VPS VIBRATIONS IN PHYSICAL SYSTEMS

DOI: 10.21008/j.0860-6897.2022.2.15
Vibrations in Physical Systems, 2022, 33(2), 2022215

EMD-Based Time-Frequency Analysis Methods of Non-Stationary Audio Signals

Marcin LEWANDOWSKI[1], **Salomea GRODZICKA**[2]

[1, 2] Warsaw University of Technology, Institute of Radioelectronics and Multimedia Technology, Faculty of Electronics and Information Technology, Nowowiejska 15/19, 00-665 Warsaw, ORCID:

Corresponding author: Marcin LEWANDOWSKI, email: marcin.lewandowski@pw.edu.pl

Abstract To ensure that any time series data is appropriately interpreted, it should be analyzed with proper signal processing tools. The most common analysis methods are kernel-based transforms, which use base functions and their modifications to represent time series data. This work discusses an analysis of audio data and two of those transforms - the Fourier transform and the wavelet transform based on *a priori* assumptions about the signal's linearity and stationarity. In audio engineering, these assumptions are invalid because the statistical parameters of most audio signals change with time and cannot be treated as an output of the LTI system. That is why recent approaches involve decomposition of a signal into different modes in a data-dependent and adaptive way, which may provide advantages over kernel-based transforms. Examples of such methods include empirical mode decomposition (EMD), ensemble EMD (EEMD), variational mode decomposition (VMD), or singular spectrum analysis (SSA). Simulations were performed with speech signal for kernel-based and data-dependent decomposition methods, which revealed that evaluated decomposition methods are promising approaches to analyzing non-stationary audio data.

Keywords: empirical mode decomposition, non-stationary audio data, time-frequency analysis.

1. Introduction

Time-frequency analysis is an essential data processing tool for analyzing and extracting information from signals. Various time-frequency analysis methods decompose a signal into important time-varying features that describe an underlying system's intrinsic behavior, determine parameters needed to construct

a system model, or confirm that the constructed model represents the actual system properly. Historically, Fourier spectral analysis has provided a general method for visualizing the global energy-frequency distributions, while spectrogram (Fourier spectral analysis in a limited time window width) is the most basic method to get time-frequency distribution. Fourier transform is valid under some extremely general conditions [1], but Fourier spectral analysis will have a physical meaning only when the analyzed system is linear and the data is strictly periodic or stationary [2]. Signals acquired for analysis, whether from physical measurements or numerical modeling, are most likely non-stationary, represent a nonlinear process, and the total signal span is most likely too short for proper analysis [2]. Wide-sense stationarity requirement is,

$$\begin{gathered} E(|X(t)^2|) < \infty, \\ E(|X(t)|) = m, \\ C\big(X(t_1), X(t_2)\big) = C\big(X(t_1+\tau), X(t_2+\tau)\big) = C(t_1 - t_2), \end{gathered} \tag{1}$$

for all t, in which $E(\cdot)$ is the expected value defined as an ensemble average and $C(\cdot)$ is a covariance function. A less rigorous definition is for piecewise stationarity when a signal is stationary within a limited time span [2]. In practice, real-world data (including audio signals) is always acquired and analyzed in a finite and limited time span. Thus, it's not stationary [3]. Nonlinearity is an intrinsic feature of many natural phenomena. It is compounded with numerical models, numerical errors, or imperfection of measurement probes, detectors, and acquisition systems. Although linear systems can approximate real-world processes, the previously mentioned complications can make the final data nonlinear. Therefore Fourier-based spectral analysis methods are of limited use and may give misleading results when non-stationary and nonlinear data is approximated with simpler models. There are two main reasons for that. First, the Fourier spectrum needs many additional harmonic components to simulate non-stationary data that are non-uniform globally. These components might make sense mathematically but not physically. Second, Fourier spectral analysis utilizes a linear combination of trigonometric functions (kernel) [2]. Therefore, whenever the form of the analyzed data is different than a pure sine or cosine function, the spectrum will contain some harmonics which are not necessarily related to an energy-frequency distribution of a signal.

Audio data, such as music or speech, which are non-stationary over time, cannot be described by a mathematical expression or approximated by a linear system. That is why recent approaches to analyzing audio data involve the decomposition of a signal into different modes in a data-dependent and adaptive way, which may provide advantages over kernel-based transforms. The empirical mode decomposition (EMD) [2] is an example of such a method and has been shown to be useful in a wide range of signal processing applications. In particular, EMD-based methods decompose a given signal into intrinsic mode functions (IMFs), which carry information at varying frequency scales by detecting local extrema and estimating upper and lower envelopes. IMFs have well-behaved Hilbert transforms, from which the instantaneous frequencies can be calculated. Thus, any event in the analyzed signal can be localized in time and frequency axes [2]. As it will be shown, the standard EMD method suffers from some problems, including mode-mixing, noise sensitivity, and data sampling. Still, some modifications of EMD were adopted to mitigate these problems.

