ISSN 1063-7850, Technical Physics Letters, 2016, Vol. 42, No. 8, pp. 865–867. © Pleiades Publishing, Ltd., 2016.
Original Russian Text © A.S. Yasin, O.N. Pavlova, A.N. Pavlov, 2016, published in Pis'ma v Zhurnal Tekhnicheskoi Fiziki, 2016, Vol. 42, No. 16, pp. 72–78.

Speech Signal Filtration Using Double-Density Dual-Tree Complex Wavelet Transform

A. S. Yasin[a, b], O. N. Pavlova[a], and A. N. Pavlov[a, c, d]*

[a] *Saratov State University, Saratov, 410012 Russia*
[b] *University of Technology, Baghdad, Iraq*
[c] *Yuri Gagarin State Technical University of Saratov, Saratov, 410054 Russia*
[d] *Kotelnikov Institute of Radio Engineering and Electronics (Saratov Branch), Saratov, 410019 Russia*
**e-mail: pavlov.alexeyn@gmail.com*

Received March 23, 2016

Abstract—We consider the task of increasing the quality of speech signal cleaning from additive noise by means of double-density dual-tree complex wavelet transform (DDCWT) as compared to the standard method of wavelet filtration based on a multiscale analysis using discrete wavelet transform (DWT) with real basis set functions such as Daubechies wavelets. It is shown that the use of DDCWT instead of DWT provides a significant increase in the mean opinion score (MOS) rating at a high additive noise and makes it possible to reduce the number of expansion levels for the subsequent correction of wavelet coefficients.

DOI: 10.1134/S1063785016080290

Wavelet filtration of noisy signals and images is a widely used method of communicated data cleaning from random fluctuations [1–3]. Traditional schemes of filtration based on a multiscale analysis using discrete wavelet transform (DWT) are characterized by high speed and have relatively simple program implementation [4]. However, they have some limitations that can be eliminated by methods of complex wavelet transform [5–9] employing basis set functions with real and imaginary parts related through Hilbert transform. A number of works in this field are based on the method of dual-tree complex wavelet transform (DTCWT) [5–7], which demonstrated high-efficiency filtration of noisy images. This method employs orthonormalized basis sets, which provides high speed and the absence of excess expansion procedures. At the same time, the use of critical sampling makes DTCWT dependent on the correction of informative wavelet coefficients, which may be accompanied by distortion of the reconstructed informative signals. For this reason, filtration of a signal transferred via communication channel is frequently performed using frames representing nonorthonormalized (excess) basis sets. At the expense of some increase in the time of data processing (which can, nevertheless, be performed online), this method offers the possibility of retaining the necessary signal information in the case of removal of "essential" wavelet coefficients or when the direct expansion is performed with insufficient accuracy because of the presence of fluctuations.

At present, much research effort is being devoted to the creation of combined algorithms for cleaning informative communications from noise and random distortions, which are based on various methods of wavelet filtration. In recent years, promising solutions have been proposed based on double-density DTCWT (DDCWT) [8, 9]. This approach employs a much more complicated algorithm, representing in fact a qualitatively new level of solving tasks of digital filtration of signals and images. The present work is devoted to studying the possibilities of using DDCWT for speech signal filtration. It is shown that this method reduces the number of expansion levels in the basis set of wavelet functions as compared to traditional DWT so as to reach maximum quality of cleaning informative communications from noise.

The DDCWT method stipulates signal expansion using a single complex scaling function and two complex wavelet functions with real and imaginary parts related through Hilbert transform. These functions obey the same conditions as those for Daubechies wavelets used in the framework of traditional DWT [10–12]. In particular, the real parts of these functions obey the following relations:

$$\varphi(t) = \sqrt{2}\sum_n h_0(n)\varphi(2t-n),$$

$$\psi_{1,2}(t) = \sqrt{2}\sum_n h_{1,2}(n)\psi_{1,2}(2t-n), \quad h_2(n) = h_1(n-1).$$

