Postprint of: Abratkiewicz K., Czarnecki K., Fourer D., Auger F., Estimation of time-frequency complex phase-based speech attributes using narrow band filter banks, W: Signal Processing Symposium (SPSympo), 12-14.09.2017, Jachranka, Poland, DOI: 10.1109/SPS.2017.8053673

Estimation of time-frequency complex phase-based speech attributes using narrow band filter banks

Karol Abratkiewicz*, Krzysztof Czarnecki†, Dominique Fourer§, François Auger‡

*Gdansk University of Technology, Faculty of Electronics, Telecommunications and Informatics
G.Narutowicza 11/12, 80-980 Gdansk, Poland, Email: karol.abratkiewicz@gmail.com

†Gdansk University of Technology, Faculty of Electronics, Telecommunications and Informatics
G.Narutowicza 11/12, 80-980 Gdansk, Poland, Email: czarnecki.krzysiek@gmail.com

§UMR STMS (IRCAM - CNRS - UPMC), Paris, France, Email: dominique@fourer.fr

‡LUNAM University, IREENA, Saint-Nazaire, France, Email: francois.auger@univ-nantes.fr

***Abstract*—In this paper, we present nonlinear estimators of nonstationary and multicomponent signal attributes (parameters, properties) which are instantaneous frequency, spectral (or group) delay, and chirp-rate (also known as instantaneous frequency slope). We estimate all of these distributions in the time-frequency domain using both finite and infinite impulse response (FIR and IIR) narrow band filers for speech analysis. Then, we present few examples including a novel type of imaging joining energy and phase acceleration in a single picture. Finally, we provide an open-source project – ccROJ – Time-Frequency C++ Framework of which we are authors and that is used for computing the presented figures.**

***Index Terms*—ccROJ, time-frequency reassignment and synchrosqueezing, STFT, instantaneous frequency, group delay, chirp-rate, FIR and IIR filters.**

I. Introduction

The time-frequency approach is mainly dedicated to the analysis of multicomponent and nonstationary signals [1], [2], [3], [4], [5], [6]. The human's speech is a representative example of them and many authors work in this area [7], [8], [9], [10], [11]. In the paper, we are also focused on this kind of signals. In general, the time-frequency analysis allows to separate signal components. Unfortunately, this separation has some limitations, which are well defined by the Heisenberg-Gabor uncertainty principle [12]. The time-frequency resolution can be controlled by the effective duration of impulse response of used filters (or analyzing windows if the Fourier transformation is applied). In both cases, the short-time Fourier transform (STFT) is obtained as a result of filtering or transforming. Then, the time-frequency distributions such as instantaneous frequency (IF), group delay (GD) [2], [3], instantaneous bandwidth (IB) [13], and chirp-rate (CR; also known as instantaneous frequency slope) can be estimated based on derivatives of the STFT complex phase [17], as it is shown in the next section of this paper.

In [1], we proposed several new estimators designated to operate in the time-frequency domain for digital signal analysis using a special recursive filtering. The main purpose of these tools is the estimation of IF and CR. Here, we extend this approach using both FIR and IIR filter banks for speech analysis, as well as we investigate other distributions such as GD [2], [3]. The FIR filters are implemented using Fast Fourier Transform (FFT). Then, the IIR filters are realized in the same way as in [1] based on a recursive implementation. Our other motivation is to demonstrate the use of an open-source project – ccROJ, that is a time-frequency C++ framework of which we are authors. Therefore, a short description of the project and a simple code listing are introduced further in this paper.

In the next section, definitions and interpretations of the considered signal properties are introduced. In Section III, many illustrative examples are presented. In Section IV, we introduce a brief description and an example of code listing of ccROJ project which contains the implementation of the presented methods in C++ programming language [14].

