Noise Reducing in Speech Signals Using Wavelet Technology

Yuriy Romanyshyn, Victor Tkachenko

***Abstract* – In this paper the features of reducing of background noise in speech signals using discrete wavelet transforms with different wavelet bases, the analysis of choosing of different wavelet bases and decomposition levels of the signal are considered.**

***Index Terms* – speech signal, discrete wavelet transform, noise, wavelet bases.**

I. INTRODUCTION

THE process of recording speech signal is often accompanied by the variety of acoustic noise. Their occurrence may be associated both with poor quality of equipment and with the presence of external noise sources. For using any method of recognition of speech signals it is important the reduction of noise, because their presence can severely affect the quality of recognition. The main directions of solving this problem are spectral methods and methods based on orthogonal discrete wavelet transforms. Due to the fact that the methods of wavelet transforms are more general compared with the spectral ones and there is quite a wide selection of used wavelet bases, the features of wavelet technology for noise reducing in speech signals are considered below.

Using the wavelet transforms for speech signal processing, including for the problem of reducing noise has not only purely mathematical basis, but the biophysical one also. Based on experimental data and analysis of the signal processing it can be substantiated that the man hearing, at least during the initial stage of processing of audio signals, implements the transform, that is equivalent to some wavelet transform [1].

Primary processing of acoustic information is carried out in the inner ear (“cochlea”). Based on experiments and the following numerical simulation it was found that the response at harmonic signal $u_\omega(t) = e^{j\omega t}$ depends not only on the frequency of the signal, but also on the geometric coordinate along the cochlea. This dependence is expressed by the following relation [1]:

$$v_\omega(t, y) = e^{j\omega t}\varphi(\omega, y), \qquad (1)$$

where $\varphi(\omega, y)$ - function that depends on the frequency ω and coordinate y.

Thus, the spectral selectivity of man hearing along the coordinate is appeared, that can be interpreted as spectral characteristic of auditory channel. In the first approximation for frequencies over 500 Hz this characteristic can be approximated by the expression [1]:

$$\varphi(\omega, y) = \varphi\left(\frac{y}{y_0} - \ln\frac{\omega}{\omega_0}\right), \qquad (2)$$

where y_0 and ω_0 - normalizing coefficients.

As a result for an arbitrary signal $u_1(t)$ output signal $u_2(t, y)$ at moment t with coordinate y is determined by the expression:

$$u_2(t, y) = \omega_0 a \int_{-\infty}^{\infty} u_1(\tau)\psi(\omega_0 a(\tau - t))d\tau , \qquad (3)$$

where $a = \exp\left(\frac{y}{y_0}\right)$; ψ – some function, which depends on function φ.

This expression, up to a multiplier, corresponds to continuous wavelet transform with scale $\frac{1}{\omega_0 a}$ and time shift t.

From the computational point of view the most widespread practical application has discrete wavelet transform (DWT) as a major alternative to discrete Fourier transform. DWT is widely used in problems of digital signal processing, including processing of speech signals. Therefore, for noise reducing in speech signals the methods based on wavelet technology are used.

The purpose of this work is researching and developing the methods of noise reducing in speech signals based on wavelet technology.

Manuscript received April 20, 2011. This work was supported by EMCAT Department (Lviv Polytechnic National University).

Yuriy Romanyshyn is with the Lviv Polytechnic National University, Ukraine (e-mail: yuriy_romanyshyn@ukrpost.ua).

Victor Tkachenko is with the Lviv Polytechnic National University, Ukraine (e-mail: viktka@gmail.com).

II. Wavelet -technology in speech signal processing

Wavelet technology at various stages of processing speech signals - noise reducing, segmentation, recognition is used.

Algorithm for noise reducing (which basically was already a classic) consists of the following steps:

1) discrete wavelet transform the signal to noise;

2) threshold processing of the wavelet coefficients (with possible adaptation);

3) reproduction signal by inverse wavelet transform.

In [2] the application of wavelet transform for task segmentation of speech signals and to noise reducing in them is considered. Wavelet transform shows the signal in scale-(frequency) time domain:

$$f(t)=\sum_k \lambda_i(k)\varphi_{ik}(t)+\sum_k\sum_{j=i}^{\infty}\gamma_j(k)\psi_{jk}(t), \quad (9)$$

where $\lambda_i(k)$ – approximation coefficients; $\gamma_j(k)$ – detail coefficients; $\varphi_{ik}(t)$ – scaling function; $\psi_{jk}(t)$ – wavelet function; k – the scale; i, j – shifts.

In the speech signal on low-noise signal / noise ratio 32 dB imposed. Noise by the sounds of machinery was created. To estimate the noise level used a fragment of the speech signal with missing information component, which noise component introduced. Due to the discrete wavelet transform to noise reducing S / N ratio increased to 37 dB when using the coefficients of detail only the first level of decomposition. To experimentally Board as the best for the speech signal (sampling frequency 11 025 Hz) wavelet basis functions Daubechies 10th order was selected.