2. Review of non-stationary data processing methods

Available methods for non-stationary data processing can be generally divided into kernel-based and data-dependent (adaptive). Kernel-based methods have an *a priori* selected basis (such as trigonometric functions in Fourier analysis or a specific wavelet function in wavelet analysis) and assume that the analyzed signal is a combination of these basis functions. Data-dependent methods decompose a signal into a set of intrinsic mode functions. Decomposition is based on the direct extraction of the energy associated with various intrinsic time scales, so these methods don't make any assumptions about the nature of the analyzed signal.

2.1. The spectrogram

Classical short-time Fourier transform (STFT) extracts a time-varying representation by localizing the signal in time using a finite window length and applying the Fourier transform to each localized segment [4]. STFT can be expressed as:

$$\mathrm{STFT}\{x(t)\}(t,\omega) = \int_{-\infty}^{\infty} x(t)\, h^{*}(t-\tau)e^{-i\omega t}dt, \tag{2}$$

where h^{*} is a specific window function. Since it relies on Fourier analysis and the window size is fixed, it is assumed that data is piecewise stationary, and the resulting analysis has a uniform time-frequency resolution. Furthermore, to precisely localize an event in time, the window width must be narrow, which on the other hand, results in poor frequency resolution.

2.2. The wavelet analysis

The wavelet transform [5] constructs a representation of the signal using a variable time window by scaling the basic wavelet function, thus creating a multiscale signal representation. Wavelet transform can be expressed with the following general definition:

$$\mathrm{W}(s,u) = \int_{-\infty}^{\infty} x(t)\frac{1}{\sqrt{s}}\psi^{*}\left(\frac{t-u}{s}\right)dt, \tag{3}$$

in which $\psi^{*}(\cdot)$ is the basic wavelet function that satisfies certain general conditions, namely, it is orthonormal, and its mean value is zero, u is a translation of the origin (which gives a temporal time location of an event) and $1/\sqrt{s}$ gives a frequency scale [2]. Wavelet analysis provides a uniform resolution for all the scales, which is the main advantage over STFT.

2.3. Empirical mode decomposition

The empirical mode decomposition is an adaptive time-frequency method for analyzing and decomposing signals into intrinsic mode functions (IMFs), where each IMF must satisfy two conditions: the number of extrema must be equal or differ by one (at most) in relation to the number of zero-crossings and the mean value of the envelope defined by the local maxima and the envelope defined by the local minima is zero [2]. Each IMF component is determined as a difference between the signal and the mean of its envelope (average between upper and lower envelope, see Figure 1):

$$u_i = s(t) - m_i, \tag{4}$$

where m_i is a signal's mean of the envelope.

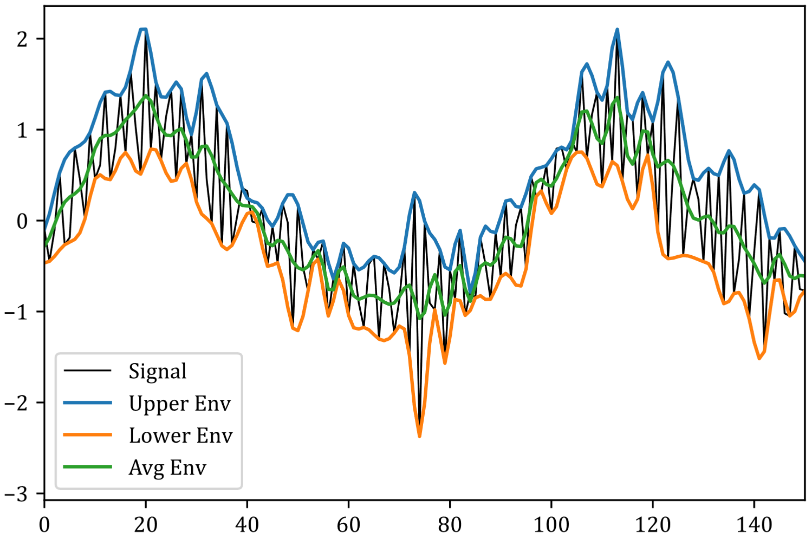

Figure 1. Average of the upper and lower envelope of the input signal.