II. Phase-based attributes

The STFT of speech signal $x(t)$ can be defined as the convolution product of this signal with an analyzing narrow band filter $g(t,\omega)$ at center angular frequency ω:

$$\begin{aligned} y_x^g(t,\omega) = (x \star g)(t,\omega) = A_x^g(t,\omega)\mathrm{e}^{\mathrm{j}\phi_x^g(t,\omega)} = \\ = \int_{\mathbb{R}} x(\tau) g(t-\tau,\omega)\, d\tau \end{aligned} \tag{1}$$

where j is the imaginary unit; $\mathrm{j}^2 = -1$; $A_x^g(t,\omega)$ and $\phi_x^g(t,\omega)$ denote, respectively, the amplitude and the phase. $g(t,\omega)$ represents both FIR and IIR filters, and can be defined as their modulated envelope (or an analysing window):

$$g(t,\omega) = h(t)\exp(j\omega t), \tag{2}$$

where $h(t)$ is the envelope (or the window; for example Blackman-Harris window). According to [17], the complex phase of this transform can be defined by:

$$\Phi_x^g(t,\omega) = \ln\left(y_x^g(t,\omega)\right) = \Lambda_x^g(t,\omega) + \mathrm{j}\phi_x^g(t,\omega), \tag{3}$$

where $\Lambda_x^g(t,\omega) = \ln\big(A_x^g(t,\omega)\big)$ is its level. This time-frequency signal representation can be used to estimate many physical signal parameters. Firstly, GD in the time-frequency domain can be defined as the partial derivative of the STFT phase with respect to frequency [2]:

$$D_x^g(t,\omega) = -\frac{\partial \Im \Phi_x^g(t,\omega)}{\partial \omega} = -\frac{\partial \phi_x^g(t,\omega)}{\partial \omega}. \tag{4}$$

Then, IF can be defined in similar manner [2]:

$$\Omega_x^g(t,\omega) = \frac{\partial \Im \Phi_x^g(t,\omega)}{\partial t} = \frac{\partial \phi_x^g(t,\omega)}{\partial t} \tag{5}$$

as the partial derivative with respect to time. There is no necessity to use the Hilbert transformation as it is presented in [15], [16] as long as we consider IF in a narrow frequency band. Both, GD and IF are used for time-frequency reassignment in order to relocate the energy in the time-frequency plane as follows [2]:

$$\Sigma_x^g(t,\omega) = \iint_{\mathbb{R}^2} E_x^g(\tau,\upsilon)\delta\big(t-\Gamma_x^g(\tau,\upsilon)\big)\delta\big(\omega-\Omega_x^g(\tau,\upsilon)\big)\mathrm{d}\tau\mathrm{d}\upsilon, \tag{6}$$

where $\Gamma_x^g(t,\omega)$ is the reassigned time defined as [5]:

$$\Gamma_x^g(t,\omega) = t + D_x^g(t,\omega), \tag{7}$$

the spectral energy is given by:

$$E_x^g(t,\omega) = \big(A_x^g(t,\omega)\big)^2 = |y_x^g(t,\omega)|^2, \tag{8}$$

and $\delta(\cdot)$ denotes the Dirac distribution. As the result, the new concentrated energy distribution $\Sigma_x^g(t,\omega)$ is estimated. This time-frequency estimate is characterized by comparatively high concentration of energy near attractors (ridges, partials).

Alternatively, the IF estimate (5) can be used in time-frequency synchrosqueezing [18], [19], [20], which can be defined for the STFT by [6]:

$$Y_x^g(t,\omega) = \int_{\mathbb{R}} y_x^g(t,\rho)e^{-j\rho t_0}\delta(\omega-\Omega_x^g(t,\rho))\,d\rho \tag{9}$$

It corresponds to the signal reconstruction formula [6]:

$$x(t-t_0) = \frac{1}{h(t_0)}\int_{\mathbb{R}} y_x^g(t,\omega)e^{-j\omega t_0}\frac{d\omega}{2\pi}. \tag{10}$$

Therefore, the synchrosqueezed STFT can be used to signal recovery as its simple integration [6]:

$$\hat{x}(t-t_0) = \frac{1}{h(t_0)}\int_{\Omega} Y_x^g(t,\omega)\frac{d\omega}{2\pi}, \tag{11}$$

where the integration area Ω can be restricted to the frequency support of the signal (or of its component) and $h(t_0) \neq 0$.

