HIGH QUALITY SYNTHESIS OF VOWELS

Leonid Chudnovsky, Vecheslav Ageyev

Institute for Information
Transmission Problem
Moscow, USSR 101447

ABSTRACT

A problem concerning synthesis of isolated Russian vowels is described. Approximation of excitation source functioning is at the centre of attension.

During vowel synthtsis attention is focused on vocal tract (modes) frequency values. Excitation source is approximated by a triangular function subjected to three jumps of a derivative in pitch period [1]. That approach doesn't provide a high quality synthesis and thus causes intelligibility degradation in additive noises. Somewhat better synthesis results are achieved for a more composite time function of the vowel excitation source [2].

Natural sounding and intelligibility of synthesized vowels can be improved due to taking into consideretion the real features of vowel excitation sources. One may get an idea of the excitation sources from vowel oscillograph traces using the inverse filtering techniques. To solve that inverse problem it is necessary to know such vocal tract parameters as a quality factor and moda frequency. A compensating method based on instant frequency measurement of a filtered speech signal has been used for moda frequency calculation [3]. It was continued by signal frequency filtering in order to extract formant oscillations. Low-pass filters have been used for the extraction of the first formant and band-pass filters for the extraction of other formants. The cut steepness of a filter respounse characteristic has accounted for no less than 48 dB/octave outside the transparence band. A quality factor of the extracted formant oscillations has been calculated using an analytical signal envelope [4]. Algoritm [4] has been modified to improve computing accuracy of a quality factor. After the extraction of formant oscillation $P_K(t)$ and the calculation of the quality factor Q_K and the vocal tract moda frequency ω_K it is possible to regenerate the moda excitation source from the following equation [2] :

$$P_K''(t)+\frac{\omega_K}{Q_K}P_K'(t)+\omega_K^2\left[1-\left(\frac{1}{2Q_K}\right)^2\right]P_K(t)=f_K(t) \qquad (1)$$

The excitation source of the formant oscillation $f_K(t)$ is related to the vowel excitation source $f(t)$ in the following way :

$$f_K(t)=\int_{-\infty}^{\infty}L_K(t')f(t-t')\,dt' \qquad (2)$$

where $L_K(t)$ - is a filter pulse response for extraction of the K -th formant oscillation.

Equation (1) may be used for speech synthesis as well.

Excitation sources of 5 Russion vowels "a", "э","o","y","и" have been experimentally studied. The extracted excitation sources of the first formant oscillation can be conventionally devided into two groups: the first group for the sounds "a" and "э", and the second group for the sounds "o","y" and "и". The first group of excitation sources represents two successive pulses with different amplitudes with the time interval 4-6 ms and each pulse duration 1-2 ms. The second pulse amplitude and its delay time with respect to the first pulse are related to the quality factor and the first moda frequency in such a way that the second pulse stops its free oscillations which appeared after the first pulse. The second group of excitation source is represented either by a single pulse with 1.5-2 ms duration or by two multy-or unidirectional pulses of the same duration with the second pulse time delay 1.5-2 ms, or by the three pulses of alternating direction with the duration 1.5-2 ms and the time delay 1.5-2 ms and 3-4 ms correspondingly. Excitation source of the sound "y" has one peculiarity. The regenerate excitation sources of the first moda and the extracted signal of the first formant oscillations are identical.

During vowel synthesis excitation pulses have been approximate by the following function :

$$\tilde{f}(t)=\eta(t+T)[1-\eta(t-T)]\cdot\exp\{-[1-(\tfrac{t}{T})^2]^{-1}\} \qquad (3)$$

where $2T$ -is a pulse duration;
$\eta(t)$ -is a unit function.

The synthesis resulted in high intelligibility of the vowels "a","э","o","y" when represented by single formant oscillation. Increase in the number of formant oscillation causes intelligibility improvement. For acceptable intelligibility of the synthesized vowel "и" it should be represented by two formants. The first mode excitation sources with the reduced pulse duration have been used for higher vocal tract moda frequencies (3). The duretion reduction factor for the K-th moda has been selected equal to the ratio ω_1/ω_K .

Fig.1 shows the excitation sources oscillograph traces of the first formant $\tilde{f}_{1a}(t)$, $\tilde{f}_{1э}(t)$, $\tilde{f}_{1o}(t)$, $\tilde{f}_{1y}(t)$, $\tilde{f}_{1и}(t)$ and of the second formant $\tilde{f}_{2и}(t)$ for the sounds "a","э","o","y", "и".

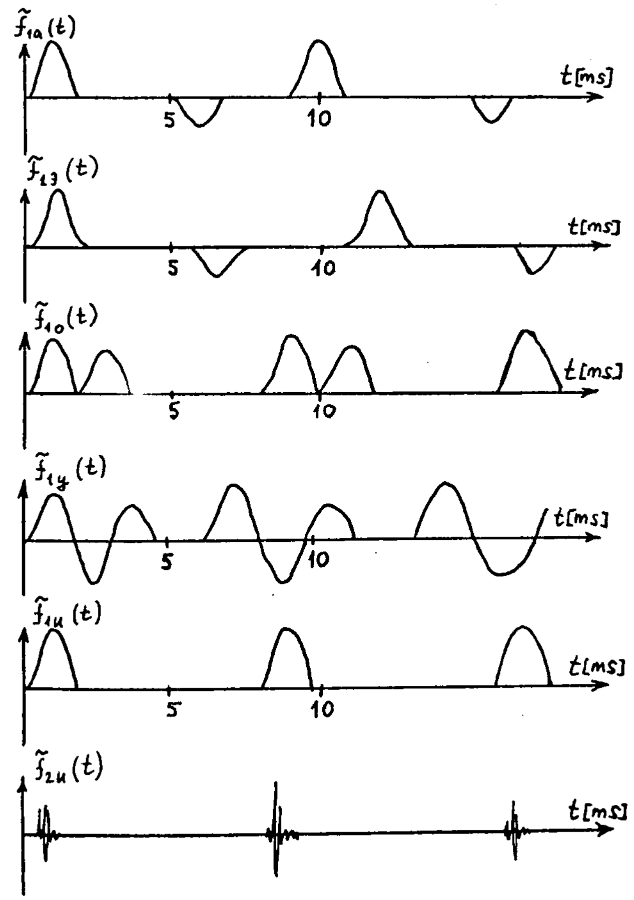

Fig. 1. The excation sources of the synthesized sounds"a","э", "o","y","и".