After the first IMF is calculated and it fulfills the previously mentioned requirements, it is subtracted from the input signal $s(t)$ and the whole process is repeated with residual signal r_1:

$$r_1 = s(t) - u_1 \tag{5}$$

Thus, the analyzed signal is decomposed into a set of n components and the last residual r_n:

$$s(t) = \sum_{j=1}^{n} u_j + r_n, \tag{6}$$

where r_n is dc offset or general trend in the analyzed signal.

2.4. Ensemble empirical mode decomposition

Classic EMD suffers from some problems, including mode-mixing, i.e., the appearance of similar frequency information shared between different IMFs. It is particularly evident with some high-frequency content bursts present in the analyzed signal.

Ensemble EMD (EEMD) attempts to overcome the mode-mixing problem by averaging the decomposition results over an ensemble of noisy versions of the original signal [6]. This can be seen in the Figure 2 where mode-mixing is almost completely reduced, but the noise remains present in IMFs, which can lead to various IMF realizations for the same input signal.

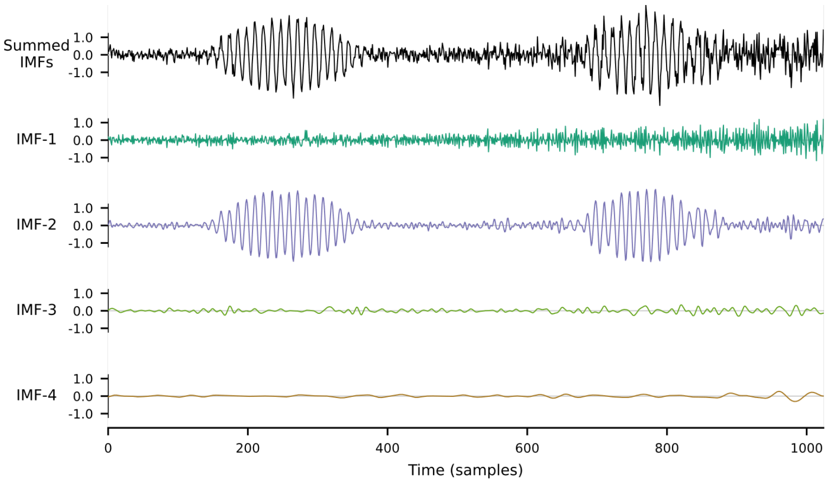

Figure 2. Mode-mixing problem of EMD is partially resolved with EEMD analysis.

In order to overcome these problems, the CEEMDAN (complete ensemble EMD with adaptive noise) method was proposed in [7] and improved in [8]. The idea is that particular noise is added at each stage of the decomposition and a unique residue is computed to obtain each IMF mode.

2.5. Variational mode decomposition

The variational mode decomposition (VMD) allows adaptive decomposition of the signal into various modes by identifying a compact frequency support around its central frequency [9]. Thus, VMD decomposes
a signal into a given number of modes, either exactly or in a least squares sense, such that each individual mode has limited bandwidth. The mode's bandwidth is estimated as the squared H^1 norm of its Hilbert complemented analytic signal with only positive frequencies, shifted to baseband by mixing with a complex exponential of the current center frequency estimate. The variational problem is solved very efficiently in
a classical ADMM approach [10]. Various applications of VMD are described, for example, in [9] and [11]. VMD decomposition of an example signal with three different frequencies and amplitudes plus noise is shown in Figure 3.

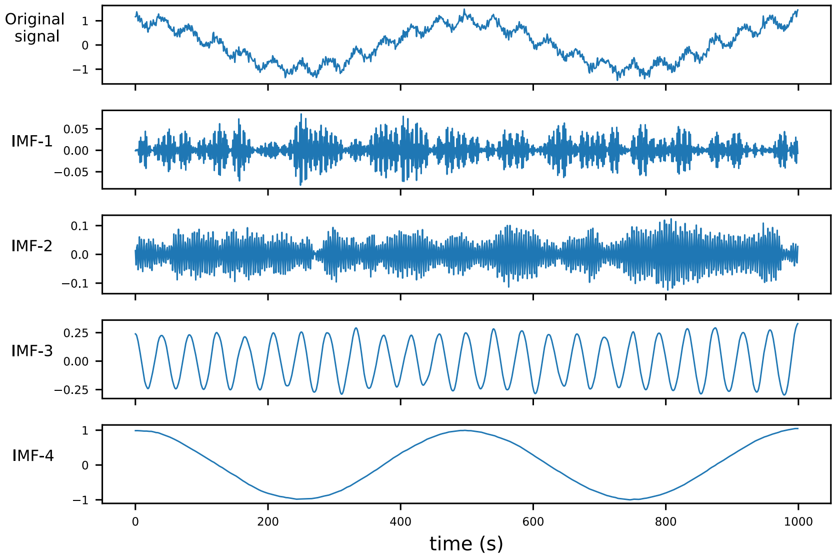

Figure 3. VMD decomposition of harmonic signal plus noise into a set of 4 IMFs.