In [3] a method of improving the speech signal using wavelet transform-based operator of energy is considered. In this and some other works as a noise signal simulated additive gauss white noise is used.

In [4] to improve speech signals using the bionic wavelet transform and recurrent neural network is considered. This method can be represented by two parts. The first step is the realization of bionic wavelet transform, the second - the using of recurrent neural network to find a set of wavelet coefficients, which by noise reducing are removed.

Two methods for noise reducing from speech signal in [5] proposed. They are based on empirical mode decomposition. Different versions of the application of wavelet technology in speech signals to noise reducing in [6], [7], [8], [9] are considered. This confirms their wide application in problems of noise reducing in creating systems of recognition of speech signals.

Application of wavelet technology, combined with spectral and cepstral coefficients in automatic speech recognition in [10] are illustrated.

III. Wavelet-transform of signal in orthogonal basis

Discrete wavelet transform of signal $S[i]$ ($i=\overline{1,m}$, m – number of signal counts) using the scaling function $\varphi(t)$, that at each scale 2^j satisfies condition orthonormalization to shifts in time to $2^{-j}k$ and $2^{-j}m$ ($k,m\in Z$) carried out:

$$\int_{-\infty}^{\infty} 2^{j/2}\varphi(2^j t-k)2^{j/2}\varphi(2^j t-m)dt=\delta_{km},$$

where δ_{km} – Kronecker symbol, Z- set of integers.

In addition, the function $\varphi(t)$ satisfies the normalization condition:

$$\int_{-\infty}^{\infty}\varphi(t)dt=1.$$

With the scaling function $\varphi(t)$ bound wave function $\psi(t)$, discrete samples which are determined by function samples $\varphi(t)$ ratio:

$$\varphi[i]=(-1)^i\psi[n+1-i];\ i=\overline{1,n},$$

where the number of counts n defined by functions $\varphi(t)$ and $\psi(t)$.

Discrete counts $\tilde{\varphi}[i]=\varphi[n+1-i]$ and $\tilde{\psi}[i]=\psi[n+1-i]$ is the discrete impulse response digital filters respectively lower and upper frequencies. Reliable signal for a given discrete functions $\varphi[i]$ and $\psi[i]$ carried out in accordance with the scheme shown in Fig. 1 [1].

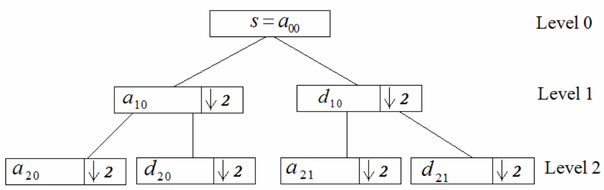

Fig. 1. Binary tree decomposition of multilevel signal

The signal sequence into a number of levels can be decomposed. At each level signal from a pool of sublevels, which correspond to the coefficients of approximation a_{jr} and detail coefficients d_{jr} (j – level number; r – number of pairs of sublevels) is generated. Each of the sublevels into two sub at a lower level can be dissected. Coefficients a_{jr} resulting digital signal filtering at the highest level of low-pass filter with impulse response $\widetilde{\varphi}[i]$, and coefficients d_{jr} – filter high-pass characteristic of $\widetilde{\psi}[i]$ followed by decimation ($\downarrow 2$).These coefficients are determined by the recurrence relations [11]:

$$a_{j+1,2r}[k] = \sqrt{2} \sum_{i=\max(1;2k+1-n)}^{\min(n;2k)} a_{jr}[i]\varphi[i+n-2k];$$

$$d_{j+1,2r}[k] = \sqrt{2} \sum_{i=\max(1;2k+1-n)}^{\min(n;2k)} a_{jr}[i]\psi[i+n-2k];$$

$$a_{j+1,2r+1}[k] = \sqrt{2} \sum_{i=\max(1;2k+1-n)}^{\min(n;2k)} d_{jr}[i]\varphi[i+n-2k];$$

$$d_{j+1,2r+1}[k] = \sqrt{2} \sum_{i=\max(1;2k+1-n)}^{\min(n;2k)} d_{jr}[i]\psi[i+n-2k];$$

$$j=0;\ r=0;\ j=1,2,...;\ r=0,1,...,2^{j-1}-1.$$

Formula for reproduction coefficients and detail coefficients at the higher level of lower-level have the form [12]:

$$a_{jr}[2k-1] = \sqrt{2} \sum_{i=k}^{k+n/2-1} (a_{j+1,2r}[i]\varphi[n+1-2i] +$$
$$+ d_{j+1,2r}[i]\psi[n+1-2i]);$$

