VPS VIBRATIONS IN PHYSICAL SYSTEMS
DOI: 10.21008/j.0860-6897.2023.2.18
Vibrations in Physical Systems, 2023, 34(2), 2023218

Optimal mother wavelet selection for a stochastic resynthesis of the sound textures

Wojciech WRONKA

AGH University of Krakow, al. Mickiewicza 30, Cracow, Poland

Corresponding author: Wojciech WRONKA, wronka@agh.edu.pl

Abstract Continuous wavelet transform is a powerful and versatile tool for signal analysis, outperforming short-time Fourier transform in the task of non-stationary, transient signals analysis. However, the method's performance is heavily influenced by the choice of a mother wavelet function, which is most often made by the experience-supported intuition, followed by the trial-and-error procedure. Numerous attempts to optimize the problem are not universal by any means, as its solution is determined by a particular application, acquired data, and other system requirements. One very specific example is wavelet-based statistical analysis, performed for the needs of the stochastic resynthesis of sound textures, which requires minimal decomposition and precise time localization of the individual acoustic events, components of a complex texture. This work presents the automated mother wavelet function optimization system, which performs the optimal selection based on the reference audio signal. The algorithm iterates through a wide set of commonly used functions and compares the wavelet packet decomposition trees in search of the single node, containing the most information possible, with the use of the entropy-based criterion. After performing the procedure, reference signal is resynthesized with coefficients of the selected wavelet function and then calculation of normalized root mean square error serves as a verification of the results. Conclusions contain both the advantages and the limitations of the proposed solution together with the possible improvements and the directions of future research.

Keywords: sound textures, wavelet packet transform, optimal basis selection, stochastic resynthesis.

1. Introduction

While there isn't a universal and accepted definition of sound texture, most sources seem to agree on their basic properties. They are usually described as a complex sound, a superposition of a multitude of shorter acoustic events, similar to each other, with somewhat random, irregular organization in time. Importantly, sound textures exhibit so-called temporal homogeneity, which makes them recognizable with time-averaged statistics [1]. The consequence of this property is presented in Figure 1, which illustrates gradually gained statistical knowledge about the given texture.

Amongst the most popular examples of such sounds are the raindrops hitting the surface, the crackling fire or the chirping crickets, but also the engine roar, thermal noise in electronic circuits or the complex, noisy urban soundscape. Textures can perform different functions; they often carry important information about the surroundings, and exposition to such sounds can have a substantial influence, positive or negative, on concentration or well-being [3]. Some of the most important applications of sound texture analysis and synthesis are noise-reduction techniques, active noise cancellation, speech recognition and synthesis, also mechanical fault [4] and disease diagnostics [5]. What is more, texture synthesis is also used for sound design in movies and games, as well as for creating virtual music instruments [6].

Given the transient and non-stationary characteristics of sound textures, known shortcomings of a traditional Fourier analysis applied to their analysis become evident. In its basic form, the transform decomposes signals into linear superpositions of stationary and infinite periodic functions, which is the opposite of non-stationary, stochastic sound textures, containing many, irregularly distributed transients. An efficient way to resolve the issue is to analyze the signal with the continuous wavelet transform (CWT). The method performs the decomposition with a compactly supported mother wavelet function, which is being scaled and dilated in the time domain. Calculation of the correlation between the wavelet function and the analyzed signal, with regard to the time and scale, creates a scalogram, a detailed representation of the signal's local similarity to the given wavelet function. In effect, wavelet transform offers a more natural and flexible compromise of the time-frequency resolution, than in the case of the short-time Fourier

 DOI: 10.21008/j.0860-6897.2023.2.18

transform (STFT). Finally, the choice of the continuous wavelet transform basis is arbitrary, making it a versatile and highly configurable tool.

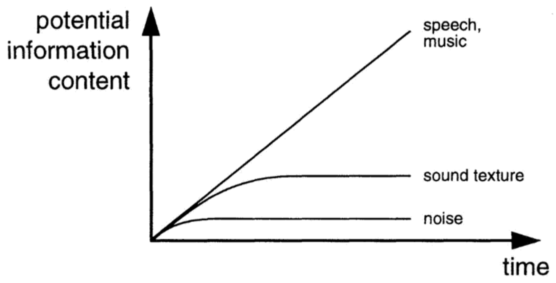

Figure 1. Illustrative graph of the perceptive information content in noise, sound textures and speech or music [2], with regard to the duration of the stimuli.