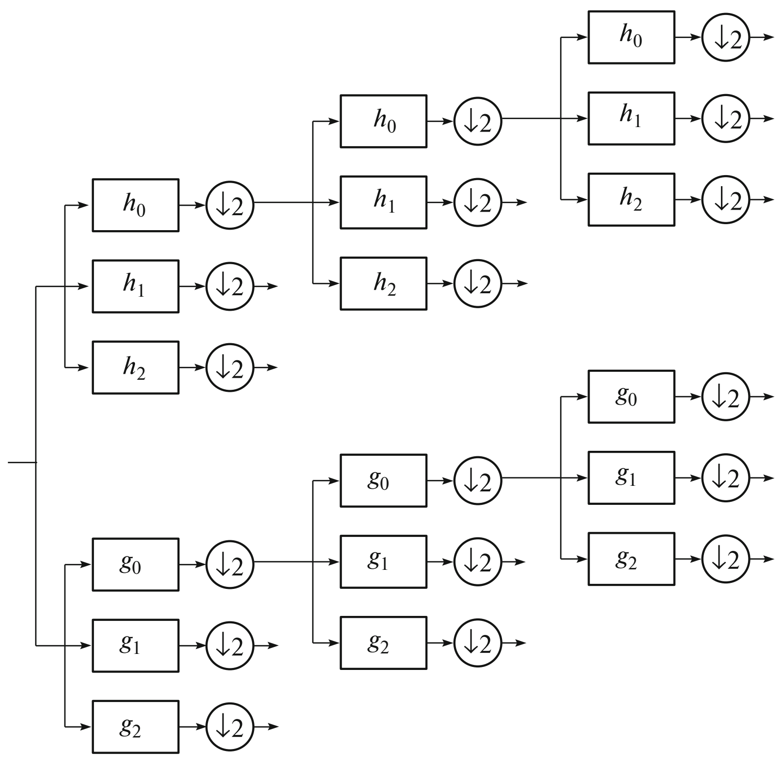

Fig. 1. General scheme of signal expansion in the framework of DDCWT method (three levels of expansion are presented). Filters h_0 and g_0 correspond to real and imaginary parts of the scaling function; filters $h_{1,\,2}$ and $g_{1,\,2}$ correspond to real and imaginary parts of wavelet functions.

Analogous relations are written for the imaginary parts. With this choice of ψ_i filters, a new level of resolution is used for gapping of only the approximating coefficients (i.e., coefficients of expansion in scaling functions φ), whereas the refining coefficients (coefficients of expansion in wavelet functions ψ_i) are completely retained.

Figure 1 shows the general scheme of signal expansion in the framework of DDCWT method. In the present work, we have employed filters developed by Selesnick [8] and compared the quality of noisy speech signal filtration achieved using DWT and DDCWT methods.

Quantitative assessment of the quality of speech signal filtration was based on the mean opinion score (MOS) rating [13], which uses a five-point grading scale to characterize the operation of a communication system used for perception of conversation or spoken material. By definition, an MOS stipulates the presence of a large number of subjective estimates of the system quality, followed by their averaging. In practice, MOS data are approximated using the PESO (perceptual evaluation of speech quality) model developed for obtaining objective MOS ratings in accordance with commonly accepted standards of the International Telecommunication Union (ITU) [14, 15].

To compare the efficiency of speech signal filtration by DWT and DDCWT, we have used samples of spoken communications (Fig. 2a) with imposed adaptive noise of varying intensity and statistics. Each sample was first subjected to multiscale analysis based on DWT, and the corresponding MOS estimates were obtained for several Daubechies wavelet families (from D^4 to D^{20}) and two variants (hard and soft) of threshold function setting for the subsequent correction of wavelet coefficients. Calculations were performed with wavelet coefficients corrected on various levels of expansion. As a result, the optimum level of expansion and the wavelet basis set were determined that provided the best quality of spoken communication filtering from additive noise.

It was established that the soft variant of threshold function setting for the correction of wavelet coefficients provides for a significantly higher MOS rating (increased by up to 50% as compared to the hard threshold setting at a signal-to-noise ratio of 0 dB). This is related to the fact that discontinuities of threshold function in the hard variant lead to more significant distortions of the reconstructed signal. The choice of optimum basis set depends on the signal analyzed, but it expedient (for obtaining higher MOS ratings) to use relatively smooth wavelets with greater domains (D^{10}–D^{20}). In addition, the quality of filtration depends on the number of expansion levels. For the example presented in Fig. 2a, the MOS value obtained using two to three levels of expansion exceeds