The level of STFT $\Lambda_x^g(t,\omega)$ can also be used for a local CR estimation. In [4], the following estimator of this parameter is proposed:

$$R_x^g(t,\omega) = -\frac{\partial \Lambda_x^g(t,\omega)}{\partial t}\Big/\frac{\partial \Lambda_x^g(t,\omega)}{\partial \omega}. \tag{12}$$

In [13], IB is defined as the absolute value of the instantaneous signal level derivative with respect to time, therefore $\partial\Lambda_x^g(t,\omega)/\partial t$ in Eq. (12) can be, more or less, interpreted in similar manner. Then, dually, $\partial\Lambda_x^g(t,\omega)/\partial\omega$ should have a dimension of time duration, therefore it can be referred to as "group duration". This approach is consistent with the definition of the CR of any LFM chirp, which is the ratio of the bandwidth to the time duration of this chirp.

In [1], the estimator (12) is effectively extended to an infinite number of estimators. We consider here only two of them, which, however, are very representative, namely:

$$\grave{R}_x^g(t,\omega) = -\frac{\partial^2 \Lambda_x^g(t,\omega)}{\partial t^2}\Big/\frac{\partial^2 \Lambda_x^g(t,\omega)}{\partial \omega \partial t} \tag{13}$$

as well as

$$\acute{R}_x^g(t,\omega) = -\frac{\partial^2 \Lambda_x^g(t,\omega)}{\partial \omega \partial t}\Big/\frac{\partial^2 \Lambda_x^g(t,\omega)}{\partial \omega^2}. \tag{14}$$

The estimators defined by Eqs. (13) and (14) are not sensitive to any slow amplitude modulation. Then, the time-frequency IF estimator can be improved using new CR estimators. For example, a new unbiased IF estimator for Eq. (12) can be obtained as follows:

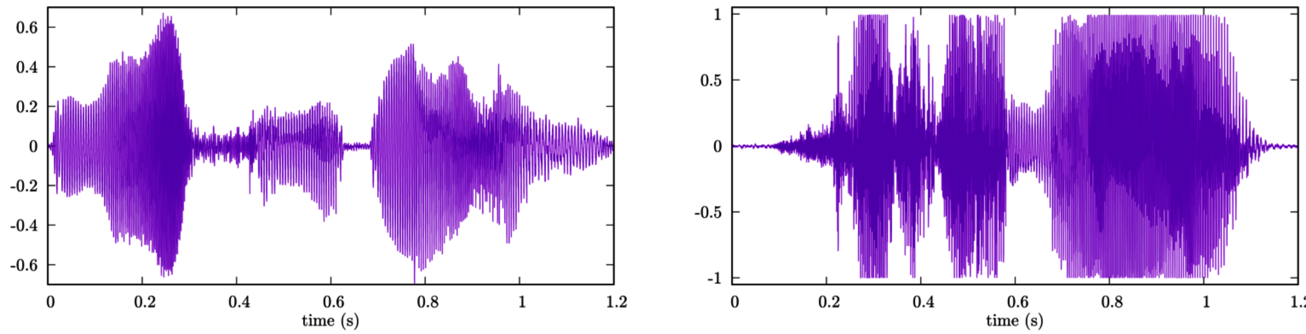
$$\begin{aligned}\hat{\Omega}_x^g(t,\omega) =& \Omega_x^g(t,\omega) + D_x^g(t,\omega)R_x^g(t,\omega) = \\ =& \frac{\partial \phi_x^g(t,\omega)}{\partial t} + \frac{\partial \phi_x^g(t,\omega)}{\partial \omega}\frac{\partial \Lambda_x^g(t,\omega)}{\partial t}\Big/\frac{\partial \Lambda_x^g(t,\omega)}{\partial \omega},\end{aligned} \tag{15}$$

what is also proposed in [1]. Finally, the synchrosqueezing (9) can be improved by replacing $\Omega_x^g(t,\omega)$ with $\hat{\Omega}_x^g(t,\omega)$. This allows to obtain more precise results.

