nasalization, radiation etc.

THE TECHNIQUE OF RESEARCH

The research method of the processes of the vocalized speech signal forming is based on analysis of power envelope spectrum of signal. Because of problem statement the speech wave analysis synchronous with the pitch was to be used, that involves application of the methods of increasing the resolution of the frequency. And at last the unstationary variant of the linear prediction was used [2]. The part of apparatus of the research complex contains mini-computer SM-4 supplied with the device of the analogous signal input. Special research software based on the program complex for processing and signal modelling included the following main modules: "visible speech" forming, calculation of the linear prediction coefficients, calculation of power spectrum by means of fast Fourier transformation, the impulses of different form excitation forming, the speech canal modelling, speech wave synthesis. The programming language is FORTRAN, the operational system is RAFOS.

A. REAL SPEECH SIGNAL

A speech signal put into computer (frequency range - 5kHz, quantification frequency - IOkHz, the number of quantification levels by amplitude - 256, signal-noisy relation hot worse than 40dB) was accumulated on the magnetic carrier of the computer. Speech material consisted of the words pronounced isolatedly by three announcer. Quasistationary parts of stressed vowels were subjected to the analysis. In calculation the prediction order was p=I2, coefficients of the linear prediction were averaged by three-four samples taken within the interval of analysis. Typical example of the evaluation of power envelope spectra is shown in Fig.I, where I - measured along the whole period of the pitch; 2 - during the closing interval; 3 - during the interval of vocal chords opening (the fragment of sound /i/ in the word "electrichestvo").

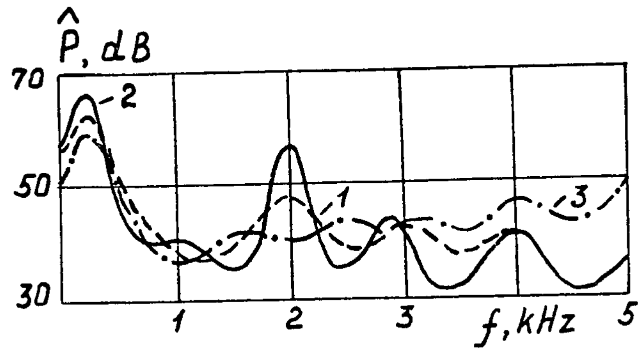

Fig.I. The power spectrum of the real signal

The experiments have shown the following [3]; I) envelope spectra during the intervals of vocal chords closing and opening are distinguished essentially by frequency and width of the spectral maximums band (formants); 2) the spectrum during the whole period of the pitch is more close to the spectrum during the interval of vocal chords closing. Thus there is sufficient variations of the voice wave parameters during the period of the pitch that is explained in general by the voice source influence on these parameters in the phase of glottis opening. The carried out calculations by the numeral evaluation of influence of the voice source on the speech canal parameters have shown that with normal conditions of pronunciation: I) the absolute changes of frequency of the formants achieve the quantity about IOOHz; 2) the absolute changes of the band width of formant-quantity about 300Hz [4].

B. THE SYNTHESIZED SIGNAL

Let us consider the model work of the synthesis of the vocalized voice signal with the excitation of the voice canal by smooth unimodal impulses. The raised cosinusoid with duration of 0,3...0,7 from the pitch period was used as the impulse. The voice canal was presented in the form of cascadely connected digital resonators corresponding to the formants [I]. In table I the parameters of the first of five formants are given, used at the synthesis of vowal /i/.

Table I

Frequency and width of the formant band

Formant	F_i, Hz	B_i, Hz
I	440	90
2	I800	50
3	2550	300
4	34I0	300
5	4400	3I0

The character of the time function of the synthesized speech signal has already shown the sufficient quality difference from real signal. The typical analysis of vowal /i/ are shown in Fig.2. The notations correspond to Fig.I. The duration of the pitch is I20 counts, the duration of the excitation impulse for Fig.2a - 40 counts, for Fig.2b - 80 counts.