$$a_{jr}[2k] = \sqrt{2} \sum_{i=k}^{k+n/2-1} (a_{j+1,2r}[i]\varphi[n+2-2i] +$$
$$+ d_{j+1,2r}[i]\psi[n+2-2i]);$$

$$d_{jr}[2k-1] = \sqrt{2} \sum_{i=k}^{k+n/2-1} (a_{j+1,2r+1}[i]\varphi[n+1-2i] +$$
$$+ d_{j+1,2r+1}[i]\psi[n+1-2i]);$$

$$d_{jr}[2k] = \sqrt{2} \sum_{i=k}^{k+n/2-1} (a_{j+1,2r+1}[i]\varphi[n+2-2i] +$$
$$+ d_{j+1,2r+1}[i]\psi[n+2-2i]);$$

$$k = \overline{1, m_j/2}.$$

Multilevel signal decomposition $s(t)$ in orthogonal wavelet basis (wavelet series) has the form [19]:

$$s(t) = \sum_{j=-\infty}^{\infty} v_j \varphi_j(t) + \sum_{i=0}^{\infty} \sum_{j=-\infty}^{\infty} w_j^{(i)} \psi_j^{(i)}(t),$$

where $\varphi_j(t)$ - shifted scaling functions for the initial decomposition; $\psi_j^{(i)}(t)$ – appropriate scaled (on i-th level) and shifted wavelet function; v_j and $w_j^{(i)}$ – expansion coefficients.

For digital signal $s[n]$ ($n = \overline{1,m}$, m – number of signal counts) equivalent wavelet series is discrete wavelet transform, in which is a multilevel signal decomposition with the calculation of each i-th level decomposition approximation coefficients $a_j^{(i)}$ by low-pass filter coefficients and detail $d_j^{(i)}$ using the High Pass Filter.

To calculate the coefficients of approximation and detail signals and playback schedules used for their respective functions DWT and IDWT mathematical package MATLAB [2].

IV. Noise reducing in speech signals using wavelet transforms

For the computational experiments speech signals from the database on the Internet [13], which were files with a record of different words and different speakers, were used. Noise signal components formed separately track several types of noise, which formed the basis of linguistic signals with additive noise for each reference signal various kinds of noise was in turn added.

The essence of the process of noise reducing is to schedule the speech signal on several levels, finding the approximation coefficients at the last level of detail coefficients at all levels, elimination (equating to zero) coefficients of detail levels on the scale that can meet the revised noise (usually those detail coefficients of wavelet decomposition module which is smaller than some specified threshold, and the required level and thresholds established experimentally). At the final stage of purification voice signal by inverse wavelet transform was synthesized. The effectiveness of noise reducing energy density by the difference signal, which was obtained after purification of the input signal with added noise determined and the obtained spectra and their difference in their wavelet coefficients was compared.

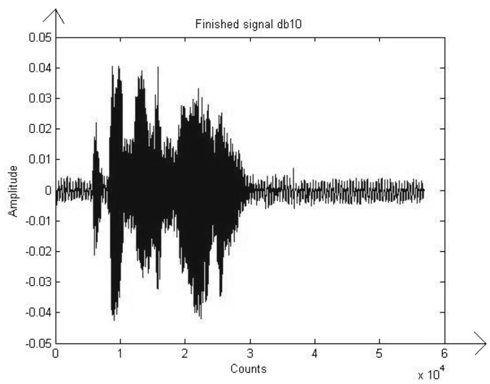

Fig. 2. The resulting signal after noise reducing by db10

Computational experiments for different signals, noises, different wavelet bases, using different levels of decomposition were conducted. In Fig. 2 an example of one of result - the Ukrainian word "married" where the added noise signal (Fig. 3) was reduced is presented.

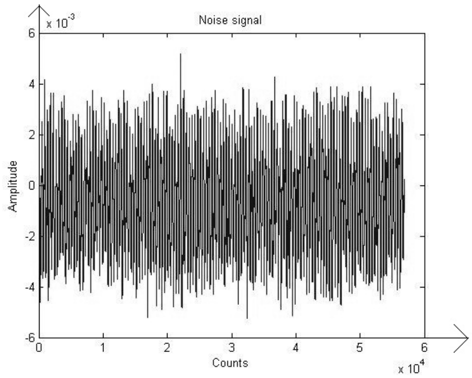

Fig 3. Interference signal

In particular, wavelet Daubechies bases order 2, 4, 6, 8, 10 was used. With their application signals cleared was received, since this was the best result that is confirmed by the obtained ratios of signal / noise ratio. Namely, the level of signal / noise ratio in noisy signals in decibels was:

