Performance evaluation of the Hilbert-Huang transform for respiratory sound analysis and its application to continuous adventitious sound characterization

Manuel Lozano [a,c,*], José Antonio Fiz [a,b,d], Raimon Jané [a,d,e]

[a] Institute for Bioengineering of Catalonia (IBEC), Baldiri Reixac, 4, Tower I, 9th floor, 08028, Barcelona, Spain

[b] Pulmonology Service at Germans Trias i Pujol University Hospital, Ctra. de Canyet, s/n, 08916, Badalona, Spain

[c] Health Sciences Research Institute of the Germans Trias i Pujol Foundation (IGTP), Ctra. de Can Ruti, Camí de les escoles, s/n, 08916, Badalona, Spain

[d] Biomedical Research Networking Center in Bioengineering, Biomaterials and Nanomedicine (CIBER-BBN)

[e] Department of Automatic Control (ESAII), Universitat Politècnica de Catalunya (UPC), Barcelona, Spain

ABSTRACT

The use of the Hilbert-Huang transform in the analysis of biomedical signals has increased during the past few years, but its use for respiratory sound (RS) analysis is still limited. The technique includes two steps: empirical mode decomposition (EMD) and instantaneous frequency (IF) estimation. Although the mode mixing (MM) problem of EMD has been widely discussed, this technique continues to be used in many RS analysis algorithms.

In this study, we analyzed the MM effect in RS signals recorded from 30 asthmatic patients, and studied the performance of ensemble EMD (EEMD) and noise-assisted multivariate EMD (NA-MEMD) as means for preventing this effect. We propose quantitative parameters for measuring the size, reduction of MM, and residual noise level of each method. These parameters showed that EEMD is a good solution for MM, thus outperforming NA-MEMD. After testing different IF estimators, we propose Kay's method to calculate an EEMD-Kay-based Hilbert spectrum that offers high energy concentrations and high time and high frequency resolutions. We also propose an algorithm for the automatic characterization of continuous adventitious sounds (CAS). The tests performed showed that the proposed EEMD-Kay-based Hilbert spectrum makes it possible to determine CAS more precisely than other conventional time-frequency techniques.

1. Introduction

Respiratory sounds (RS) are multicomponent, nonlinear, and non-stationary signals. In general, RS signals are comprised of normal RS and may contain superimposed abnormal RS, such as continuous (CAS) and discontinuous (DAS) adventitious sounds, as well as different types of noise, such as clicks, background talking, or heart sounds. Normal RS are random in nature, whereas CAS are quasi-periodic waveforms with a duration of more than 80-100 ms and a fundamental frequency of over 100 Hz, and DAS are transient and short sounds (around 20 ms), with high frequency components (above 300 Hz) [1,2]. Therefore, RS are complex signals made up of a set of components, each one having different time-frequency features.

Due to the different and variable characteristics of RS, time-frequency distributions (TFDs) have become the

*Corresponding author at: Institute for Bioengineering of Catalonia (IBEC), Baldiri Reixac, 4, Tower I, 9th floor, 08028 Barcelona, Spain. Tel.: þ34 695119250.
E-mail addresses: mlozanogar@gmail.com (M.Lozano), jafiz@msn.com (J.A.Fiz), rjane@ibecbarcelona.eu (R.Jané).

the most commonly used and straightforward techniques for RS characterization. In CAS analysis, spectrogram has been the most widely used TFD [3-7], despite its poor and window-dependent resolution. Nevertheless, more advanced TFDs have recently been proposed for CAS analysis, either through combining wavelet decomposition with third order spectra features [8], or by deriving a temporal-spectral dominance spectrogram from the short-time Fourier transform [9].

As opposed to CAS analysis, DAS analysis requires TFDs with higher time resolution than spectrogram. Wavelet-based techniques, such as scalogram, have been widely used for DAS detection [10-12]. In addition, DAS have also been analyzed by means of nonlinear techniques, such as kurtosis and fractal dimension as measures of gaussianity and complexity, respectively [13-15].