III. Numerical Examples

In this section, we present results of the analysis of 2 speech signals. The first one being considered is a recording sampled at the rate equal to 22050 Sa/s in which an adult human female pronounces the sentence: *"You've got mail"*. This signal is analyzed using the windowed FFT with the Blackman-Harris window, whose width is equal to 22.05 ms. In the second one, which is sampled at the rate equal to 11025 Sa/s, an adult human male pronounces *"Sacre Bleu!"*. This recording is clearly distorted and is analyzed using the recursive ODE filters, whose time spread and order are equal to, respectively, 5 ms and 5. Both are represented with a 16 bits of precision using floating point numbers.

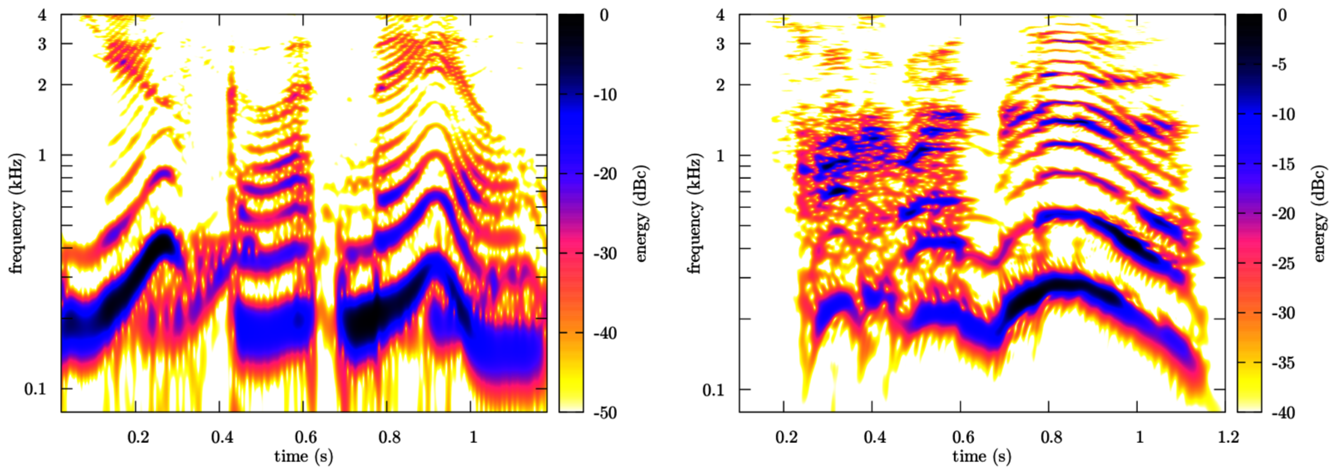
In Fig. 1, the speech signals are presented in the time domain. Next figures depict the distributions in the time-frequency domain. In Figs. 2 and 3, respectively, the classical and the reassigned energy distributions are shown. The reassigned energy is mapped close to the attractors associated with each component [5]. It results in a high energy concentration. In Fig. 4, CR estimates obtained using Eqs. (13) and (14) are presented. The color corresponds to estimated values according to the associated color boxes. Some improvements of this imaging are the combination of both CR and energy level in a single picture. In this approach the color denotes the CR while the saturation corresponds to the energy level. The results are presented in Fig. 5.