2.6. Masked and iterated masked EMD

Some drawbacks of the EMD were briefly discussed in Sect. 2.3 and Sect. 2.4 along with a description of noise-assisted methods (EEMD and CEEMDAN) which alleviate these drawbacks. Noise-assisted methods create an ensemble of many sifts processes, each with different or the same noise injected. Then, the final IMFs are computed as an average across this ensemble. Still, these methods suffer from some degree of mode-mixing, unwanted residual noise, or splitting the signal between modes [12].

Recently, another approach was proposed for improving EMD decomposition. Masked EMD [13] works by injecting a masking signal into input before sifting. This reduces mode-mixing by making the sift ignore signal content slower than the frequency of the masking signal [12]. With a mask, any signal content with frequencies much lower than the masking frequency will be ignored by the sift in that iteration and replaced by the mask. The mask is finally removed, which allows recovering intermittent activity correctly [12]. The choice of masking signals remains an area of active research, still, it remains a manual process in many cases. This requires experience and may introduce subjective bias [13-17].

The problem with the choice of masking signal was recently reduced with iterated masked EMD (itEMD) proposed in [12]. itEMD is a sifting method that automates the choice of mask signal frequencies based on the data. It can automatically identify oscillations and minimizes mode-mixing without specifying masking signals.

2.7. Hilbert-Huang transform

The Hilbert-Huang transform describes how the energy or power within a signal is distributed across frequency and time. It uses Hilbert spectral analysis (HSA) [2,18] as a method for examining the input signal's instantaneous frequency as a function of time. Thus, this transform is well suited for examining non-stationary data. Instantaneous frequency can be calculated with Hilbert transform, and then the signal can be expressed by:

$$s(t) = Real\left\{\sum_{j=1}^{n} a_j(t)\exp\left(i\int \omega_j(t)\,dt\right)\right\}, \tag{7}$$

where $a_j(t)$ and $\omega_j(t)$ are instantaneous amplitude and frequency, respectively. To make this analysis possible, the input signal needs to meet specific conditions, which are satisfied by each of the IMFs decomposed with EMD-based methods. The final result is a frequency-time distribution of signal amplitude (or energy), which permits the identification of localized features. As an example, the Hilbert-Huang transform of the nonlinear and non-stationary signal is shown in Figure 4 below.

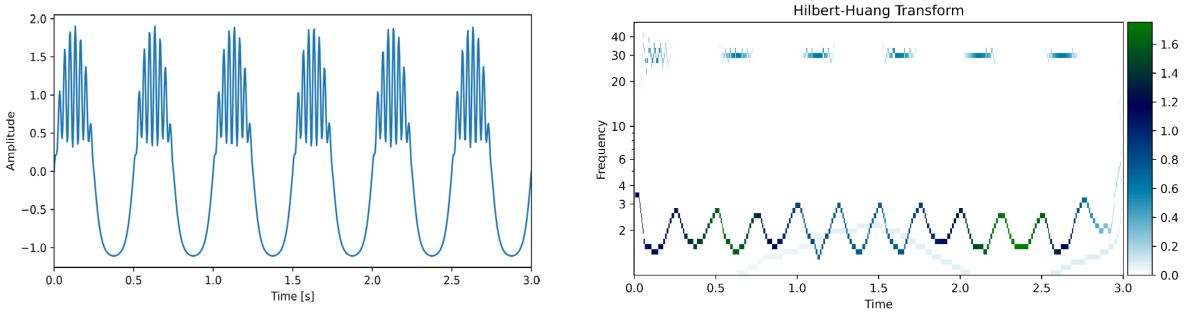

Figure 4. Nonlinear and non-stationary synthetic signal and Hilbert-Huang transform for EMD decomposition.

2.8. Other methods

The synchrosqueezed transform (SST) [19] improves over the wavelet analysis by calculating instantaneous frequencies and "squeezing" them through a reassignment algorithm, namely, shifting them to the center of the time-frequency region [20]. This leads to sharper time-frequency representations than the STFT and wavelet analysis, which are often limited by the finite sampling lengths and can create spectral smearing. In addition, the sharpening essentially prunes the unnecessary wavelet coefficients, thus leading to a sparser representation.

