UDC 681.84.087.4

Zsymanski Z.
Automated Control Systems Department,
Lviv Polytechnic National University,
S.Bandery Str., 12, Lviv, 79013, Ukraine,
SWSPiZ, Sienkiewicza 9, Lodz, Poland,
E-mail: office@swspiz.pl

MINIMIZATION OF AMPLITUDE DISTORTIONS FOR TIME-SCALE COMPRESSION OF SPEECH SIGNALS

The algorithm of time-scale compression of speech signal, based on temporal reconstruction with the help of functions of temporal transforming, has been improved. Existing is signal amplitude distortions are reduced by the use of cross-correlation function and creating of intermediate composite segment.

Keywords – correlation, speech signal, time-scale compression.

1. Introduction

The analysis of existing methods and algorithms of time-scale modification of speech signals shows that the most effective are ones based of the signal modification made in time scale. Compression or expansion of signal is performed utilizing the peculiarities of the transformed speech. In [1] a new adaptive approach is developed which uses the functions of temporal transforming (FTT), describing the peculiarities of speech signals structure change while speeding or slowing of speech tempo by human speaker. This approach allows to significantly enlarge the time-scale modification factor preserving at the same time speech naturality and legibility of time compressed speech. At the same time a problem with arising noise caused by amplitude and phase distortion is not efficiently solved.

2. Problem statement

Developed in [1] FTT define the consequence of speech singal elementary segments (ES) during the process of its reduction or expansion (an example of unvoiced consonant sound's FTT is shown on Fig.1). If some ES belong to the part of the sound with the same FTT, each of them could be excluded and in the place of ES conjunction amplitude and phase distortions (click) appear. To minimize a quantity of such clicks it is recommended to exclude the neighboring segments. For example (Fig.1), if we need to shorten the duration of the sound by 3 ES, it is necessary to exclude any 3 consecutive ES with numbers from 9 till 14, because each of them has the same value of FTT. At the same time the choice of excluded ES will influence on the level of distortion and the whole process should be optimized.

There are some publications devoted to the use of cross-correlation functions to define the best point of signal conjunction during speech time-scale modification process [2,3]. Suzuki and Misaki [2] have developed an algorithm of the best conjunction point determination by maximization of cross-correlation function of two neighboring segments:

$$R(\tau) = \sum_{m=0}^{T_s-1} x(i+m+\tau)x(j+m)\,, \qquad \text{for} \qquad -S \le \tau \le S\ , \tag{1}$$

where: x(k) – input signal, i – top of a first segment, j – top of the second segment, τ - time-lag, S – ranges of search for correlation function. The time-lag T_s which maximizes the correlation function $R(\tau)$ determines the best point of signal conjunction.

A grave disadvantage of this method is an impossibility to use it in real time modification system because we do not know ahead a duration of excluded part of the initial signal. Because of that the resulting signal duration does not correspond to the duration of initial signal divided by modification factor.

So, the **goal** of this article is the modification of above described algorithm to make it possible to use it in real time time-scale modification system based on FTT.

3. The algorithm description

Let the duration of elementary segment is N , and we need to exclude M segments. The basic idea of the proposed modification is to find the speech section $(x(i), x(i+NM))$ which has the maximal correlation function for segments $x(i-2N)$ and $x(i+NM)$ with time lad equal to N. The duration of each segment is $2N$.

$$R(\tau)=\sum_{m=0}^{2N-\tau-1} x(i+m+\tau)x(j+m)\ , \tag{2}$$

where $j = I + (M+1)N$.

Conjunction of elementary segments (after the excluding of the appropriate section) is performed by the use of special composite segment:

$$y(k)=x(i-N+k)w(k)+x(j+k)(1-w(k)) \quad \text{for} \quad 0\le k\le N-1 \tag{3}$$

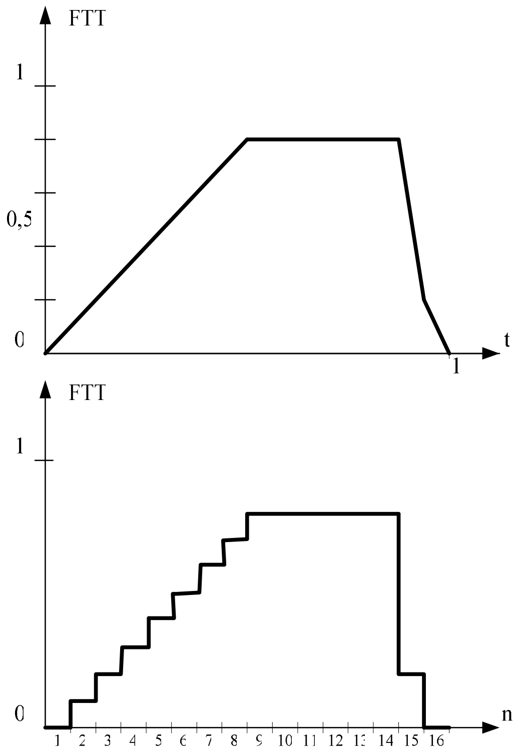
where $w(k)$ is a weighting function:

$$w(k)=\frac{N-k}{N} \qquad \text{for} \qquad 0\le k\le N-1$$

The $R(\tau)$ is maximized by all possible sections $(x(i), x(i+NM))$, which allows as a result to minimize arising amplitude and phase distortions. For the example (Fig.1), these possible sections (for *M=3*) are as follows with ES numbers: (9,10), (10,11), (11,12), (12,13), (13,14).

The equation (3) could be simplified to:

$$R(\tau)=R(N)=\sum_{m=0}^{N-1} x(i+m+N)x(j+m) \tag{4}$$



Finally, the proposed algorithm includes 3 following steps:

- Determining of all existing sections with a duration of each equal to *M-1* elementary segments, which belong to the part of speech signal with the same FTT value;
- On the base of (4) the section $(x(l), x(l+(M-1)N))$ with the maximal $R(\tau)$ for segments $x(l-1)$ and $x(l+MN)$ with the time lag N is determined;
- On the base of (3) the composite segment consisting of $x(l-1)$ and $x(l+MN)$ elementary segments is composed.

4. Conclusion

The algorithm described above was examined and tested by auditors on the real speech signal. During the experiment a duration of each ES was equal to 5 ms, the frequency of speech signal discretization was 16 kHz, a type of waiting function – triangular, time-scale modification factor varied from 1,2 till 1,5.

All auditors underlined the decrease of noise level and a better speech naturality. At the same time a legibility of time-compressed speech was the same as previous.

References

[1] Рашкевич Ю.М. Перетворення часового масштабу мовних сигналів. Львів.- Академічний експрес, 1997, 140 с.

[2] R. Suzuki, M.Misaki. Tine-scale Modification of speed signals using cross-correlation fiuctions. IEEE Transactions on Consumer Electronics, Vol. 38, No. 3, 1992, pp. 357-363.

[3] M.Covell, M. Witgott, M. Slaney. MACH1: nonuniform Time-scale modification of speech. IEEE Int. Conference on Acoustics, Speech and Signal Processing, Proceedings of the 1998 IEEE ICASSP, Vol. 6, 1998, pp. 349-352.