Besides Fourier and wavelet-based techniques, one of the most relevant parameters of time-frequency analysis for nonlinear and non-stationary signals, such as RS, is the instantaneous frequency (IF), which consists of the frequency content of a signal at each time instant [16]. The concept of IF has led to the definition of TFDs that highly concentrate the energy of a signal along its IF, which makes it possible to identify signal components more precisely. Several IF estimators have been proposed, such as the phase derivative of the analytic signal associated with a real signal [17,18], zero-crossing [19], or adaptive IF estimators based on data modeling [20-22]. However, the most common IF estimators are based on TFDs [16,23], which give IF estimates with lower variance.

Quadratic TFDs, such as the Wigner-Ville distribution (WVD), were defined based on the IF concept with the aim of improving the resolution and concentration of energy of spectrogram [16]. However, a major drawback of the WVD is the presence of cross-terms, which complicate IF estimation, especially in multicomponent signals. In order to reduce cross-terms, other quadratic TFDs have been proposed as filtered versions of the WVD, using different time-frequency smoothing kernels, such as the smoothed pseudo-WVD [24] or the reduced interference distributions [25], which reduce cross-terms while maintaining high resolution.

In addition to these smoothing approaches for cross-term reduction, other techniques have been proposed to increase the signal energy concentration and resolution of different TFDs. For example, the adaptive short-time Fourier transform [26,27] uses a variable window length adapted to signal characteristics in order to improve the resolution of spectrogram. Moreover, reassignment techniques [28,29] are alternative approaches for the enhancement of TFDs, especially the synchrosqueezing transform [30], which allows mode retrieval in multicomponent signals.

Beyond the calculation of the aforementioned TFDs for IF estimation, strictly speaking, estimating the IF only makes sense for monocomponent or narrowband signals [18]. For that reason, estimating IF from the peaks of TFDs in multicomponent signals requires an extra step to extract and isolate different components before IF estimation methods can be applied to each component. For this purpose, a conventional approach consists of segmenting the TFD of a multicomponent signal using image processing techniques, including local peak detection and component linking [31] and blind source separation [32], among others.

The Hilbert-Huang transform (the HHT) [33,34] has been proposed as a new adaptive technique for the analysis of nonlinear and non-stationary signals. The technique consists of combining empirical mode decomposition (EMD) and Hilbert spectral analysis to obtain an alternative TFD of a signal, called the Hilbert spectrum (the HS), as a function of its IF and instantaneous amplitude (IA).

The HHT has some advantages over TFD-based IF estimation methods, which is why it was chosen for RS analysis in this paper. Since EMD is an adaptive and direct decomposition technique, it makes it possible to retrieve the modes of a multicomponent signal, called intrinsic mode functions (IMFs), without any a priori knowledge of the signal characteristics. In addition, HHT-based IF estimation is performed by means of differentiation; therefore, the HHT does not suffer from the uncertainty principle and simultaneously provides both high time and high frequency resolutions. Moreover, since IF and IA sequences are separately calculated for each component, we can work independently in either a time-frequency or a time-energy domain, without having to process an entire TFD. Furthermore, although the properties of the HHT have led to its application to a number of biomedical signals [35-38], it has rarely been used for RS analysis, as there are only a few studies, mainly focusing on DAS detection [39-41]. However, we found in our previous studies that the HHT also performed well in CAS detection [42,43], which inspired us to analyze its performance for CAS characterization in depth and explore its advantages over spectrogram, which has traditionally been the most commonly used technique for this purpose.

Another reason for which this study was carried out was that most proposed HHT-based methods for RS analysis [39-42] used the original EMD, which has a mode mixing (MM) effect. The MM effect consists either of an IMF containing components of widely different frequencies or of a signal component appearing in different IMFs [44]. Due to this MM, we found that EMD, when applied to some RS signals, resulted in poor separation of RS signal components [42]. Nevertheless, the original EMD has been used in other RS analysis approaches [45-48]. Among the proposed solutions for MM, the ensemble EMD (EEMD) [44,49] and the noise-assisted multivariate EMD (NA-MEMD) [50] are some of the most well-established and widely used methods, but they have rarely been applied to RS analysis [43,51]. Moreover, the implementation and performance of these methods depend on each application and a detailed analysis of the MM effect and the performance of EEMD and NA-MEMD in RS signals is lacking.