On the other hand, this advantage of nearly unrestricted choice of decomposition basis might be a serious difficulty, as there isn't a universal solution to the problem. Several previous approaches to the problem focused on either examining parametrized mother wavelet performance [7], maximizing the sparsity measure [8] using genetic algorithm (GA) or comparing a wavelet functions performance in a very specific task [9].

This work presents a specialized approach to the optimal mother wavelet selection for the task of stochastic resynthesis of sound textures and is an essential preliminary step of a statistical analysis of the signals to be resynthesized. It is based on a reference signal decomposition into a wavelet packet tree and maximization of the information contained in a single node, using the entropy criterion [10]. The algorithm iterates through a wide set of commonly used wavelet functions and the method's performance is compared for several types of real-world, recorded sound textures.

2. Method

Developed optimization procedure searches for the best wavelet function to offer minimal decomposition. It is based on a discrete method of the wavelet packet transform [11], which decomposes the signal into subsets of wavelet function coefficients by repeatedly splitting the signal into a low and a high-pass component at each level of the transform. The result can be presented in a form of a decomposition tree, as shown in Figure 2a. Each node of the tree represents the coefficients vector and each of the terminal nodes' wavelet functions form an orthogonal basis.

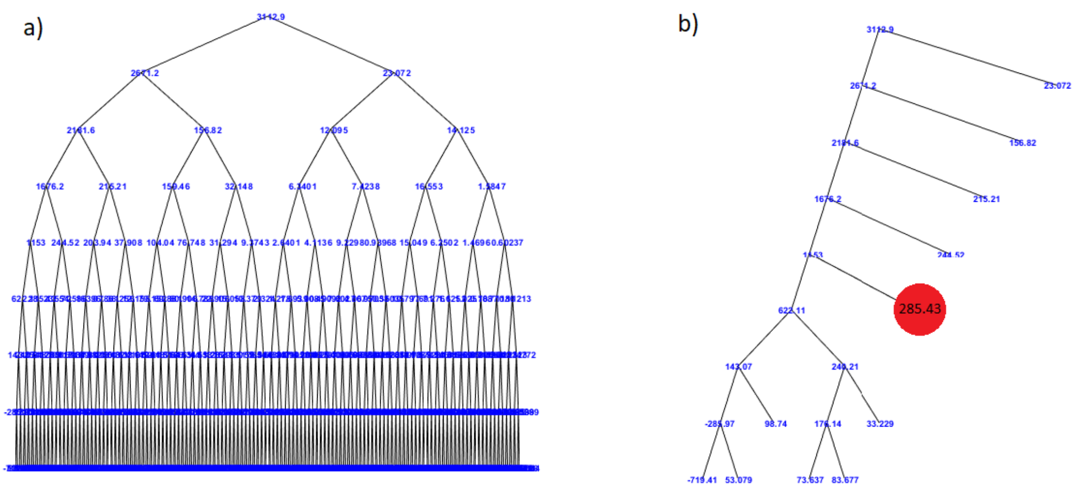

Figure 2. Wavelet packet tree (a) and its reduced, optimal form (b). The red dot indicates the node with the highest Shannon entropy value.

Then, the wavelet packet tree can be reduced to form an optimal decomposition basis by merging the children nodes containing more information than its parent. It involves comparing the individual nodes with the criterion based on a Shannon entropy formula:

$$H(X) = -\sum_i P(X = x_i)\ln\big(P(X = x_i)\big), \tag{1}$$

where each node's wavelet coefficient vector is treated as an empirical probability distribution of a random variable X. Remaining tree nodes form an optimal decomposition basis (Fig. 2b), with wavelet functions of different scales.

The algorithm iterates through the given set of the wavelet and scaling function pairs, related to the high and low-frequency components. It compares the maximal relative entropy of the terminal nodes of each decomposition. As a result, the best node and the corresponding wavelet function are selected, which provide the most information about the signal structure. Table 1 depicts the wavelet families used for optimization, together with their brief descriptions. All of the used wavelet functions, 62 in total, have the important property of being orthogonal or biorthogonal and a compact support.