Another method is Singular Spectrum Analysis [21] which decomposes the original time series into the sum of a small number of independent and interpretable components such as a slowly varying trend, oscillatory components and a structureless noise. SSA is based on the singular value decomposition (SVD) of a specific matrix constructed upon the time series. Neither a parametric model nor stationarity-type conditions have to be assumed for the time series. This makes SSA a model-free method and hence enables SSA to have a very wide range of applicability.

Other methods are Intrinsic-Time Scale Decomposition (ITD) [22] or Fourier Decomposition method (FDM) [23].

3. Application of EMD-based methods to speech audio data

The speech signal is generated by a complex psycho-acoustic process developed as a result of thousands of years of human evolution. It contains a multitude of information like the speaker's age, height, emotion, accent, health and physiological disorders, identity, etc., which give rise to the various fields of Speech Processing [24-25]. However, speech is a highly nonlinear and non-stationary signal [26], and hence extracting such information is not a trivial task [25,27]. EMD-based methods could reveal more information about speech signals than traditional Fourier and wavelet-based approach

The analyzed signal was a letter 'A' spoken by a female and recorded at 48 kHz in 16 bits. Waveform and Fourier spectrum of speech signal are shown in Figure 5. STFT and wavelet analysis are shown in Figure 6 (left and right panel, respectively). All transforms reveal that this speech signal has fundamental frequency and harmonics at about 200, 400, 600, 800, 950, 1050 and 1100 Hz. However, this kind of analysis cannot indicate how components interact with each other or how their parameters change with time.

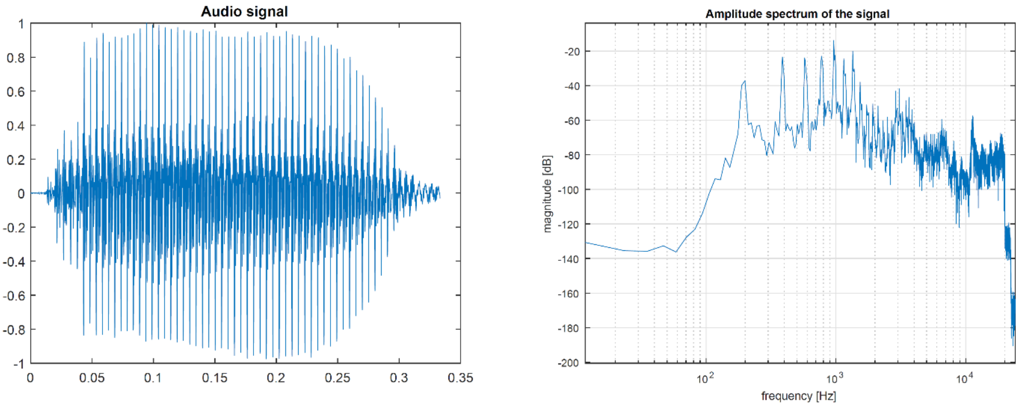

Figure 5. Speech signal – female voice, letter 'A, its waveform in left panel, and FFT in right panel.

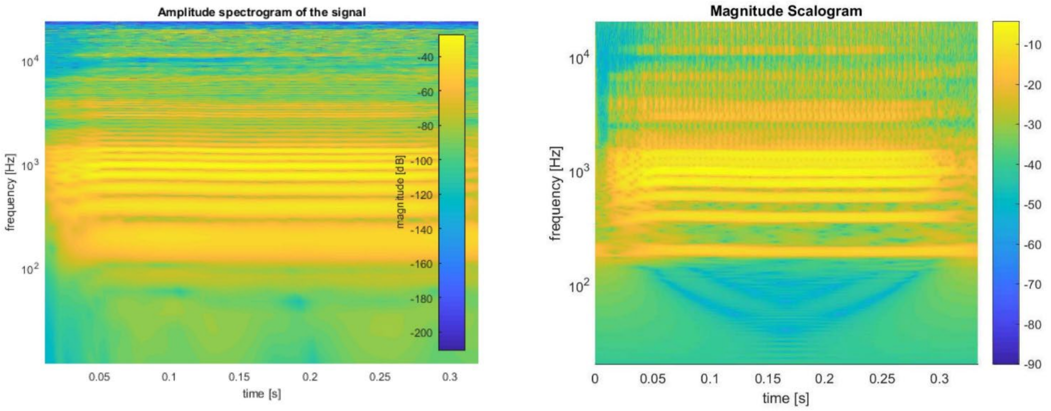

Figure 6. Short-time Fourier transform (left panel) and wavelet transform (right panel) of speech signal (letter 'A').