(a) first signal

(b) second signal

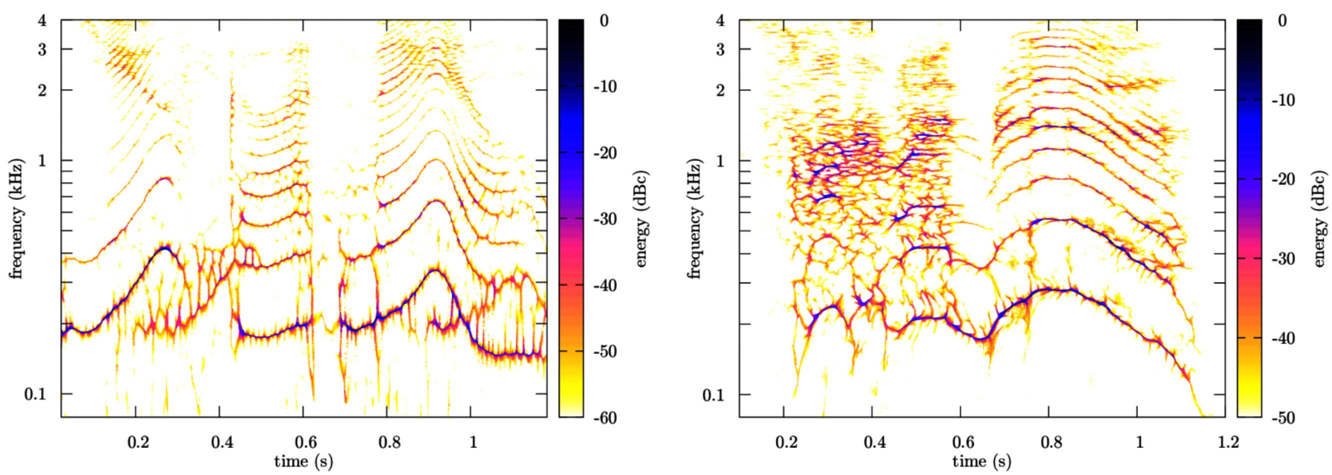
Fig. 1. Signals in the time domain.



(a) first signal

(b) second signal

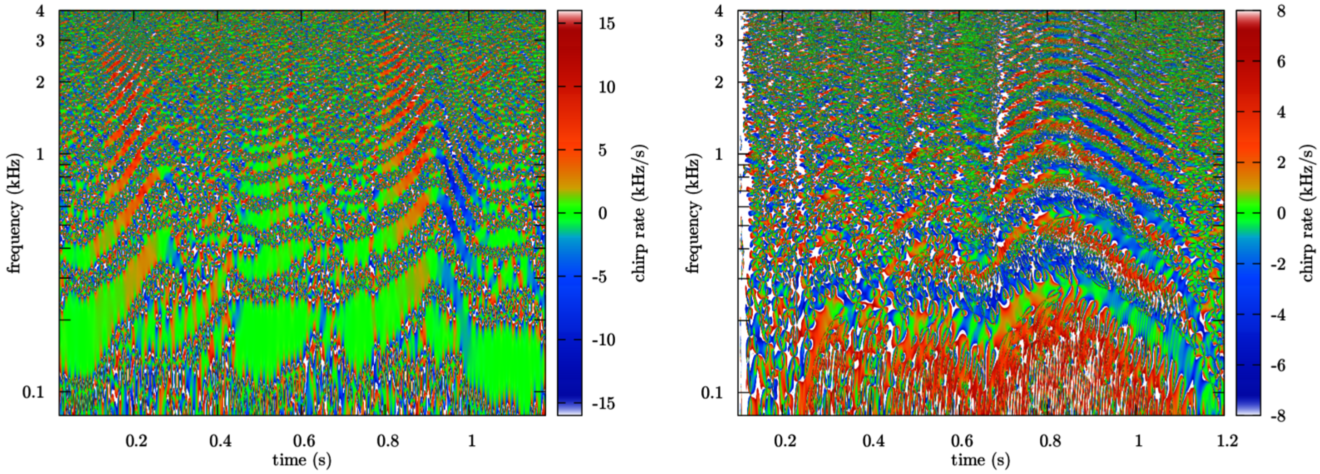
Fig. 2. Classical spectrograms.



