Advances in Science and Technology
Research Journal
Volume 12, Issue 1, March 2018, pages 251–259 *Research Article*
DOI: 10.12913/22998624/87028

ADVANCED TIME-FREQUENCY REPRESENTATION IN VOICE SIGNAL ANALYSIS

Dariusz Mika[1], Jerzy Józwik[2]

[1] The State School of Higher Education, The Institute of Technical Sciences and Aviation, 54 Pocztowa Street, 22-100 Chełm, Poland, e-mail: dmika@pwsz.chelm.pl

[2] Lublin University of Technology, Mechanical Engineering Faculty, Department of Production Engineering, 36 Nadbystrzycka Street, 20-618 Lublin, Poland, e-mail: j.jozwik@pollub.pl

Received: 2017.07.13
Accepted: 2018.02.01
Published: 2018.03.01

ABSTRACT

The most commonly used time-frequency representation of the analysis in voice signal is spectrogram. This representation belongs in general to Cohen's class, the class of time-frequency energy distributions. From the standpoint of properties of the resolution, spectrogram representation is not optimal. In Cohen class representations are known which have a better resolution properties. All of them are created by smoothing the Wigner-Ville'a distribution characterized by the best resolution, however, the biggest harmful interference. The used smoothing functions decide about a compromise between the properties of resolution and eliminating harmful interference term. Another class of time-frequency energy distributions is the affine class of distributions. From the point of view of readability of analysis of the best properties are known so called Redistribution of energy caused by the use of a general methodology referred to as *reassignment* to any time-frequency representation. Reassigned distributions efficiently combine a reduction of the interference terms provided by a well adapted smoothing kernel and an increased concentration of the signal components.

Keywords: signal analysis, spectrogram, time-frequency analysis, time-frequency representation, Cohen's class, Wigner-Ville'a distribution

INTRODUCTION

A time-frequency representation (TFR) is widely used in the analysis of non-stationary signals such as,: human speech signals, ECG signals, geophysical signals. Analyzed signal is represented as a joint function of time and frequency – rather than as a function of time or frequency [2, 3]. Such an analysis should constitute an important tool for understanding many processes and phenomena within problems of estimation, detection or classification. A unified way of presenting different kinds of TFR was followed by L. Cohen in the mid-1960s (in a context of quantum mechanics), what has become known as Cohen's class since then [16, 17]. There has been a rapid growth of interest in this subject. The diversity of theoretical and practical viewpoints from which they could be approached and the numerous known results would make a complete synthesis of this subject a voluminous document [13–15, 26–28]. Moreover, succinct [9, 13–15, 17, 33, 36, 39, 40] or detailed [18, 22, 23] publications already exist, and the reader is invited to refer to them.

Time-frequency analysis (TF) can be used for acoustic signal analysis, including speech signals [4]. TF analysis is used as a tool in various types of techniques such as: speech coding, speech synthesis, speech recognition and speaker recognition. These techniques are defined by a common term speech processing. Speech processing mostly performs two fundamental operations:

Feature Extraction [6] and Classification [7]. In scientific work [13] the authors presented survey of various feature extraction techniques in speech processing such as Fast Fourier Transforms, Linear Predictive Coding, Mel Frequency Cepstral Coefficients, Discrete Wavelet Transforms, Wavelet Packet Transforms and their applications in speech processing. Very often for effective analysis speech signals they are used sparse time frequency representations [19, 24, 35]. In scientific work [35] a comparison between atomic decomposition methods and time-frequency distributions with respect to speech signals is presented. The authors demonstrated that the highest resolution of the analysis is achieved with the application Positive Time-Frequency Distributions. In scientific work [38] the authors conducted a comparison of speech signal analysis with the use of discreet Fourier transformation (DFT) and discreet cosines transformation DCT. The authors showed that the spectrograms plotted using DCT are clearer than the spectrograms plotted using same point DFT. Demonstrated that spectrogram using DCT is characterized by a higher resolution in relation to DFT. In the context of speech signal recognition, artificial neural networks are also used. The authors [34] compare various popular signal representations such as short-time Fourier transform (STFT) with linear and Mel scales, constant-Q transform (CQT) and continuous Wavelet transform (CWT), and assess their impact on the classification performance of environmental sound datasets using convolutional neural networks. It was shown that Mel-STFT spectrograms were consistently good performers across the variations tested. Time – frequency analysis is also used in biomedical diagnostics and analysis.