EMD decomposition gives slightly different results, shown in Figure 7. Each decomposed mode is a characteristic AM-FM modulated signal, but it is clearly seen on left panel in Figure 7 that IMF1-IMF3 represents one of the main drawbacks of the classic EMD, namely mode-mixing (the same narrow-band energy is contained in all three components over time).

Ensemble EMD improves the analysis by adding some noise to each of the components before the sifting process. All IMFs are AM-FM narrow-band representations of the whole signal, but each component's instantaneous frequency is still not stable and is smeared over a frequency range near 1kHz (see Figure 8).

VMD method gives the best time-frequency resolution of decomposed IMFs, shown in Figure 9. The main energy is focused around 1kHz and there is evident AM-FM modulation in each component.

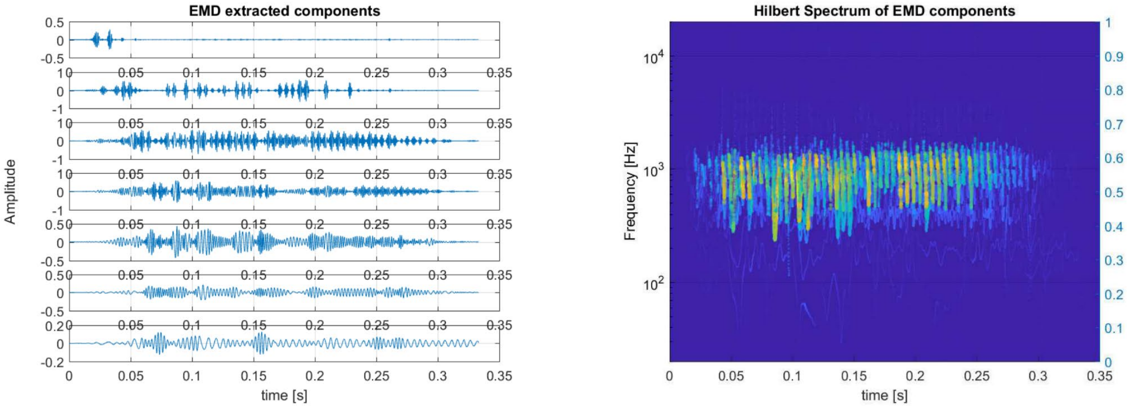

Figure 7. EMD decomposition and HSA of speech signal (letter 'A' spoken by a female).

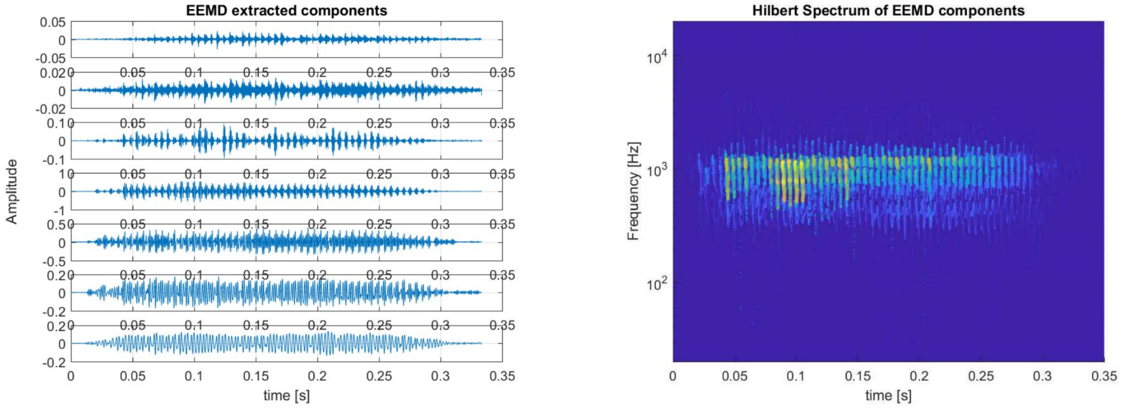

Figure 8. EEMD decomposition and HSA of speech signal (letter 'A' spoken by a female).