Let us compare Fig.2 and Fig.3 where the spectrum of impulse reaction of voice canal with parameters of table I is given. The calculation of this spectrum has been made also by means of written technology under the excitation of canal model by the single impulse. The comparing results illustrate that cosinusoidal impulse of excitation distort the spectrum of frequency of the synthesized signal. The typical results have been obtained under the other forms of the excitation impulse triangle, in particular.

The carried out experiments and modelling say about nonadequancy of model excitation of the canal by smooth unimodal impulse. Besides, you may conclude that excitation of digital model of the voice canal by the single impulse is more close to real process of speech formation.

THE SYNTHESIS MODEL OF SPEECH SIGNAL

The carried out investigations allow to formulate two main demands to the synthesis model of vocalized speech signal: I) the excitation of the voice canal should be produced by short at the moments of vocal chords closing; 2) during the interval of the vocal chords opening the change of the voice canal parameters must be carried out, modelling, the influence of the voice source upon the speech wave parameters. The model illustrated in Fig.4 meets these requirements. It is the development of the model examined in [I]. According to the given model the voice canal in the form of a filter

$$V(z) = \frac{G}{1 - \sum_{k=1}^{p} a_k z^{-k}}$$

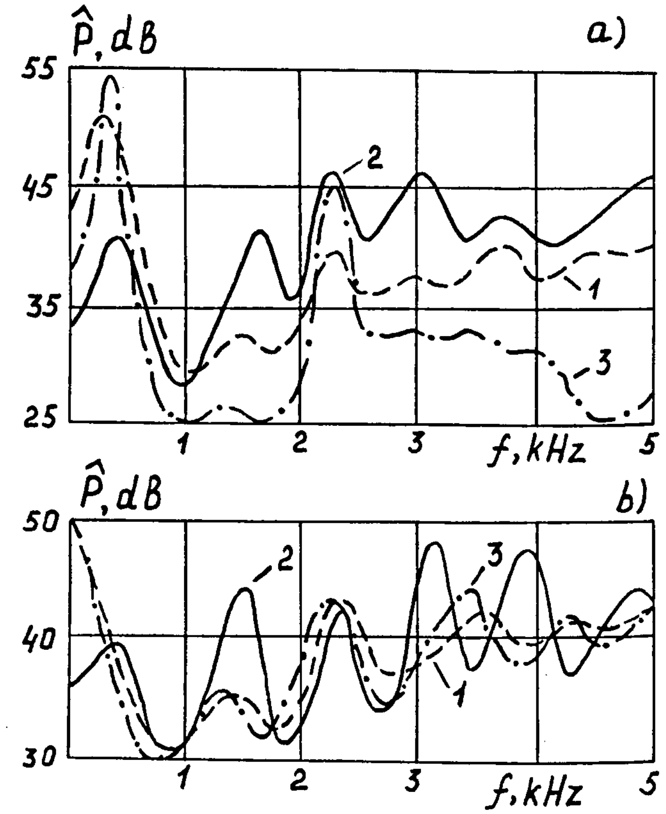

Fig.2. The power spectra of a synthesized signal with cosinusoidal source of the excitation

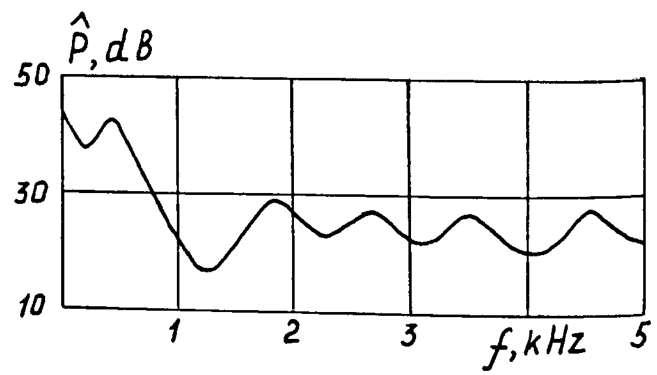

Fig.3. The impulse reaction spectrum of the voice canal

is excited by $\delta(n)$ signal at time moments t_1, t_3 etc. The parameters of the canal are changed in accordance with the function $f(n)$ during the interval of opening $[t_2, t_3]$. For the formant model, in particular, frequency F_i and width of strip B_i are these parameters counted in $V(z)$.