In phoniatrics time-frequency representations (primarily spectrograms and scaleograms) are used for diagnosis of diseases of the vocal organs [1, 5, 9, 19, 20, 11, 35, 43, 45]. The authors [21] presented an overview of the methodology of automatic detection of pathological changes in the voice. In scientific work [37] the authors demonstrated detection and discrimination of voice disorders in using modulation frequency analysis. The authors [19] used TF representations from the class Cohen for vocal fold's onset signal for diagnosis of different phonation disorders evoked by pathological changes. The vibration signals are acquired by direct optical inspection of the glottis using an endoscope and a high speed CCD camera system. In order to analyze the speech signal, TF representations from the Cohen class were used along with cone kernel distribution to ensure maximum smoothness over time. The authors show that even small pathological changes in the vocal folds are visible on the time-frequency plane, which allows sensitive detection of affects and helps to diagnose. Authors of many works in order to identify diseases and pathological changes in the voice used a discrete wavelet transformation DWT [1, 41] and support vector machine-based classification method as feature classification tools [1, 5, 6, 11]. In scientific work [1, 21, 44] demonstrated that the most effective algorithm (100% recognition efficiency) is a system composed of wavelet packet transforms along with feature dimension reduction by linear discriminant analysis and a support vector machine-based classification method.

MATHEMATICAL BASICS OF ANALYSIS

The concept of affine time-frequency representation was introduced for the first time in 1985 [8] and is based on wavelet transform. In fact, the wavelet transform is directly connected to affine time-frequency representations by means of a smoothing operation in the time-frequency plane, which explains the great importance of affine time-frequency representations in signal analysis. First construction of these distributions was based on group theory [6, 7], a powerful tool in signal analysis, which did not, however, stir up an enthusiasm comparable with that caused by the study of Cohen's class distributions [16]. Second approach [28] relies on the affine smoothing of certain distributions of Cohen's class.

Spectrogram often used in signal analysis is the squared modulus of the short time Fourier transformation (STFT) of signal *x(t)* and can be expressed as (1):

$$S_x(t,\nu) = \\ = \left|\int_{-\infty}^{+\infty} x(u)h^*(u-t)e^{-j2\pi\nu u}du\right|^2 \quad (1)$$

where: t – time, u – shift in time domain, ν – frequency, $h(t)$- time window and '*' is conjugate operator

Spectrogram belongs to the class of time-frequency energy distributions. The purpose of energy distributions is to distribute the *energy* of the signal over the two description variables: time and frequency. Among the desirable proper-

ties of an energy time-frequency distribution, two of them are of particular importance: *time and frequency covariance* (2–3):

$$y(t) = x(t - t_0) \;\rightarrow\; S_y(t,\nu) = S_x(t - t_0, \nu) \tag{2}$$

$$y(t) = x(t)\exp(j2\pi\nu_0 t) \;\rightarrow\; S_y(t,\nu) = S_x(t, \nu - \nu_0) \tag{3}$$

where: t – time, t_0 – shift in time domain,
ν – frequency,
ν_0 – shift in frequency domain.

These properties guaranty that, if the signal is delayed in time and modulated, its time-frequency distribution is translated of the same quantities in the time-frequency plane. This group of transformations is called the *Weyl-Heisenberg* group [3]. It has been shown that the class of energy time-frequency distributions verifying these covariance properties possesses the following general expression (4):

$$C_x(t,\nu;\phi) = \iiint_{-\infty}^{+\infty} e^{j2\pi\xi(s-t)}\phi(\xi,\tau)\, x(s+\tau/2)\, x^*(s-\tau/2) e^{-j2\pi\nu\tau} d\xi ds d\tau \tag{4}$$

where: $\phi(\xi, \tau)$ is two-dimensional function called *parameterization function*,
(ξ, τ) is same times called *Doppler frequency* and *delay* and s is shift in time domain,
t – time

This class of quadratic time-frequency distributions, which preserves time and frequency shifts is called Cohen's class [21]. It can also be written as (5):

$$C_x(t,\nu;\mathbf{\Pi}) = \iint_{-\infty}^{+\infty} \mathbf{\Pi}(t-s,\nu-f)\, W_x(s,f)\, ds\, df \tag{5}$$

where f is a shift in frequency domain and (6):

$$\mathbf{\Pi}(t,\nu) = \iint_{-\infty}^{+\infty} \phi(\xi,\tau)\, e^{-j2\pi(\nu\tau-\xi t)} d\xi d\tau \tag{6}$$

is the two-dimensional Fourier transfer of the parameterization function $\phi(\xi, \tau)$ and

$$W_x(s,f) = \int_{-\infty}^{+\infty} x(s+\tau/2)\, x^*(s-\tau/2)\, e^{-j2\pi f\tau} d\tau \tag{7}$$

is Wigner-Ville distribution (WVD) of signal $x(t)$. In the case where Π is a smoothing function, this expression allows one to interpret C_x as a smoothed version of the WVD; consequently, such a distribution will attenuate in a particular way the interferences of the WVD.