Table 1. Summary of the wavelet functions used in the optimization procedure.

Family name	Description	Wavelet functions used
Haar	Simplest and oldest wavelet, built of quantised steps. Symmetric and useful for detecting non-continuity.	'haar'
Daubechies	Extremal phase and highest number of vanishing moments for a given support width. Associated scaling filters are minimum-phase filters. Parametrized with the number of vanishing moments.	'db1', 'db2', 'db3', 'db4', 'db5', 'db6', 'db7', 'db8', 'db9', 'db10'
Symlet	Least asymmetry and highest number of vanishing moments for a given support width. Associated scaling filters are near linear-phase filters. Parametrized with the number of vanishing moments	'sym2', 'sym3', 'sym4', 'sym5', 'sym6', 'sym7', 'sym8', 'sym9', 'sym10'
Coiflet	Highest number of vanishing moments for both the scaling and the wavelet function for a given support width. Parametrized with the number of vanishing moments.	'coif1', 'coif2', 'coif3', 'coif4', 'coif5'
Biorthogonal	Biorthogonal spline wavelets with linear phase, for which symmetry and exact reconstruction are possible with FIR filters. Parametrized by the number of vanishing moments for the reconstruction and decomposition filters respectively.	'bior1.1', 'bior1.3', 'bior1.5', 'bior2.2', 'bior2.4', 'bior2.6', 'bior2.8', 'bior3.1', 'bior3.3', 'bior3.5', 'bior3.7', 'bior3.9', 'bior4.4', 'bior5.5', 'bior6.8'
Reverse biorthogonal	Biorthogonal spline wavelets with linear phase, for which symmetry and exact reconstruction are possible with FIR filters. Parametrized by the number of vanishing moments for the reconstruction and decomposition filters respectively.	'rbio1.1', 'rbio1.3', 'rbio1.5', 'rbio2.2', 'rbio2.4', 'rbio2.6', 'rbio2.8', 'rbio3.1', 'rbio3.3', 'rbio3.5', 'rbio3.7', 'rbio3.9', 'rbio4.4', 'rbio5.5', 'rbio6.8'
Discrete Meyer	FIR based approximation of the Meyer wavelet, infinitely regular orthogonal wavelet.	'dmey'
Fejer-Korovkin	Minimal difference between a valid scaling filter and the ideal sinc lowpass filter, especially useful in discrete (decimated and undecimated) wavelet packet transforms. Parametrized with the number of filter coefficients.	'fk4', 'fk6', 'fk8', 'fk14', 'fk18', 'fk22'

3. Results

The set of reference signals used to test the optimization procedure comprises 5 audio files, with a sampling frequency of 48 kHz and a bit depth of 24. All files are monophonic, and their length was trimmed to 131072 samples, which equals the 17th power of two, as required by the wavelet packet transform. The length of the audio file corresponds to about 2.73 seconds of recording, which proved enough to ensure similar results with different realizations of the same sound texture.

After performing the optimization, the input signal can be reconstructed using the selected wavelet function and the vector of coefficients. This allows not only to audibly and visually inspect the result but also to measure the optimization performance with a root mean square error normalized by the root mean square of a reference signal (NRMSE). Figures 3 – 8 contain the results in form of the reference and reconstructed signal's waveform, the chosen wavelet function, and the relative entropy of all terminal nodes, which helps in the assessment of how the information is spread across different nodes of the decomposition tree.