(a) first signal

(b) second signal

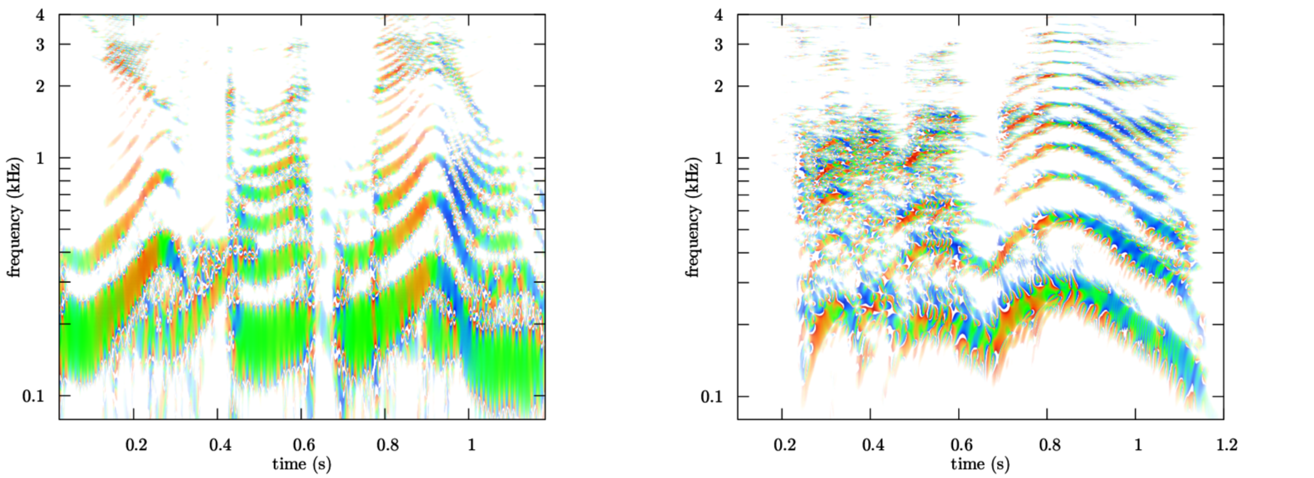
Fig. 3. Time-frequency reassigned spectrograms.



(a) first signal

(b) second signal

Fig. 4. Phase accelerograms in the time-frequency domain.



(a) first signal

(b) second signal

Fig. 5. Compositions of the spectrogram and phase accelerogram.

IV. ccROJ PROJECT

The all estimators presented in this paper are implemented as a part of the open source ccROJ project [14], of which we are authors. This project provides efficient algorithms written in C++ and dedicated to GNU/Linux operating system for FIR as well IIR filter analyzing banks. Overall, this framework provides the following functionalities.

- Signal time-frequency decomposition using filter banks and FFT including the chirplet transform [21], [22].
- Estimators of time-frequency distributions (including spectral energy, IF, GD, and CR).
- Reassignment functions operating in various domains.
- Fast signal reconstruction based on its STFT including the improved time-frequency synchrosqueezing.
- Pulses, noise, filter, and window generators.
- Signal, image, and array data objects including configuration structures. They contain many useful methods for data manipulating, transforming, saving, etc. as well as operator overloading.
- Fourier, Hilbert, and time-frequency Hough transform.
- median filtering in the time-frequency domain including an efficient implementation.
- Signal processing procedures in the time domain including convolution and correlation.
- Test programs for tutoring and scripts for result drawing.

A. Sample program

Let us consider a simple demonstrative program in C++ language which uses the ccROJ framework in version 0-40 (Listing. 1). At the beginning, an LFM chirp signal lasts 1 second is generated (Listing. 2). Then, an ODE analyzer which consists of 5 IIR filter banks is defined (Listing. 3). Finally, the classical spectrogram and the phase accelerogram are estimated and saved to files (Listing. 5). The resultant data can be plotted using attached *Gnuplot* scripts. Comments are highlighted in green, elements of ccROJ, strings, and C++ keywords are colored, respectively, in red, magenta, and blue.

```
print_roj_info ();   /* print ccROJ header */

/* check ccROJ version */
require_roj_version (0, 40);
```

Listing 1. The version checking.

```
/* roj_signal_config is a structure dedicated to
   define basic signal parameters: */
roj_signal_config sig_conf;
sig_conf.rate = 1000.0; /* sampling rate */
sig_conf.length = 1000; /* number of samples */
sig_conf.start = -0.5; /* initial instant */

/* new empty signal */
roj_complex_signal* signal_ptr = new
  roj_complex_signal (sig_conf);

/* signal generation */
double crate = 1000.0; /* chirp-rate */
for(int n=0; n<sig_conf.length; n++){

  /* time instant */
  double t = sig_conf.start
   + (double)n/sig_conf.rate;

  /* M_PI and pow() are elements of math.h */
  double arg = M_PI * pow(t, 2.0) * crate;

  /* field m_waveform gives access to signal
  samples; cexp and 1I are from complex.h */
  signal_ptr->m_waveform[n] = cexp(1I * arg);
} /* now signal is ready to use */
```