This class is of significant importance since it includes a large number of the existing time-frequency energy distributions. The type of parameterization function ϕ used (or smoothing function Π) determines the type of signal representation. The most frequently used representations include: Pseudo-Wigner-Villa distribution PWVD, Smoothed-pseudo Wigner-Villa distribution SPWVD, Rihaczek distribution, Margenau-Hill distribution, Choi-Williams distribution, Born-Jordan distribution and Zhao-Atlas-Marks distribution called also Cone-Shaped Kernel distribution [32].

The Cohen's class, is based on the properties of covariance by shifts in time and in frequency. In order to favor a time-scale approach of the signal, one can also choose to put forward, among these desirable properties, the *covariance by translation in time and dilation* [30]. The corresponding group of transforms, counterpart the Weyl-Heisenberg group, is the *affine group.* It can be expressed as (8):

$$\mathbf{\Psi}_x(t,\nu;\mathbf{\Pi}) = \iint_{-\infty}^{+\infty} \mathbf{\Pi}\left(\frac{s-t}{a}, af\right) W_x(s,f)\, ds\, df \tag{8}$$

The set of such representations defines the affine class, which is the class of time-frequency energy distributions covariant by translation in time and dilation. As in the case of Cohen class smoothing function type Π determines the type of representation. The most popular are: affine smoothed pseudo Wigner distribution ASPWD, Bertrand distribution, D-Flandrin distribution, Unterberger distribution (active and passive) [4, 6, 7, 23, 29, 30, 42].

Bilinear time-frequency distributions offer a wide range of methods designed for the analysis

of non-stationary signals. Nevertheless, a critical point of these methods is their readability, which means both a good concentration of the signal components and no misleading interference terms. Some efforts have been made recently in that direction, and in particular a general methodology referred to as *reassignment* [34, 32]. The original idea of reassignment was introduced in an attempt to improve the spectrogram. Indeed, as any other bilinear energy distribution, the spectrogram is faced with an unavoidable trade-off between the reduction of misleading interference terms and a sharp localization of signal components. The reassignment method concentrate the averaged energy of signal not at the geometrical center but rather at the center of gravity of the domain. Reasoning with a mechanical analogy, the local energy *Time*-distribution for example $\Pi(t-s, \nu-f)\, W_x(s,f)$ (as a function of s and f) can be considered as a mass distribution, and it is much more accurate to assign the total mass (i.e. the spectrogram value) to the center of gravity of the domain rather than to its geometrical center. The reassignment principle can be used for any distributions belonging to Cohen and affine class. Reassigned distributions efficiently combine a reduction of the interference terms provided by a well adapted smoothing kernel and an increased concentration of the signal components achieved by the reassignment.

METHODOLOGY AND RESULTS OF ANALYSIS

The time-frequency representations presented above were applied to speech signals with disease syndromes. The used samples of recordings come from a person suffering from larynx cancer. Recordings were made using a simple audio recorder with a sampling frequency of 10kHz. TF analysis was performed in Matlab using the Time-Frequency Toolbox. For comparison, the results of the TFR analysis were presented using both the Cohen and the affine class TFR and their reassignment version. In addition to the spectrogram and its reassignment, PWVD and SPWVD reassignment versions (Fig. 2, 3, 4) were presented in the Cohen class. In the affine class, Morlet's skalogram with reassignment and ASPWD (Fig. 5) were presented. In order to reduce processing time, processing was performed on samples with a reduced sampling rate of up to 2 kHz, so the received TFR images are limited to 1 kHz (the frequency scale is a relative scale). The time scale corresponds to the number of samples. The time of the analyzed signal was 5 seconds. In the preprocessing stage, the recorded speech signal was processed into an analytical signal using Hilbert transform. TFR transformations were performed using the Keizer window.

Figure 1 shows the Wigner-Ville distribution of the analyzed speech signal, which is the "basis" for the TFR representation used. Harmful interference is shown in the image. TFR representations are obtained by appropriate smoothing of WV representation in the time and frequency domain.

The classical method of presentation of a sound signal in the form of a spectrogram (Fig. 2a) is obtained by simultaneous time and frequency smoothing. This results in the removal of harmful interference clearly visible on the WV representation (Fig. 1), but at the expense of reduced resolution and legibility of the received image.