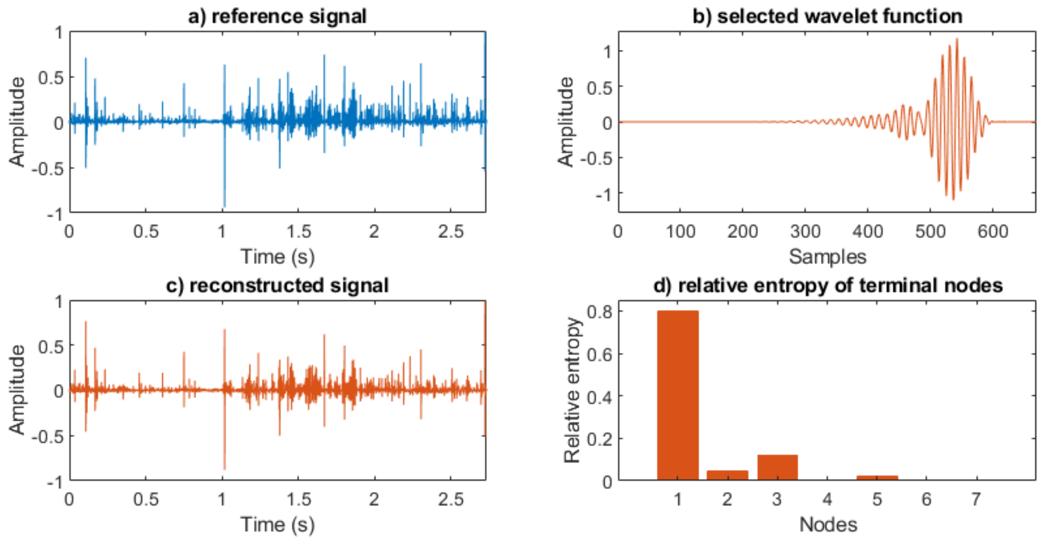

Figure 3. Selected wavelet ('fk22' in the node (3,0)) and the reconstruction of the vinyl record's noise. The relative entropy of the best node is 0.8018 and the reconstruction NRMSE is 0.3066.

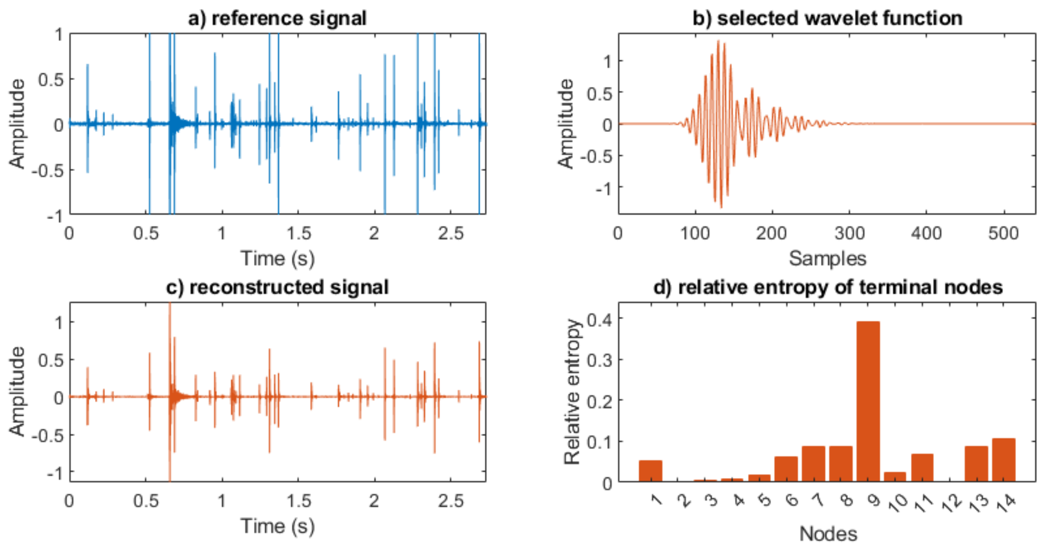

Figure 4. Selected wavelet ('db9' in the node (2,1)) and the reconstruction of the sound of the crackling fire. The relative entropy of the best node is 0.3911 and the NRMSE is 0.7195.