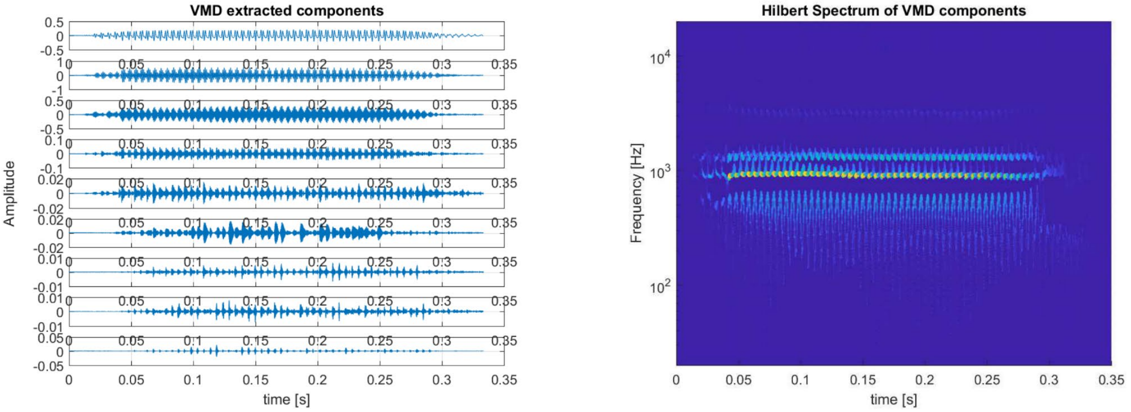

Figure 9. VMD decomposition and HSA of speech signal (letter 'A' spoken by a female).

4. Conclusions

This paper briefly describes the analysis and processing of data (speech signal as an example) from a nonlinear and non-stationary point of view, as opposed to the traditional short-time linear and stationary analysis. Audio data, such as music or speech, are inherently non-stationary over time and cannot be described by a mathematical expression or approximated by a linear system. This motivates AM-FM representation of audio data, which models audio as being constituted of an ensemble of AM-FM signals.

A speech signal is only one of many audio data that cannot be approximated with a linear model. For example, quantization noise is always modeled as a white noise added to the system, but it is a strictly non-stationary, input-dependent, and nonlinear error signal. Another example is time series data of state variables in all systems with feedback (modulators, adaptive filters, prediction algorithms, IIR filters) which are prone to changing their parameters over time. Methods based on EMD decomposition are naturally empirical and each approach described in a paper has some drawbacks. Thus, these kinds of time-frequency analysis methods remain an active research area.

Future work will focus on the investigation of Holo-Hilbert spectral analysis (HHSA) [28], which dives deeper into a decomposition of signals and can reveal so far unknown information in data. HHSA has been successfully adopted in biomedical and neuroscience data analysis [29–32], so it could also be used in audio data analysis.

Acknowledgments

This work was supported by the Warsaw University of Technology, Poland, under Grant 1820/10/201/POB2/2021.

Additional information

The authors 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. S. Bochner; Fourier Integrals: Introduction to the Theory of Fourier Integrals. By EC Titchmarsh. Oxford, Clarendon Press, 1937; Science, 87, 2260, 370, 1938.
2. N.E. Huang et al.; The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis, Proc. R. Soc. Lond. Ser. Math. Phys. Eng. Sci., 1998, 454(1971), 903–995. DOI:10.1098/rspa.1998.0193.
3. P.J. Brockwell, R. A. Davis; Time series: theory and methods. Springer science & business media, 2009.
4. J. Allen; Short term spectral analysis, synthesis, and modification by discrete Fourier transform; IEEE Trans. Acoust. Speech Signal Process., 1977, 25(3), 235–238.
5. Y. Meyer; Wavelets and Operators: Volume 1; Cambridge University Press, 1992.
6. Z. Wu, N. E. Huang; Ensemble empirical mode decomposition: a noise-assisted data analysis method; Adv. Adapt. Data Anal., 2009, 1(1), 1–41.
7. M.E. Torres, M.A. Colominas, G. Schlotthauer, P. Flandrin; A complete ensemble empirical mode decomposition with adaptive noise; In: IEEE international conference on acoustics, speech and signal processing (ICASSP), 2011, 4144–4147.
8. M.A. Colominas, G. Schlotthauer, M.E. Torres; Improved complete ensemble EMD: A suitable tool for biomedical signal processing; Biomed. Signal Process. Control, 2014, 14, 19–29.
9. T. Liu, Z. Luo, J. Huang, S. Yan; A comparative study of four kinds of adaptive decomposition algorithms and their applications; Sensors, 2018, 8(7), 2120.
10. K. Dragomiretskiy, D. Zosso; Variational mode decomposition; IEEE Trans. Signal Process., 2013, 62(3), 531–544.
11. V.R. Carvalho, M.F. Moraes, A.P. Braga, E.M. Mendes; Evaluating five different adaptive decomposition methods for EEG signal seizure detection and classification; Biomed. Signal Process. Control, 2020, 62, 102073.
12. M.S. Fabus, A.J. Quinn, C.E. Warnaby, M.W. Woolrich; Automatic decomposition of electrophysiological data into distinct nonsinusoidal oscillatory modes; J. Neurophysiol., 2021, 126(5), 1670–1684.
13. R. Deering, J.F. Kaiser; The use of a masking signal to improve empirical mode decomposition; In: Proceedings ICASSP'05, IEEE International Conference on Acoustics, Speech, and Signal Processing, 2005., 4, iv–485.
14. A.J. Quinn, et al.; Within-cycle instantaneous frequency profiles report oscillatory waveform dynamics; J. Neurophysiol., 2021, 126(4), 1190–1208.
15. Y. Yang, J. Deng, D. Kang; An improved empirical mode decomposition by using dyadic masking signals; Signal Image Video Process., 2015, 9(6), 1259–1263.
16. O.B. Fosso, M. Molinas; EMD mode mixing separation of signals with close spectral proximity in smart grids; 2018 IEEE PES innovative smart grid technologies conference Europe (ISGT-Europe), 2018, 1–6.
17. S. Cole; B. Voytek; Cycle-by-cycle analysis of neural oscillations; J. Neurophysiol., 2019, 122(2), 849–861.