The PWVD representation (Fig. 3a) obtained by smoothing only in the frequency do-

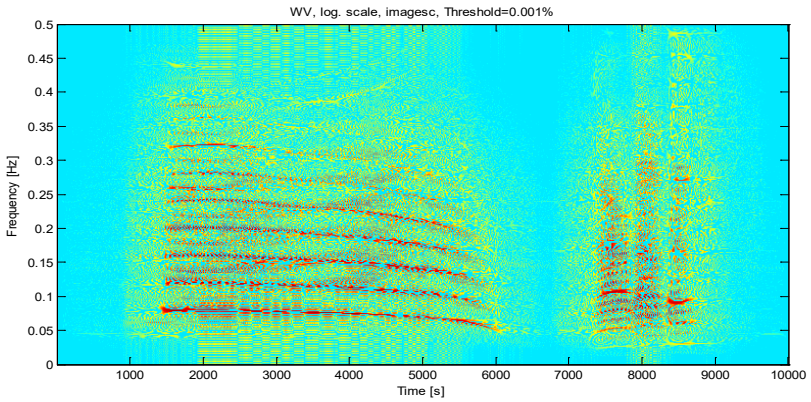
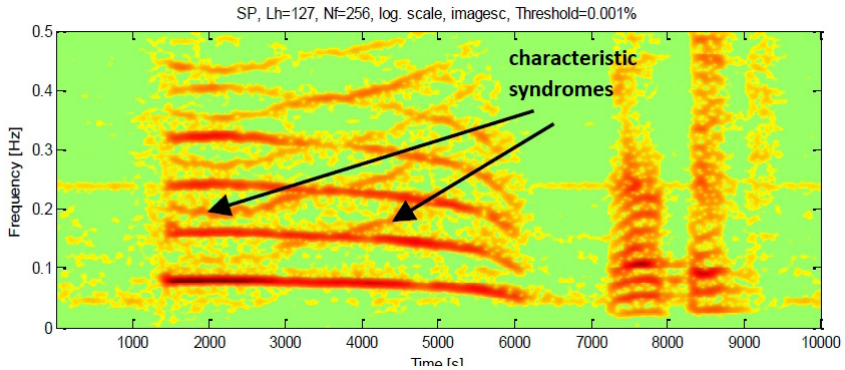


Fig. 1. Wigner-Ville distribution of the analyzed speech signal along with harmful interference

a)



b)

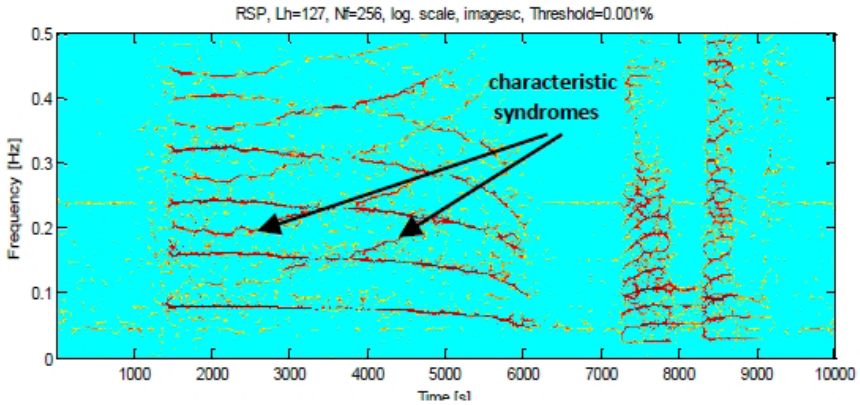


Fig. 2. The classical method of presentation of a sound signal is: a) a spectrogram, and b) its reassignment. Arrows indicate characteristic syndromes

main reduces the degree of harmful interference, however, the readability of the resulting image is still insufficient.

The controlled degree of smoothing both in the time domain and frequency characteristic for the SPWVD representation (Fig. 4a) results in a significant increase in the readability of the obtained images relative to PWVD, despite visible harmful interference.

Similar to the affine class, simultaneous time and scale smoothing in the field of wavelet transformations (Fig. 5a) (Morler's Skalogram) causes low readability of the received image despite the removal of harmful interferences. Controlled smoothing separately in time and scale in the case of ASPWD (Fig. 5b) significantly improves readability of the resulting image.