The aim of this study is to provide an in-depth evaluation of the performance of the HHT for RS analysis, which led us to calculate the HS with high resolution as an alternative representation to improve on the performance of spectrogram, especially for CAS characterization. The study is divided into two parts. First, we analyze the MM effect of EMD in recorded RS signals and evaluate the performance of EEMD and NA-MEMD

to reduce this effect using a number of quantitative parameters (section 4). Second, we evaluate the performance of three different IF estimators to obtain a suitable EEMD-Kay-based HS for CAS characterization (section 5), and we propose a new method for the automatic segmentation and characterization of CAS based on the HS processing (section 6). This algorithm was tested using a set of synthetic and recorded CAS signals, which allowed us to compare the performance of the HS and spectrogram.

2. Study dataset

2.1. Recorded RS signals

Recorded RS were obtained from the Pulmonary Function Testing Laboratory of Germans Trias i Pujol University Hospital in Badalona, Spain. All recordings were acquired from 30 patients with asthma. Four piezoelectric contact microphones (TSD108, Biopac Systems, Inc.) were placed on the surface of the patients' backs, on either side of the spinal cord, at the base and near the upper lobe of the right/left lung. All sensors were attached to the skin using adhesive rings. Airflow signals were recorded simultaneously with sound signals using a pneumotachograph (TSD107B, Biopac Systems, Inc.). All signals were sampled at 12,500 samples/s, using a 16-bit analogue-to-digital converter (MP150, Biopac Systems, Inc.). After digitalization, the sound signals were decimated by a factor of 4 to 3,125 samples/s and the respiratory phases were automatically detected using the airflow signal as the reference signal. After cycle segmentation, we selected RS from 636 inspiratory phases, with 353 normal sounds signals and 283 CAS signals, including both monophonic and polyphonic CAS.

2.2. Synthetic signals

In order to test the performance of different IF estimators and the proposed CAS characterization algorithm, we generated several synthetic CAS signals. Monophonic CAS signals with slightly variable IFs were modeled as sinusoid frequency modulated signals:

$$c_{1,k}(t) = sin[2\pi f_c(k)t + 0.6\,sin(2\pi 15t)] \qquad (1)$$

$$f_{1,k}(t) = \frac{1}{2\pi}\frac{\partial \Phi_{1,k}}{\partial t} = \frac{1}{2\pi}\frac{\partial(2\pi f_c(k)t + 0.6\,sin(2\pi 15t))}{\partial t} = f_c(k) + 9\,cos(2\pi 15t)$$
$$(2)$$

$$t = 0..0.3\,s, k = 1..4, f_c = [80,150,250,550]$$

where $f_{1,k}$ are the IF sequences of signals $c_{1,k}(t)$ (Fig. 1-a). Monophonic CAS signals with frequency sweeping were modeled as linear frequency modulated signals:

$$c_{2,k}(t) = sin[2\pi f_c(k)t + 2\pi\mu(k)t^2] \qquad (3)$$

$$f_{2,k}(t) = \frac{1}{2\pi}\frac{\partial \Phi_{2,k}}{\partial t} = \frac{1}{2\pi}\frac{\partial(2\pi f_c(k)t + \pi\mu(k)t^2)}{\partial t} = f_c(k) + 2\mu(k)t$$
$$(4)$$

$$t = 0..0.25\,s, k = 1..4, f_c = [80,150,250,550],$$
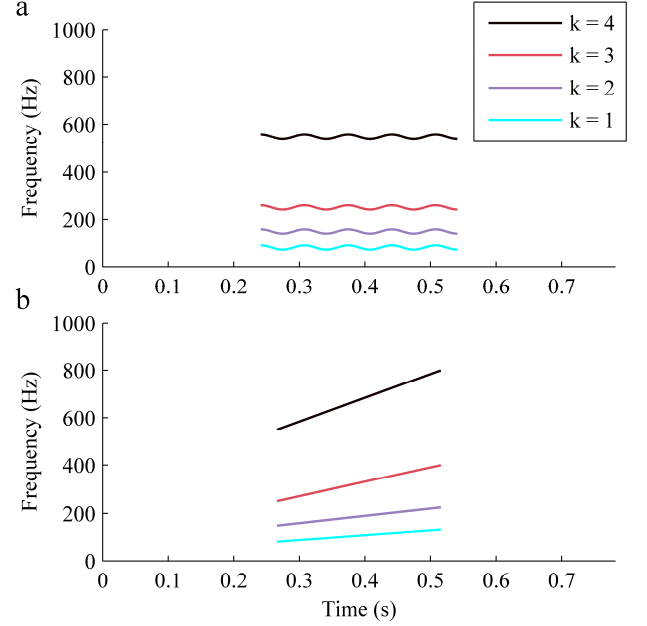$$\mu = [100,150,300,500]$$