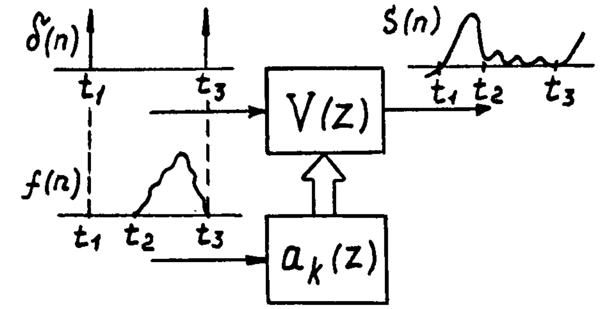

Fig.4. The synthesis model

Fig.5 illustrates the power envelope spectra synthesized by means of the suggested model signal (Fig.I). The parameters of formant during the interval of the vocal chords closing are given in table I. The duration of the pitch period was equal to I20 counts; the duration of the opening period was equal to 80 counts; the absolute changes of frequency of 5 formants during the opening interval - I00Hz; the absolute changes of the band width - 200Hz. Comparing figures I, 2, 3 and 5 we see that a given model at the spectral level is much closer to the real process of speech formation than a traditional model.

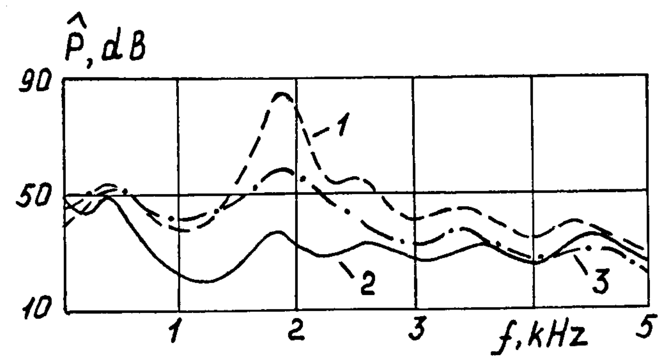

Fig.5. The power spectra synthesized by means of the suggested model signal

In the time domain the character of the synthesized speech wave is also close to real signal. It may be supposed that such model of synthesis will allow to raise naturality of synthesized speech, to model some individual peculiarities of voice. The examination of these assumptions is not included in this paper.

CONCLUSION

In consequence of carried out experimental investigations it is shown that the imagination about the excitation of the voice canal by smooth unimodal impulses of air stream are not adequate. The excitation of the voice canal by single impulses during the period of time corresponding to vocal chords closing and modelling of change of speech wave parameters as the smooth unimodal function during the phase of vocal chords opening is more close to real process of speech formation. In our opinion well known, so called, impulses of the voice source are the form of power accumulation for the next excitation of the voice canal at the moment of vocal chords closing.

REFERENCES

[I] L.R.Rabiner, R.W.Schater, "Digital processing of speech signals", New Jersey.

[2] Y.Makhoul, "Spectral analysis of speech by linear prediction", IEEE Trans. Audio and Electroacount, I973, vol.AU-2I, N3, p.I40-I48.

[3] В.Б.Гитлин, А.М.Сметанин и др., "Выбор интервала измерений частоты и ширины формант", материалы Всесоюзной школы-семинара АРСО-I0, Тбилиси, октябрь 1978, с. 20-22.

[4] В.Б.Гитлин, А.М.Сметанин, "О повышении точности измерения параметров формант", "Проблемы построения систем понимания речи", М.: Наука, I980, с.I09-II5.