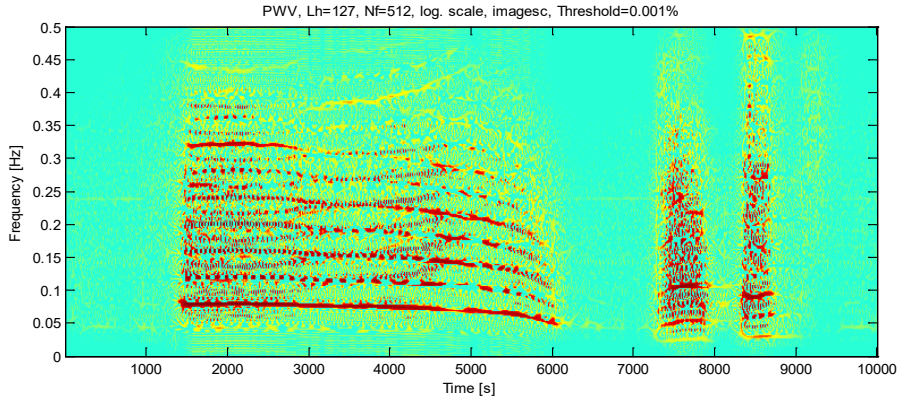
The use of the reassignment version for the TFR representation significantly increases the resolution of the images obtained, thereby increasing the readability. The lesions are visible in a slightly fuzzy manner on the spectrogram images (Fig.1a), SPWVD and ASPWD on reassignment images (Fig. 2b, 3b, 4b and 5c) are clearly distinguishable which makes them easy to detect.

In the context of medical diagnostics, such an improvement of readability of the received TFR images can contribute to improved detection and evaluation of existing speech disorders. The complexity of the numerical operations occurring with such a representation results, however, in a significant lengthening of the analysis time, which excludes on-line processing mode applications.

CONCLUSIONS

Representations belonging to the class of time-frequented energy distributions are well known theoretically. However, their use in science and technology is mainly limited to a simple spectrogram, which is characterized by the lack of control over the degree of smoothing the signal. The resulting TFR image is therefore blurry, and for some applications (especially where the nuances in TFR play an important role) may not be sufficient. The use of other representations belonging to this class, and especially in conjunction with their reassignment versions, allows a very accurate analysis of the signal at the time-frequency plane.

a)



b)

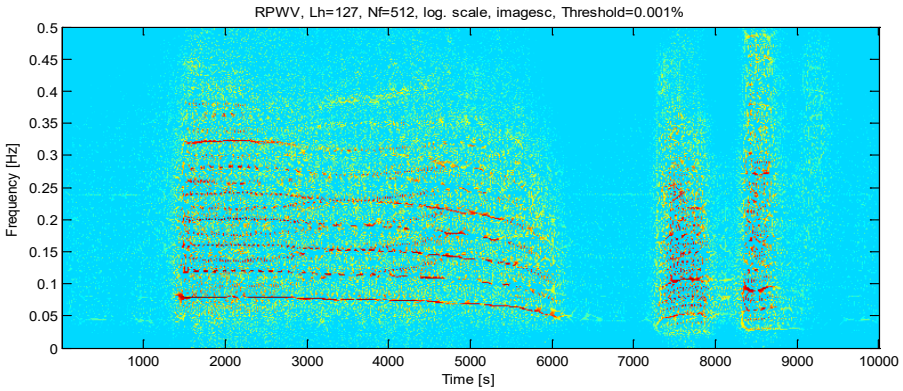
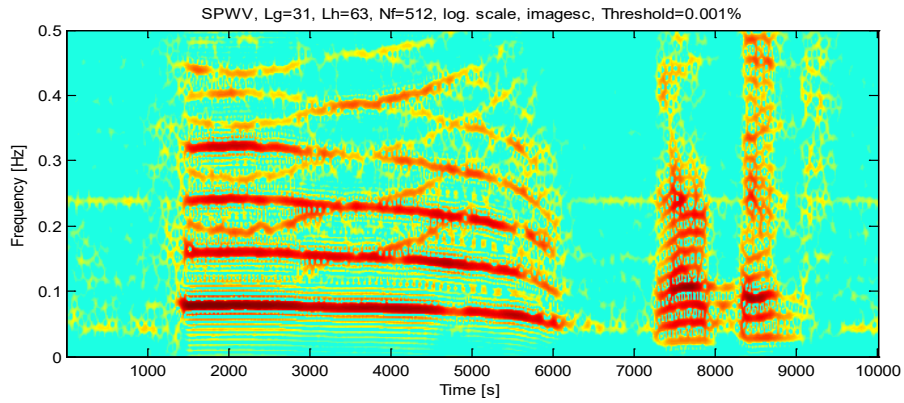


Fig. 3. Pseudo Wigner-Ville distribution a) and b) its reassignment version

a)



b)