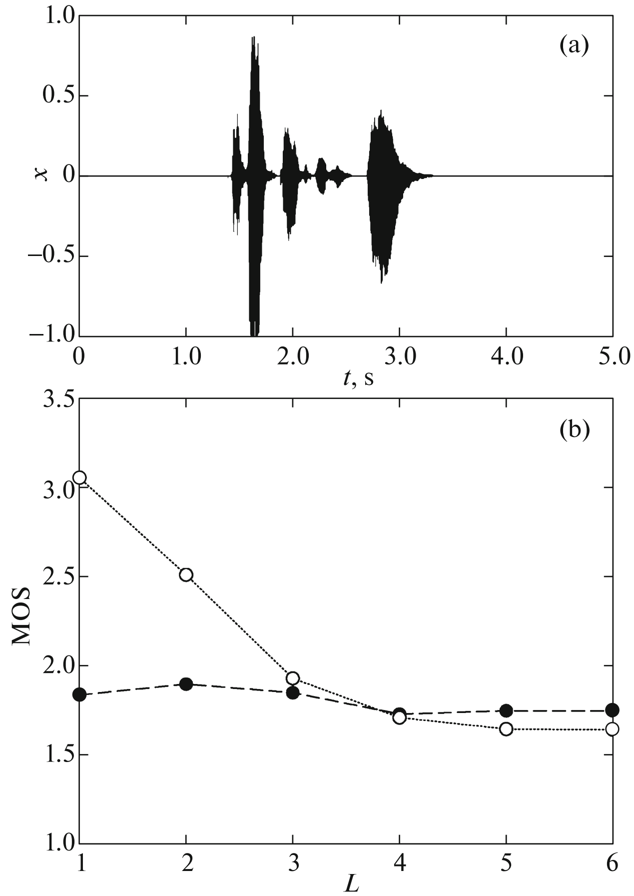

Fig. 2. (a) Example of the analyzed noisy spoken test signal (short phrase "Hello, how do you do") and (b) results of the wavelet filtration of added noise by methods using DWT (black circles) and DDCWT (open circles).

the value obtained with a single level (Fig. 2b, black circles). Similar results were obtained with other examples of experimental data. It should be noted that the maximum MOS value for the DWT method in Fig. 2b is about 1.9, which is indicative of a low quality of speech signal filtration at the given noise intensity.

Analogous calculations were performed for DDCWT method. In the given case, a single variant of wavelet basis set [8] was used and the quality of filtration was determined for various numbers of expansion levels of the noisy speech signal. Figure 2b (open circles) shows a typical plot of the MOS value versus number of expansion levels. In contrast to the case of DWT filtration, the maximum MOS value achieved with DDCWT amounts to 3.1, which is indicative of significant increase in this characteristic of filtration quality. In other examples, the increase in MOS rating was also significant and showed evidence of unambiguous advantage of the DDCWT method. An important circumstance is that, in the given example, the MOS maximum is already attained on the first level of expansion, rather than for two to three levels as in the case of DWT. This circumstance was also confirmed in other examples. The number of the level on which the maximum MOS rating was achieved varied depending on the noise intensity, but in all cases the maximum for DDCWT method was achieved for a lower expansion level as compared to that in the case of DWT. This circumstance allows the number of expansion levels for wavelet filtration to be reduced, thus partly compensating for increased time consumed at the stage of information cleaning from noise with the aid of excessive basis sets and complex wavelet functions.

Acknowledgments. This work was supported in part by the Russian Science Foundation, project no. 14-12-00324.

REFERENCES

1. Y. Meyer, *Wavelets: Algorithms and Applications* (SIAM, Philadelphia, 1993).
2. M. Jansen, *Noise Reduction by Wavelet Thresholding* (Springer-Verlag, New York, 2001).
3. D. L. Donoho, IEEE Trans. Inform. Theory **41**, 613 (1995).
4. W. H. Press, S. A. Teukokolsky, W. T. Vetterling, and B. P. Flanney, *Numerical Recipes in C: The Art of Scientific Computing* (Cambridge University Press, Cambridge. 1992).
5. N. G. Kingsbury, Appl. Comput. Harmon. Anal. **10**, 234 (2001).
6. I. W. Selesnick, R. G. Baraniuk, and N. G. Kingsbury, IEEE Signal Process. Mag. **22**, 123 (2005).
7. I. W. Selesnick, IEEE Signal Process. Lett. **8**, 170 (2001).
8. I. W. Selesnick, IEEE Trans. Signal Process. **52** (5), 1304 (2004).
9. D. Bhonsle and S. Dewangan, Int. J. Sci. Res. Publ. **2** (7), 1 (2012).
10. I. Daubechies, *Ten Lectures on Wavelets* (SIAM, Philadelphia, 1992).
11. I. M. Dremin, O. V. Ivanov, and V. A. Nechitailo, Phys. Usp. **44**, 447 (2001).
12. A. N. Pavlov, A. E. Hramov, A. A. Koronovskii, E. Yu. Sitnikova, V. A. Makarov and A. A. Ovchinnikov, Phys. Usp. **55**, 845 (2012).
13. P. C. Loizou, *Speech Enhancement: Theory and Practice*, 2nd ed. (CRC Press, Boca Raton, 2013).
14. J. G. Beerends and J. A. Stemerdink, J. Audio Eng. Soc. **40** (12), 963 (1992).
15. T. Thiede, W. C. Treurniet, R. Bitto, C. Schmidmer, T. Sporer, J. G. Beerends, C. Colomes, M. Keyhl, G. Stoll, K. Brandenburg, and B. Feiten, J. Audio Eng. Soc. **48** (1/2), 3 (2000).

Translated by P. Pozdeev