18. A.V. Oppenheim; Discrete-time signal processing; Pearson Education India, 1999.
19. Daubechies, J. Lu, H.-T. Wu; Synchrosqueezed wavelet transforms: An empirical mode decomposition-like tool; Appl. Comput. Harmon. Anal., 2011, 30(2), 243–261.
20. F. Auger, et al.; Time-frequency reassignment and synchrosqueezing: An overview; IEEE Signal Process. Mag., 2013, 30(6), 32–41.
21. J.B. Elsner, A.A. Tsonis; Singular spectrum analysis: a new tool in time series analysis; Springer Science & Business Media, 1996.
22. M. G. Frei, I. Osorio; Intrinsic time-scale decomposition: time–frequency–energy analysis and real-time filtering of non-stationary signals; Proc. R. Soc. Math. Phys. Eng. Sci., 2007, 463(2078), 321–342. DOI:10.1098/rspa.2006.1761.
23. P. Singh, S.D. Joshi, R.K. Patney, K. Saha; The Fourier decomposition method for nonlinear and non-stationary time series analysis; Proc. R. Soc. Math. Phys. Eng. Sci., 2017, 473, 2199, 20160871.
24. L.R. Rabiner, R. W. Schafer, et al.; Introduction to digital speech processing; Found. Trends® Signal Process., 2007, 1(1–2), 1–194.
25. J. Benesty, M.M. Sondhi, Y. Huang, et al.; Springer handbook of speech processing, vol.1. Springer, 2008.
26. D. Kapilow, Y. Stylianou, J. Schroeter; Detection of non-stationarity in speech signals and its application to time-scaling; Sixth European Conference on Speech Communication and Technology, EUROSPEECH 1999, Budapest, Hungary, September 5-9, 1999. DOI:10.21437/Eurospeech.1999-503
27. R.S. Holambe, M.S. Deshpande, Advances in non-linear modeling for speech processing. Springer Science & Business Media, 2012.
28. N.E. Huang et al.; On Holo-Hilbert spectral analysis: a full informational spectral representation for nonlinear and non-stationary data; Philos. Trans. R. Soc. Math. Phys. Eng. Sci., 2016, 374, 2065, 20150206. DOI:10.1098/rsta.2015.0206.
29. P.-L. Lee et al.; The Full Informational Spectral Analysis for Auditory Steady-State Responses in Human Brain Using the Combination of Canonical Correlation Analysis and Holo-Hilbert Spectral Analysis; J. Clin. Med., 2022, 11(13), 3868.
30. N. Moradi, P. LeVan, B. Akin, B.G. Goodyear, R.C. Sotero; Holo-Hilbert spectral-based noise removal method for EEG high-frequency bands; J. Neurosci. Methods, 2022, 368, 109470.
31. W.-K. Liang, P. Tseng, J.-R. Yeh, N.E. Huang, C.-H. Juan; Frontoparietal beta amplitude modulation and its interareal cross-frequency coupling in visual working memory; Neuroscience, 2021, 460, 69–87.
32. C.-H. Juan et al.; Revealing the dynamic nature of amplitude modulated neural entrainment with Holo-Hilbert spectral analysis; Front. Neurosci., 2021, 15, 673369.