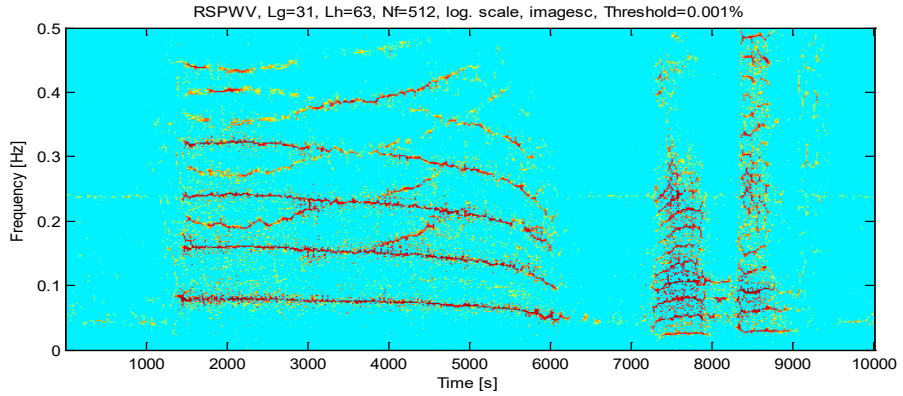


Fig. 4. Smoothed Pseudo Wigner-Ville distribution a) and b) its reassignment version

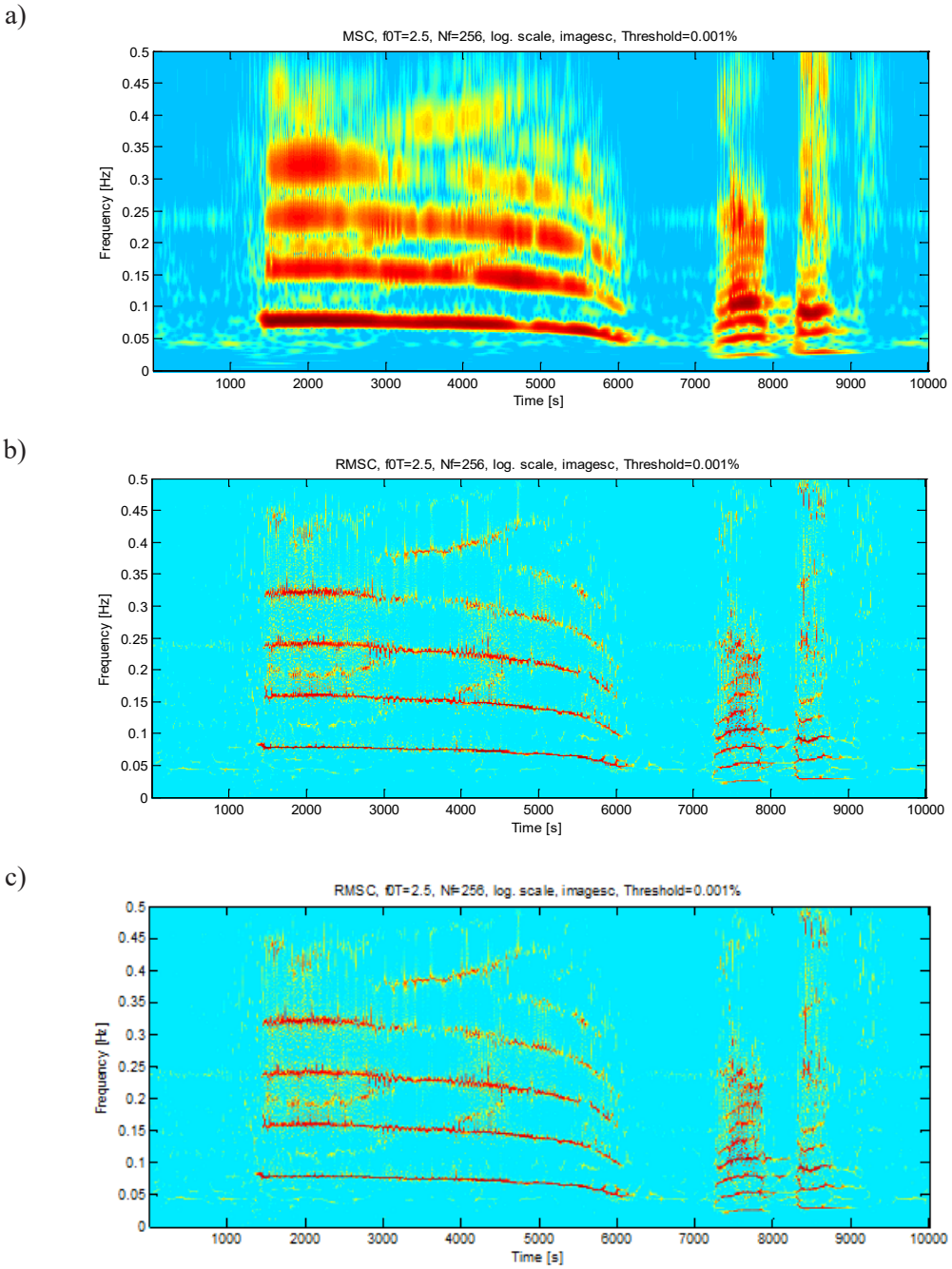


Fig. 5. Example of representation with affine class: a) Morlet scalogram, b) Affine Smoothed Wigner distribution, and c) reassignment version of Morlet scalogram