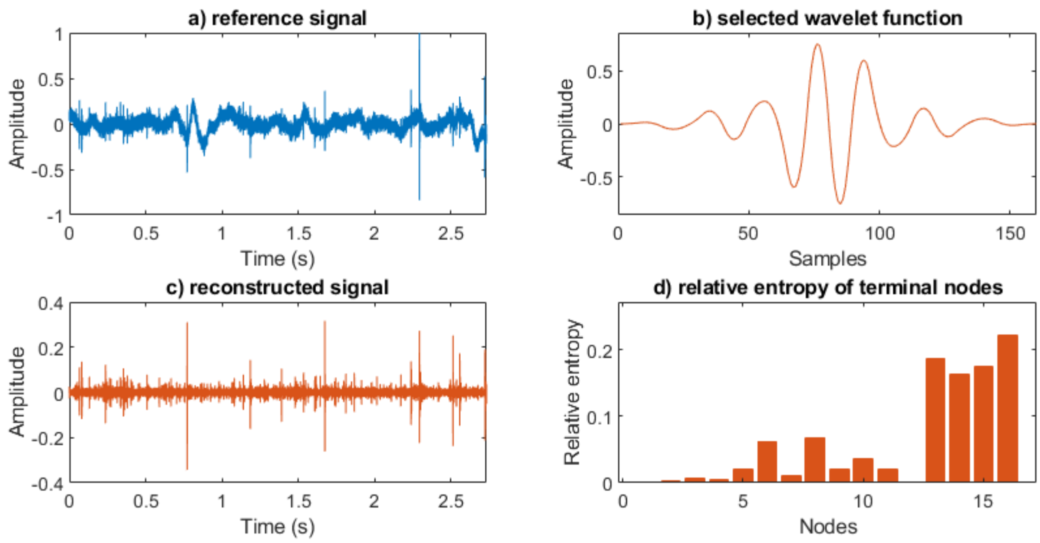

Figure 5. Selected wavelet ('rbio3.3' in the node (1,1)) and the reconstruction of the sound of the rain. The relative entropy of the best node is 0.2217 and the reconstruction NRMSE is 0.9958.

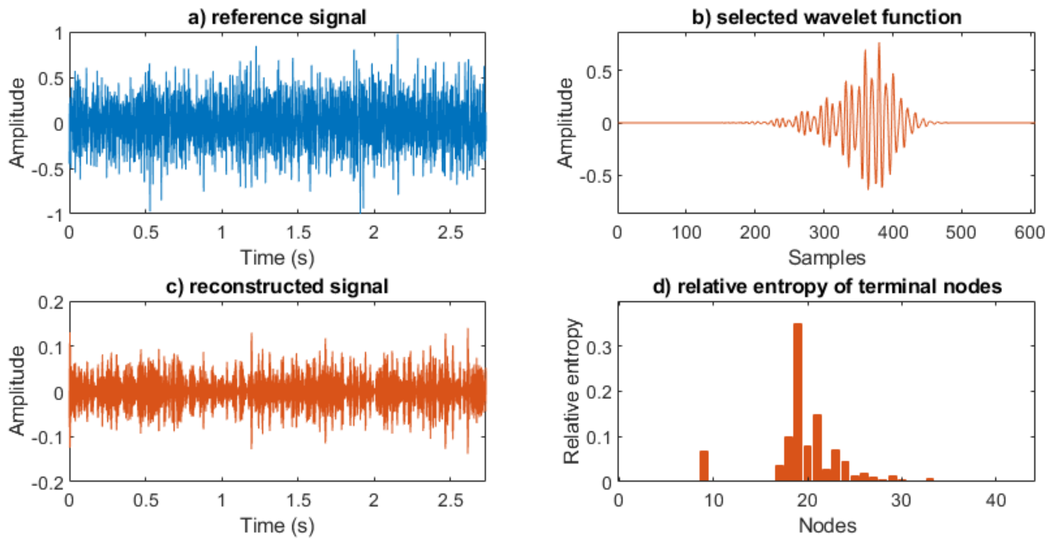

Figure 6. Selected wavelet ('db10' in the node (8,6)) and the reconstruction of the constant hum inside an airplane. The relative entropy of the best node is 0.3498 and the reconstruction NRMSE is 0.9998.

The accuracy of the chosen solution cannot be easily estimated, but it definitely should be based on multiple measures. Comparison of the signals' waveforms allows for an intuitive comparison, however, it certainly lacks important details. A deeper insight could be gained by the analysis of the entropy distribution between terminal nodes. Many peaks, which are comparable in height (i.e. Fig. 5d), suggest similarly important, but disjoined components of the signal and undermine the reasonability of choosing an individual wavelet function. On the other hand, one, distinctive peak in the entropy distribution confirms the quality of the selection.

However, even with the obvious candidate for the best-suited wavelet function, the reconstructed signal and the reconstruction NRMSE don't necessarily reflect such situation. It is crucial to emphasize, that the signal is re-synthesized with wavelets of only one scale, which means inevitable discrepancies with the original, complex sound texture. For example, the algorithm neglects the low-frequency components of the signal, especially observed in reference signals' waveforms in Figs. 5 and 8, as it primarily searches for the best shape of the wavelet function.

