The investigation of gullet speech spectrum by means of the recursive filters system.

Alexander U. Kornilov(1)

(1) Tomsk State University of Control Systems and Radioelectonics, Tomsk, Russia, Kornilov@hotmail.ru

Abstract

The using of recursive filters system for speech spectrum investigation in oncological patient speech rehabilitation process, advantages and disadvantages of such system are described in this paper.

1. Introduction

Gullet speech is formed in the process of oncological patient rehabilitation, when larynx is cut out as the result of operation[3,4]. Such patients have broken connection between the lungs and vocal tract. As the result the usual mechanism of voiced speech generation by means of the system consisting the lungs, larynx, vocal tract is lost. Usually such patients are taught to speak using gullet bubble as a pressure source and gullet contraction as a source of voiced speech. Patients are taught to collect air in gullet to form the pressure source similar to the lungs. Speech therapeutists are to give the quantitative and qualitative assessment to that speech. One of methods is described in this paper.

2. Fourier transform and speech spectrum investigation

Fourier transform is usually used in speech spectrum investigation. Such method is very useful as it helps to investigate speech spectrum with good precision. In some cases this method is difficult to use. The stability of gullet voice pitch is one of the main parameters controlling during rehabilitation. The main feature of the gullet voice pitch is significant deviation from mean value unlike the normal speech. So the length of frame for Fourier transform has to be short. It results in poor frequency resolution. Strong noise component on formant frequencies is another feature of gullet speech. It makes difficult the analysis of current processes and results in useless mean spectrum.

Although there is a way which improves the gullet speech analysis and to get rid of restriction imposed on Fourier transform.

The distinctive feature of this investigation is using resonance recursive digital signal processing unlike usually used non-recursive ones. The using of recursive digital signal filters is better decision in this case as it gives the possibility to estimate speech parameters in real time as they require less processing capacity unlike non-recursive ones.

3. System of recursive filters for speech spectrum investigation

At the same time recursive filters have some disadvantages: non-linear phase response and necessity in additional stability investigations[1].

In general resonance filter have the following declaration:

\[ H(z) = \frac{1}{(\frac{z - z_1}{z - z_2})^n}, \quad (1) \]

where \( z_1, z_2 \) is a complex conjugate pair of poles , and \( 2n \) is an order of resonance filter.

Let’s estimate the filter discrimination.

\[ H(e^{j\alpha}) = \frac{1}{(e^{j\alpha} - z_1) \cdot (e^{j\alpha} - z_2)^n}, \quad (2) \]

The change of \( z_1=\rho (\cos(\alpha)+j \sin(\alpha)) \) with the following change of \( \cos(\alpha) \) on \( (e^{j\alpha}+e^{-j\alpha})/2 \) results in

\[ H(\rho, \alpha) = \frac{1}{e^{2j\alpha} \cdot (1 - \rho \cdot (e^{j\alpha} + e^{-j\alpha}) \cdot e^{j\alpha} + \rho^2 e^{-2j\alpha})^n}, \quad (3) \]

where \( \rho \) varies in range of \([0,1]\), \( \alpha \) varies in range of \([0,\pi]\) and represents itself coordinates of the pole pair in z-field. \( e^{2j\alpha} \) gives the phase shift on one period and can be ignored. Thus this representation gives fully stable 2\( n \) order recursive filter.

The parameter \( \rho \) is responsible for the filter discrimination, and \( \alpha \) is a resonance frequency.

All resonance filters have one and the same mean of amplitude-frequency response module in the pass band. In case the modules are unequal we’ll not get correct representation of instant signal spectrum on the set of such filters. Let’s get A as a model value. Then each filter should be multiplied by a coefficient \( A \) divided by the maximum value of amplitude-frequency response As amplitude-frequency response achieves its maximum on its resonance frequency, so we should find this coefficient using \( \alpha=\alpha_{\text{res}} \) in (3). It is important that we can get good filter discrimination, but less reserve of stability using \( \rho \rightarrow 1 \). The reserve of stability is as important as there are effects of finite word length...
and quantization noise. The width of the pass band of each filter is chosen in conditions of highest possible pulsations of sum of modules amplitude-frequency response of whole set. An ideal set of filters has responds $K(j\omega)=1$.

For real set of filters

$$H(j\omega) = \sum_i H_i(j\omega_i), \quad (4)$$

If frequencies take up regular positions, so

$$\omega_i = \frac{\pi}{N} \cdot i, \quad (5)$$

where $N$ is a general quantity of channels.

The following system of filters was used in gullet speech investigation:

For lower frequencies (0-500 Hz) the system consisted of 200 filters was used. The filters have the following look:

$$H(z) = \frac{1}{((z - \rho \cdot (\cos(\frac{\pi}{N}))) + j\sin(\frac{\pi}{N})) \cdot (z - \rho \cdot (\cos(\frac{\pi}{N})) - j\sin(\frac{\pi}{N})))} \quad (6)$$

where $n=2$ is a filter of the fourth order, $N$ — is quantity of filtration channels, and $\rho=0.955$ is filter discrimination.

The pass band width may be estimated according to -3 dB level:

$$0.707 = |H(e^{j\omega})| \quad (7)$$

The pass band width is $\Delta\omega=9.5$ Hz for sampling at 1 kHz. The transient band width may be estimated in the same way according to -20dB level. The transient band width is 17 Hz. The phase frequency response in the transient band and pass band changes from $-\pi$ to $-2\pi$. So the maximum phase shift doesn’t exceed $\pi$ for any filters. The pulsation of amplitude-frequency response module doesn’t exceed 1%. The estimation of oscillation setting time in filters (base on pass band width) results in 22 ms[2], so that it is possible to track the shortest phonations.

In real situation of generating such set we must consider with decrease of stability filter reserve during approaching to half sampling frequency and zero frequency. It results in necessity in truncation of the first and the last $p$ channels. The order $p$ can be defined based on root values of real constructed filters.

4. The results of applying of system of recursive filters

The pictures number 1, 2, 3 represent the result of processing of sound [i:] in three steps: healthy voice, the voice of a patient in the beginning of rehabilitation and when the rehabilitation is over.

5. Estimation of the recourses, required for spectrum investigation by means of system of recursive filters.

It should be noticed that filtration realizes in real time and doesn’t require too much calculating recourses. The most of software cannot suggest constructing of spectrum in real time as it requires fast Fourier transform (FFT) to be accomplished in every point of sample. In mentioned example calculating resources on every report for filter algorithm is $N*4=800$ multiplications of numbers with floating point. Frequency resolution in this case is 2.5 Hz.
Without taking into consideration the choice of segment length Fast Fourier transform requires at least \( N/2 \log_2 N \) complex multiplications or \( 4\left(N/2 \log_2 N\right) \) multiplications of numbers with floating point what results in 2700 multiplication. The estimation of addition operations gives significant superiority of filtration algorithm over Fast Fourier transform at the expense of great amount of addition operation in the last one. On real applications it is possible to achieve quintuple and even tenfold extension of productivity in compare with Fast Fourier transform at the expense of the fact, that the segment length of Fast Fourier transform must exceed the number of channels for border effect calculations.

The investigations showed that using of recursive filters had perspective in study of gullet speech. The constructed system already now allows estimating frequencies, instability of pitch frequency, noise- signal ratio in real time.

6. References