The use of such TFR representations, among others in medical or technical diagnostics, can contribute to more effective detection of disease syndromes and malfunctions. However, due to the complexity of the numerical operations involved in creating these representations (especially their reassignment), there is a need to develop (optimize) computational algorithms that will shorten the analysis time and allow for online mode application.

REFERENCES

1. Arjmandi M.,K., Pooyan M. An optimum algorithm in pathological voice quality assessment using wavelet-packet-based features, linear discriminant analysis and support vector machine, Biomedical Signal Processing and Control, 7 (1), 2012, 3–19.
2. Auger F. Representations temps-frequence des signaux nonstationnaires: synthese et contributions. PhD thesis, Ecole Centrale de Nantes, France 1991.

3. Auger F., Flandrin P. Improving the Readability of Time-Frequency and Time-Scale Representations by the Reassignment Method. IEEE Transactions on Signal Processing, 43(5), 1995, 1068–89.
4. Auger F., Flandrin P., Goncalves P., Lemoine O. Time-Frequency Toolbox. User Guide, 1995–1996.
5. Behroozmand R., Almasganj F. Optimal selection of wavelet-packet based features using genetic algorithm in pathological assessment of patients' speech signal with unilateral vocal fold paralysis, Computers in Biology and Medicine, 37(4), 2007, 474–485.
6. Bertrand J., Bertrand P. A class of affine Wigner functions with extended covariance properties. J. Math. Phys., 33(7), 1992, 2515–2527.
7. Bertrand J., Bertrand P. Affine time-frequency distributions, Boashash B., Ed., Time-Frequency Signal Analysis – Methods and Applications, Longman Cheshire, Melbourne (Australia), 1992, 118–140.
8. Bertrand J., Bertrand P. Representation des signaux a large bande, La Recherche Aérospatiale, 5, 1985, 277–283.
9. Boudreaux-BAartels G. F. Mixed time-frequency signal transformations, Poularikas A., Ed., The Transforms and Applications Handbook, IEEECRC, Press, Boca Raton, FL,1995, 887–962.
10. Braunschweig T., Thomä R. S., Trautwein U., Wittenberg T. Time-Frequency Analysis of Vocal Fold's Onset, IEEE Instrumentation and Measurement Technology Conference Ottawa, Canada, 2, 1997, 516–521.
11. Carvalho R.,T.,S., Cavalcante C.,C., Cortez P.,C. Wavelet Transform and Artificial Neural Networks Applied to Voice Disorders Identification, Third World Congress on Nature and Biologically Inspired Computing (NaBIC), 2011, 371–376.
12. Cheng Jun, Liu Feng, Yi Kechu. Time frequency representations for the analysis of speech signals, Conference Paper. TENCON '93. Proceedings. Region 10 Conference on Computer, Communication, Control and Power Engineering, IEEE, 1993.
13. Claasent T. A. C. M., Mecklenbrauker W. F. G. The Wigner distribution – A tool for time-frequency analysis. Philips J. Res., 35 (3), 1980, 217–250.
14. Claasent T. A. C. M., Mecklenbrauker W. F. G. The Wigner distribution – A tool for time-frequency analysis. Philips J. Res., 35 (4/5), 1980, 276–300.
15. Claasent T. A. C. M., Mecklenbrauker W. F. G. The Wigner distribution – A tool for time-frequency analysis. Philips J. Res., 35 (6), 1980, 372–389.
16. Cohen L. Generalized phase-space distribution functions. J. Math. Phys., 7, 1966, 781–786.
17. Cohen L. Time-Frequency Analysis, Prentice Hall, Englewoods Cliffs, NJ, 1995.
18. Cohen L. Time-Frequency Distributions – A Review. Proceedings of the IEEE, 77(7), 1989, 941–948.
19. Davies M.E., Daudet L. Sparse audio representations using the MCLT, Signal Processing. 86(3), 2006, 457–470.
20. Dennis J., Tran H. D., Li H. Spectrogram image feature for sound event classification in mismatched conditions., IEEE Signal Processing Letters, 18, (2), 2011, 130–133.
21. Farooq O., Datta S. Phoneme recognition using wavelet based features, Information Sciences, 150(1–2), 2003, 5–15.
22. Flandrin P. Temps-fréquence, Hermes, Paris, 2nd edition, 1998.
23. Flandrin P. Time-Frequency/Time-Scale Analysis, Academic Press, San Diego, CA, 1999.
24. Gardner T. J., Magnasco M. O. Sparse time-frequency representations, PNAS, Proceedings of the National Academy of Sciences of the United States of America. 103(16), 2006, 6094–6099.
25. Ghoraani B., Krishnan S. Time–Frequency Matrix Feature Extraction and Classification of Environmental Audio Signals, IEEE Transactions on Audio, Speech, and Language Processing, 19(7), 2011, 2197–2209.
26. Glowacz A. Diagnostics of direct current machine based on analysis of acoustic signals with the use of symlet wavelet transform and modified classifier based on words, Maintenance and Reliability. 16(4), 2014, 554–558.
27. Glowacz A., Glowacz Z. Diagnosis of stator faults of the single-phase induction motor using acoustic signals. Applied Acoustics. 117, 2017, 20–27
28. Glowacz A., Glowacz Z. Recognition of rotor damages in a DC motor using acoustic signals, Bulletin of the Polish Academy of Sciences Technical Sciences, 65(2), 2017, 187–194.
29. Gonalvs P. and Baraniuk R. Pseudo affine wigner distributions and kernel formulation, Submitted to IEEE Transactions on Signal Processing, 1996.
30. Grossmann A., Morlet J. Decomposition of Hardy functions into square integrable wavelets of constant shape, SIAM J. Math. Anal., 15(4), 1984, 723–736.
31. Hibare R., Vibhute A. Feature Extraction Techniques in Speech Processing: A Survey, International Journal of Computer Applications, 107 (5), 2014, 1–8.
32. Hlawatsch F., Auger F. Time-Frequency Analysis, Digital Signal and Image Processing. ISTE Ltd and Wiley & Sons, Inc 2008.
33. Hlawatsch F., Bourdox-Bartels G. F. Linear and quadratic time-frequency signal representations, IEEE Signal Process. Mag., 9(2), 1992, 21–67.
34. Huzaifah M. Comparison of Time-Frequency Representations for Environmental Sound Classifica-