The oscillograph traces of the synthesized single-formant vowels "a","э","o","y" and the two-formant vowel "и" are shown in Fig.2.

Natural sounding improvemt of the synthesized vowels is achieved with due regard for time variation of the excitation source parameters of each moda $f_K(t)$. Test data analysis has shown that tne vowels excitation sources are subjected to different transformations , i.e. abrupt transformations with the time interval of 30-100 ms and slow period-by-period transformations. The vowels excitation sources which differ in their voice onset time with open or close vocal bands are well differentiated.

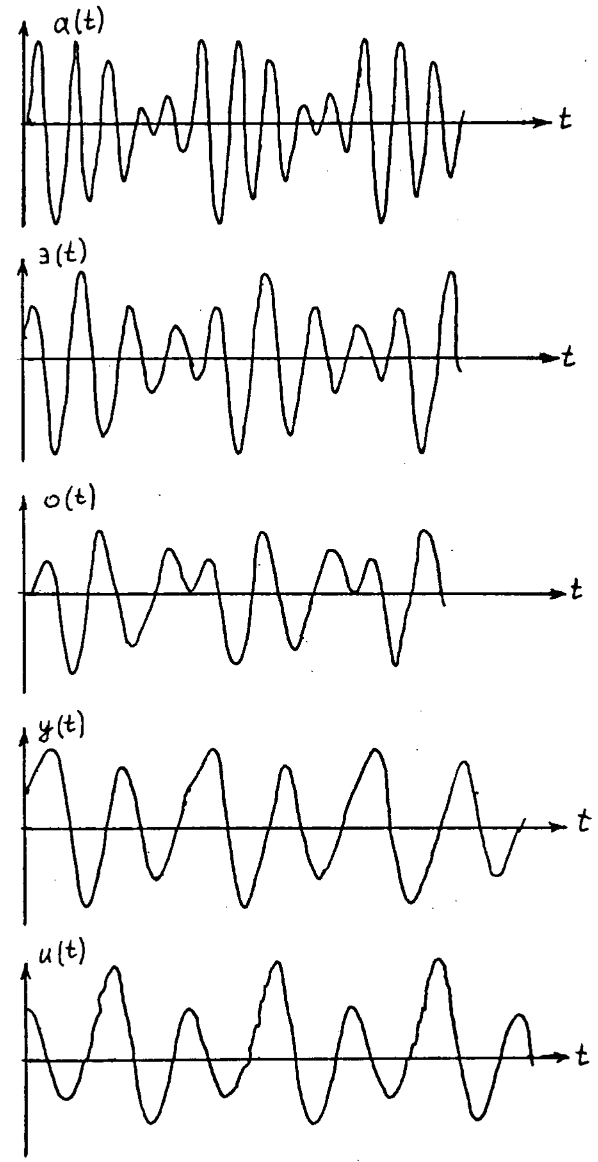

Fig. 2. The synthesized sounds "a", "э","o","y","и".

Fig.3. shows the extracted excitation sources of the first moda $f_{1a}(t)$ and $f_{1и}(t)$ of the vowls "a" and "и". The phonation of the vowel "a' initiates with close and of the vowel "и" with open vocal bands. Due to the extracted excitation source it has been found out that the voice onset time with open vocal bands and the cessation of phonation (Fig.3) have the same time structure and are practically speaker independent. To achieve the vowels high quality synthesis with due regard for the source signal variation the function has been approximated with the help of the tables.

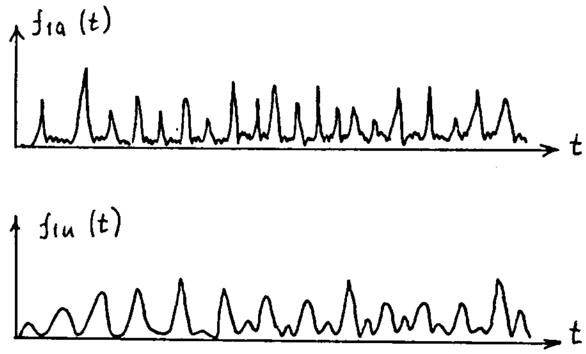

Fig. 3. The regenerated excitation sources of the first moda of the sounds "a" and "И".

The usage of excitation sources peculiarities and their relationship with vocal tract paramters gives an opportunity to achieve the high quality synthesis of vowels and speech as whole.

REFERENCES

[1] Дж.Фланаган "Анализ,синтез и восприятие речи",М.,Связь,1968.

[2] В.Н.Сорокин "Теория речеобразования",М., Радио и Связь,1985.

[3] Л.С.Чудновский, В.М.Агеев "Анализатор речевых сигналов", авторское свидетельство №1275527, МКИ G10 L 9/00,БИ №45 1986.

[4] В.С.Пичкур, А.Ф.Приставка "Определение текущих параметров частотных составляющих речевых сигналов", тезисы докладов и сообщений 13-ой Всесоюзной школы-семинара "Автоматическое распознавание слуховых образов", Новосибирск, 1984.