TABLE I
PREDICTED SOLUTION TIME

	Db10	Db 8	Db 6	Db 4	Db 2
Word1	26.55	26.52	26.53	26.53	26.51
Word 2	50.66	50.63	50.63	50.64	50.64
Word 3	51.07	51.04	51.05	51.06	51.05
Word 4	62.67	62.64	62.65	62.66	62.66

Following algorithm procedures for processing signals (from noise to clean signal) is proposed:

1. Input signals (standard and noise).
2. Determination of the maximum level of noise signal and set the threshold based on it.
3. Adding noise signal to the reference signal.
4. Determination of the ratio signal / noise in the noised signal.
5. Schedule noisy signal obtained by Daubechies wavelet bases (for bases in turn 2, 4, 6, 8, 10).
6. Removing noise component from the signal.
7. Restoration of signal using the inverse wavelet transform.
8. Determination of the ratio signal / noise in the signal cleared.
9. Output of the results.

IV. Conclusion

During the experiments we used wavelet bases 2, 4, 6, 8, 10 of Daubechies family and the method validation results using the signal / noise ratio obtained in the process of cleaning and noise signals was proposed. It was determined that the best results in solving the problem is the use of wavelet db10.

References

[1]. Daubechies I. Ten Lectures on Wavelets // CBMS-NSF Series in Applied Mathematics. Philadelphia: SIAM Publications, 1992. 357 p.

[2] G.Dobrushkin, V.Danilov. Application of wavelet transform for noise removal and segmentation of speech signals / / Scientific news "KPI". 2010 / 2. S. 34-42.

[3] M. Bahoura, J. Rouat, Wavelet Speech Enhancement based on the Teager Energy Operator, Signal Processing Letters, vol.8, Issue: 1, pp.10-12, 2001.

[4] M. Talbi, L. Salhi, W. Barkouti and A. Cherif, Speech Enhancement with Bionic Wavelet Transform and Recurrent Neural Network, 5th International Conference: Sciences of Electronic, Technologies of Information and Telecommunications SETIT 2009 March 22-26, 2009 – TUNISIA – 9 p.

[5] K. Khaldi, A.-O. Boudraa, A. Bouchikhi and M.T.-H. Alouane, Speech Enhancement via EMD, EURASIP Journal on Advances in Signal Processing, Volume 2008, Article ID 873204, 8 p.

[6] Q. Fu, E.A. Wan, A Novel Speech Enhancement system based on Wavelet Denoising, Center of Spoken Language Understanding, OGI School of Science and Engineering at OHSU, 9 p., February 14, 2003

[7] M. Bahoura and J. Rouat, "Denoising by Wavelet Transform: Application to Speech Enhancement". Canadian Acoustics, Vol. 28, No. 3, pp 158-159, 2000.

[8] Y. Ghanbari, M.R. Karami, S.Y. Mortazavi, A New Speech Enhancement System Based on the Adaptive Thresholding of Wavelet Packets, 13th ICEE2005, Vol. 1, Zanjan, Iran, May 10-12, 2005, 6 p.

[9] A.V. Lastochkin, V.Yu. Kobelev, The Denoising Method Based on the Wavelet Processing Adapted for Sharp Signals, DSPA-2000, 2 p.

[10] M.C.A. Korba, D. Messadeg, R. Djemili, H. Bourouba, Robust Speech Recognition Using Perceptual Wavelet Denoising and Mel-frequency Product Spectrum Cepstral Coefficient Features, Informatica 32 (2008), 283-288.

[11]. Yu Romanyshyn, W. Gudym. Compression of speech signal based on discrete wave transformations / Radioelectronics and Telecommunications. Bulletin of the National University "Lviv Polytechnic", № 428. - Lviv, 2001. - S. 22-27.

[12] A. Pereberin. About the systematization of wavelet transforms / / Computational Methods and Programming. - 2001. - T. 2. - S. 15-40.

[13]. http://www.speech.com.ua/russian.html.

Victor Tkachenko was born in L'viv, Ukraine, in September 9 1986. In 2008 got Master's degree of computer science at Lviv Polytechnic National Univercity, Ukraine.

At 2008 become PhD student at Lviv Polytechnic National University at EMCAT department. The field of study is speech recognition task.

Publications: V.Tkachenko, Yu.Romanyshyn. Noise reducing in speech signals using wavelet technology // IEEE CADSM'2011. – Polyana, 2011. – P. 446; V. Pavlysh, Yu. Romanyshyn, V. Tkachenko. Software tools of construction, training and using of hidden Markov models in MATLAB system. Proceedings of the IXth International Conference CADSM'2009. - Lviv-Polyana: Publishing House of the Lviv Polytechnic National University, 2009. - P.125; V. Pavlysh, Yu. Romanyshyn, V. Tkachenko. Preliminary segmentation of speech signals for the tasks of their recognition. //IEEE MEMSTECH'2009. - Polyana, 2009. - P.144