tion using Convolutional Neural Networks, [Online]. Available: https://arxiv.org/abs/1706.07156.

35. Kern A., Nagy O., Stoop R. Sparse Time-Frequency Analysis of Speech Signals, International Symposium on Nonlinear Theory and its Applications (NOLTA2005) Bruges, Belgium, 2005, 545–548.
36. Krolczyk G.M., Nieslony P. Maruda R.W., Wojciechowski S. Dry cutting effect in turning of a duplex stainless steel as a key factor in clean production. Journal of Cleaner Production, 142, 2017, 3343–3354.
37. Markaki M., Stylianou Y. Voice Pathology Detection and Discrimination Based on Modulation Spectral Feature, IEEE Transactions on audio, speech, and language processing, 19(7), 2011, 1938–1948.
38. Mergu Rohini R., Dixit Shantanu K. Multi-Resolution Speech Spectrogram, International Journal of Computer Applications, 15 (4), 2011, 28–32.
39. Nieslony P., Krolczyk G.M., Wojciechowski S., Chudy R., Zak K., Maruda R.W. Surface quality and topographic inspection of variable compliance part after precise turning, Applied Surface Science, 434, 2018, 91–101.
40. Nieslony P., Krolczyk G.M., Zak K., Maruda R.W., Legutko S. Comparative assessment of the mechanical and electromagnetic surfaces of explosively clad Ti–steel plates after drilling process. Precision Engineering, 47, 2017, 104–110.
41. Parraga A. Application of wavelet packet transform in the analysis and classification of pathological signs and voices, Master's thesis, Federal University of Rio Grande do Sul, Brazil, School of Engineering, 2002.
42. Rioul O., Flandrin P. Time-scale distributions: A general class extending wavelet transform. IEEE Trans. Signal Process., 40(7), 1992, 1746–1757.
43. Saenz-Lechon N., Godino-Llorente J. I., Osma-Ruiz V., Gómez-Vilda P. Methodological issues in the development of automatic systems for voice pathology detection, Biomedical Signal Processing and Control, 1(2), 2006, 120–128.
44. Trivedi N., Kumar V., Singh S., Ahuja S., Chadha R. Speech Recognition by Wavelet Analysis, International Journal of Computer Applications, 15(8), 2011, 27–32.
45. Umapathy K., Krishnan S., Parsa V., Jamieson D. G. Discrimination of pathological voices using time-frequency approach, IEEE Trans. Biomed. Eng., 52(3), 2005, 421–430.