Listing 2. LFM signal generation.

```
/* universal filter generator creation */
roj_filter_generator *filter_gen = new
  roj_filter_generator (sig_conf.rate);

/* filter generator configuration */
filter_gen->set_type(ROJ_ODE_FILTER); /* code */
filter_gen->set_spread(0.01); /* time spread */
filter_gen->set_order(5); /* filter order */

/* parameters for setting filter distribution
   along frequency axis (evenly) */
roj_array_config arr_conf;
arr_conf.min = -500.0; /* minimal frequency */
arr_conf.max = 500.0; /* maximal frequency */
arr_conf.length = 1000; /* number of filters */

/* create TF filter analyzer, frequency axis
  configuration and pointer to filter generator
  are given as its constructor arguments */
roj_ode_analyzer *analyzer_ptr = new
  roj_ode_analyzer(arr_conf, filter_gen);

/* signal setting to ODE analyzer */
int hopsize = 5; /* hopsize in samples */
analyzer_ptr->set_signal(signal_ptr, hopsize);
/* now the analyzer is ready to work */
```

Listing 3. ODE analyzer definition.

```
/* estimation of spectral energy and storing
   them in roj_real_matrix object */
roj_real_matrix *se_ptr =
  analyzer_ptr->get_spectral_energy();

/* save estimate to text file */
se_ptr->save("spectral-energy.txt");

/* estimation of chirp-rate and storing
   them in roj_real_matrix object */
roj_real_matrix *cr_ptr =
  analyzer_ptr->get_chirp_rate(CR_D_ESTIMATOR);

/* get matrix configuration as roj_image_config*/
roj_image_config img_conf = cr_ptr->get_config();

for(int n=0; n<img_conf.time.length; n++){
  /* get time of n-th instant */
  double t = cr_ptr->get_time_by_index(n);

  for(int k=0; k<img_conf.frequency.length; k++){
    /* get center frequency of k-th channel */
    double f = cr_ptr->get_frequency_by_index(n);

    /* get estimated values and print to stdout */
    double cr = cr_ptr->m_data[n][k];
    printf("%g\t%g\t%g\n", t, f, cr);
  }
}
```

Listing 4. Energy and chirp-rate estimation, saving, and printing.

```
/* the memory should be cleaned
   if the program does not end */
delete analyzer_ptr;

/* these objects can be removed after analyzer
   initiation and signal introduction   */
delete signal_ptr, filter_gen;

/* the resultant objects can
   also be removed in this way:  */
delete se_ptr, cr_ptr;
```

Listing 5. The memory cleaning.

V. Conclusion

In this paper, we have presented many estimators of signal attributes such as: IF, GD, and CR, which are obtained using narow band filter banks and based on STFT complex phase. We have investigated them in the context of the analysis of human's speech which is an example of nonstationary and multicomponent signal. Moreover, we have shown how these estimates can be used for time-frequency reassignment and synchrosqueezing.

Then, we have introduced the open source ccROJ project [14], whose we are authors. This project contains the C++ implementation of all algorithms presented in this paper. We have presented a simple sample program. We have also introduced a novel type of imaging joining energy level and phase acceleration in a single picture. This method is also attached as a *Gnuplot* script to ccROJ project. Our future plans in this field cover the use of the presented estimators for an adaptive analysis in the time-frequency domain as well as the continuation of ccROJ project maintenance and development.

Acknowledgment

The authors would like to thank dr. Mirosław Rojewski for his involvement in our research related to signal analysis in the time-frequency domain.