The general observation is that the sounds with sparser time distribution and more pronounced transients (such as the sound of fire or vinyl record noise) were assigned wavelet functions more successfully. This is equally visible in the chosen node entropy and the NRMSE of the reconstruction, as well as in the reconstructed sound file. The most difficult textures are composite sounds, with the decomposition tree containing many terminal nodes, such as the sound of the city bus (Fig. 8).

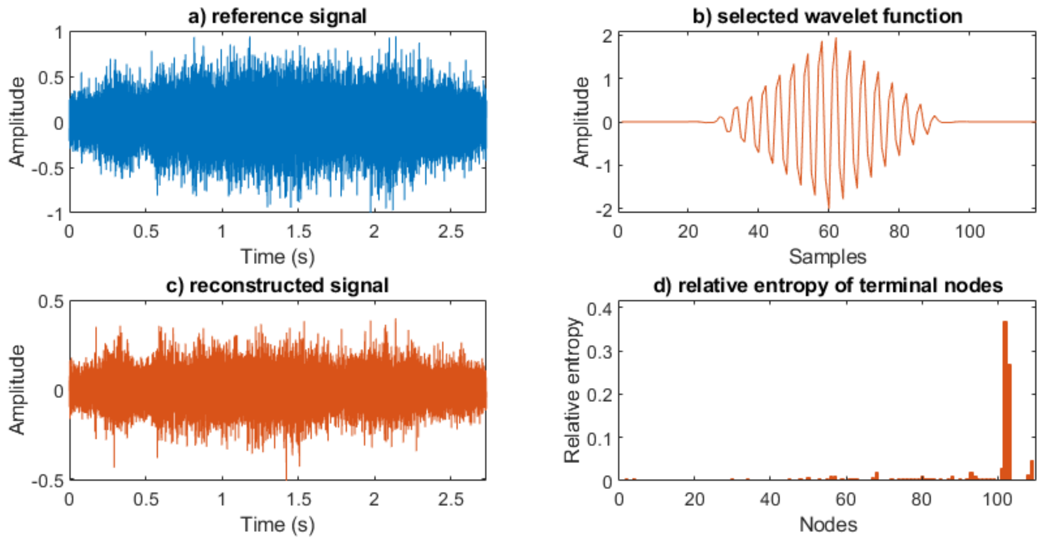

Figure 7. Selected wavelet ('bior2.8' in the node (3,1)) and the reconstruction of the sound of sea waves. The relative entropy of the best node is 0.3670 and the reconstruction NRMSE is 0.9664.

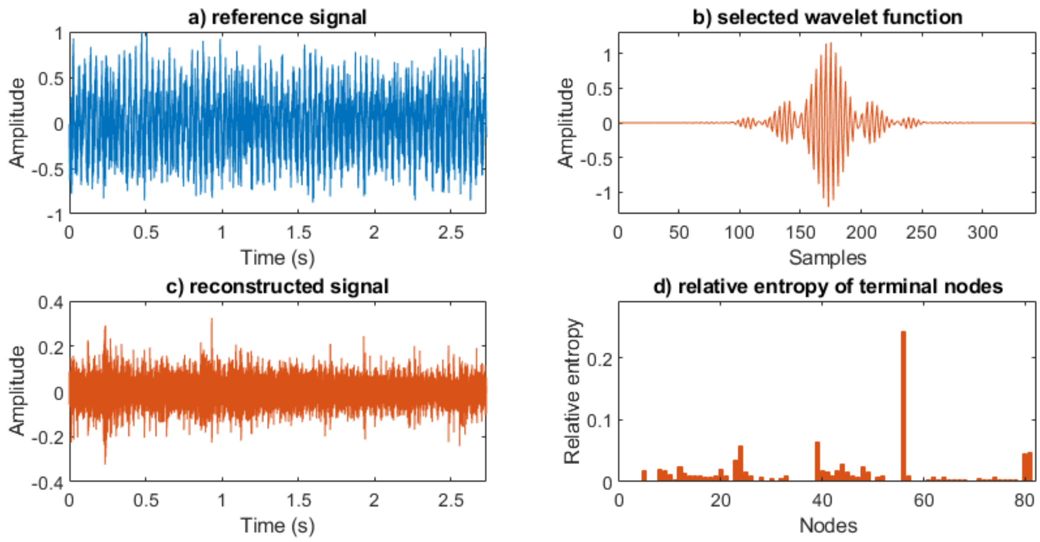

Figure 8. Selected wavelet ('rbio2.6' in the node (3,1)) and the reconstruction of the sound of the city bus. The relative entropy of the best node is 0.2425 and the reconstruction NRMSE is 0.9928.

4. Conclusions and further research

The result of the work is an efficient algorithm, which compromises its complexity with a reasonable performance. The presented approach gives a logical foundation for the selection of wavelet function for the stochastic resynthesis of sound textures. However, there are some relevant limitations and, more importantly, possible solutions and improvements.

An essential requirement for the method is to reduce the decomposition basis to the individual wavelet function. However, very often the information was spread across several nodes of the decomposition tree, which could be observed in the distribution of the entropy of the terminal nodes. The next step to consider is to explore the correlation between the coefficient vectors of different wavelet functions and possibly combine matching wavelets.

Wavelet packet transform seems to be a well-suited tool for the task, however, it has some important disadvantages. Decimation performed with each level of the decomposition greatly reduces the resolution, and implementation of the maximal overlap discrete wavelet transform (MODWT) could indicate how this flaw influences the final results. Also, WPT chooses decomposition scales independently of the source signal. An ideal solution should adapt to the signal's frequency content, as is the case in the empirical wavelet transform (EWT) [12]. Lastly, the complex, multi-level structure of sound textures especially difficult in the analysis could potentially be reflected in hierarchical dependencies between different wavelet functions.

Additional information

The author(s) declare: no competing financial interests and that all material taken from other sources (including their own published works) is clearly cited and that appropriate permits are obtained.

References

1. J. H. McDermott, E.P. Simoncelli; Sound texture perception via statistics of the auditory periphery: evidence from sound synthesis; Neuron, 2011, 71(5), 926–940; DOI: 10.1016/j.neuron.2011.06.032
2. S.A. Nicolas, K. Popat; Analysis and synthesis of sound textures; In: Computational auditory scene analysis, CRC Press, 1998, 293–308
3. J.J. Alvarsson, S. Wiens, M.E. Nilsson; Stress recovery during exposure to nature sound and environmental noise; Int. J. Environ. Res. Public Health, 2010, 7(3), 1036–1046; DOI: 10.3390/ijerph7031036
4. M. Ruiz et al.; Wind turbine fault detection and classification by means of image texture analysis; Mechanical Systems and Signal Processing, 2018, 107, 149–167; DOI: 10.1016/j.ymssp.2017.12.035
5. G. Sharma, K. Umapathy, S. Krishnan; Audio texture analysis of COVID-19 cough, breath, and speech sounds; Biomed. Signal Process. Control., 2022, 76, 103703; DOI: 10.1016/j.bspc.2022.103703
6. C. Roads; Microsound; The MIT Press, 2004
7. I. De Moortel, S.A. Munday, A.W. Hood; Wavelet analysis: the effect of varying basic wavelet parameters; Solar Physics, 2004, 222(2), 203–228
8. P.W. Tse, D. Wang; The automatic selection of an optimal wavelet filter and its enhancement by the new sparsogram for bearing fault detection; Mechanical Systems and Signal Processing, 2013, 40(2), 520–544; DOI: 10.1016/j.ymssp.2013.05.018
9. Y. Jang et al.; The optimal selection of mother wavelet function and decomposition level for denoising of DCG signal; Sensors, 2021, 21(5), 1851; DOI: 10.3390/s21051851
10. R.R. Coifman, M.V. Wickerhauser; Entropy-Based Algorithms for Best Basis Selection; IEEE Transactions on Information Theory, 1992, 38(2), 713–718; DOI: 10.1109/18.119732
11. M.A. Cody; The wavelet packet transform: Extending the wavelet transform; Dr. Dobb's Journal, 1994, 19, 44–46
12. J. Gilles; Empirical wavelet transform; IEEE Transactions on Signal Processing, 2013, 61(16), 3999–4010; DOI: 10.1109/TSP.2013.2265222