References

[1] D. Fourer, F. Auger, K. Czarnecki, S. Meignen, and P. Flandrin, *Chirp rate and instantaneous frequency estimation: application to recursive vertical synchrosqueezing*, IEEE Signal Processing Letters, 2017.
[2] K. Kodera, C.D. Villedary, R. Gendrin, *A new method for the numerical analysis of non-stationary signals*, Physics of the Earth and Planetary Interiors, vol. 12, no. 2–3, pp. 142–150, 1976.
[3] B. Boashash, Time-frequency signal analysis and processing: a comprehensive reference, Academic Press, 2015.
[4] K. Czarnecki, *The instantaneous frequency rate spectrogram*, Mechanical Systems and Signal Processing, vol.66–67, pp. 361–373, 2016.
[5] F. Auger, P. Flandrin, *Improving the readability of time-frequency and time-scale representations by the reassignment method*, IEEE Transactions on Signal Processing, vol. 43, no. 5, pp. 1068–1089, 1995.
[6] D. Fourer, F. Auger, P. Flandrin, *Recursive versions of the Levenberg-Marquardt reassigned spectrogram and of the synchrosqueezed STFT*, 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 4880–4884, 2016.
[7] D. Nelson, *Cross-spectral methods for processing speech*, The Journal of the Acoustical Society of America, vol. 110, no. 5, pp. 2575–2592, 2001.
[8] S.A. Fulop, K. Fitz, *Algorithms for computing the time-corrected instantaneous frequency (reassigned) spectrogram, with applications*, The Journal of the Acoustical Society of America, vol. 119, no. 1, pp. 360–371, 2006.
[9] T. Oberlin, S. Meignen, V. Perrier, *Second-Order Synchrosqueezing Transform or Invertible Reassignment? Towards Ideal Time-Frequency Representations*, IEEE Transactions on Signal Processing, vol. 63, no. 5, pp. 1335–1344, March, 2015.
[10] D. Wang and J. H. L. Hansen, *F0 estimation for noisy speech by exploring temporal harmonic structures in local time frequency spectrum segment*, proc. 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 6510-6514, 2016.
[11] G. Tryfou, M. Omologo, *A reassigned based singing voice pitch contour extraction method*, proc. 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 321–325, 2017.
[12] D. Gabor, *Theory of communication. Part 1: The analysis of information*, Journal of the Institution of Electrical Engineers – Part III: Radio and Communication Engineering, vol. 93, no. 26, pp. 429–441, 1946.
[13] A.E. Barnes, *The calculation of instantaneous frequency and instantaneous bandwidth*, Geophysics, vol. 57, no. 11, pp. 1520–1524, 1992.
[14] F. Auger, K. Czarnecki, D. Fourer, *ccROJ - Time-Frequency C++ Framework*, SoftwareX, 2017. (paper is submitted) Project webpage: https://github.com/dsp-box/ccROJ
[15] B. Boashash, *Estimating and Interpreting The Instantaneous Frequency of a SignalPart 1: Fundamentals*, Proceedings of the IEEE vol. 80, no. 4, pp. 520–538, 1992.
[16] S.L Hahn, Hilbert transforms in signal processing, Artech House, 1996.
[17] S.L Hahn, *On the uniqueness of the definition of the amplitude and phase of the analytic signal*, Signal Processing, vol. 83, no. 8, pp. 1815–1820, August, 2003.
[18] F. Auger, P. Flandrin, Y. T. Lin, S. McLaughlin, S. Meignen, T. Oberlin, H. T. Wu, *Time-Frequency Reassignment and Synchrosqueezing: An Overview*, IEEE Signal Processing Magazine, vol. 30, no. 6, pp. 32–41, November, 2013.
[19] S. Meignen, T. Oberlin, and S. Mclaughlin, *A new algorithm for multicomponent signal analysis based on Synchrosqueezing: with an application to signal sampling and denoising*, IEEE Transactions on Signal Processing, vol. 60, no. 11, pp. 5787–5798, 2012.
[20] I. Daubechies, S. Maes, *A nonlinear squeezing of the continuous wavelet transform based on auditory nerve models*, Wavelets in Medicine and Biology, pp. 527–546, 1996.
[21] S. Mann, S. Haykin, *The Chirplet Transform: Physical Considerations*, IEEE Transactions on Signal Processing, vol. 43, no. 11, pp. 2745–2761, November, 1995.
[22] K. Czarnecki, D. Fouan, Y. Achaoui, S. Mensah, *Fast bubble dynamics and sizing*, Journal of Sound and Vibration, vol. 356, pp. 48–60, 2015.