Fig. 1. **Theoretical IF of synthetic CAS.** IF laws of monophonic CAS signals with slightly variable IF, $c_{1,k}(t)$ (a), and monophonic CAS signals with frequency sweeping, $c_{2,k}(t)$ (b).

where $f_{2,k}$ are the IF sequences of signals $c_{2,k}(t)$ (Fig. 1-b).

Finally, polyphonic CAS signals were formed by combining $c_{1,k}(t)$ and $c_{2,k}(t)$ as follows:

$$c_{3,k}(t) = \begin{cases} c_{1,k}(t) & t \in [0,0.025] \\ c_{1,k}(t) + c_{2,k+1}(t) & t \in (0.025,0.275] \\ c_{1,k}(t) & t \in (0.275,0.3] \end{cases} \qquad (5)$$
$$t = 0..0.3\,s, k = 1..3$$

We obtained 11 different synthetic CAS signals in total. Each synthetic CAS signal was added to a recorded RS signal containing normal RS at different signal-to-noise ratios (SNRs), thus simulating real CAS that superimposed on normal RS. Since normal RS usually have a sharp energy drop at about 200-250 Hz [1,2], synthetic CAS signals containing components below 200 Hz ($c_{1,k}(t)$, $c_{2,k}(t)$, and $c_{3,k}(t)$ for k=1,2), which overlap with normal RS, were added at SNRs from -4 dB to 12 dB, in increments of 2 dB. However, synthetic CAS signals containing components above 200 Hz ($c_{1,k}(t)$, $c_{2,k}(t)$, and $c_{3,k}(t)$ for k>2) were added at SNRs from -8 dB to 12 dB, in increments of 2 dB. As a result, a total of 109 synthetic CAS signals were obtained, including 80 monophonic and 29 polyphonic CAS signals.

3. Overview of the HHT

The HHT consists of two steps, EMD and the Hilbert transform. The central step of the HHT is EMD, which decomposes a multicomponent signal $s(t)$ into a set of zero mean narrowband components (IMFs), for which meaningful IF and IA can be calculated at any point by

means of the Hilbert transform. The main advantage of EMD is that it is a direct and adaptive decomposition technique, which extracts each IMF directly from the original signal by means of a sifting process [33]. As a result of this process, the signal $s(t)$ can be expressed as a linear combination of its components as follows:

$$s(t) = \sum_{i=1}^{n} IMF_i(t) + r_n(t) \qquad (6)$$

where n is the number of extracted IMFs and $r_n(t)$ is the residue of $s(t)$. Having decomposed $s(t)$ by EMD, IF and IA can be calculated by the phase derivative and envelope of the analytic signal of each IMF and, therefore, $s(t)$ can be expressed as a function of its IF and IA as follows:

$$s(t) = \sum_{i=1}^{n} a_i(t) \cos(\int 2\pi f_i(t)dt) + r_n(t) \qquad (7)$$

where $f_i(t)$ and $a_i(t)$ are the IF and IA of the i-th IMF, respectively. Building on expression (7), we can rearrange IF and IA in a three-dimensional TFD of the amplitude, the HS.

4. Evaluation of the EMD step of the HHT in RS signals

4.1. The MM effect of EMD in RS signals

Ideally, each IMF of a multicomponent signal would contain a few different frequency components of the signal. However, due to the MM effect of EMD, some components may appear within different IMFs, thus leading to some IMFs containing components of widely different frequencies.

Assessing the MM effect in multicomponent random signals, such as RS, is a complex task, since there is no a priori knowledge of the signal component characteristics. Nevertheless, this effect can be clearly observed in CAS signals because they are sinusoidal-like waveforms with well-defined fundamental frequencies. With the aim of illustrating the MM effect in RS signals, we applied EMD to a recorded polyphonic CAS signal with two overlapped components at around 140 Hz and 275 Hz. The resulting IMFs, which were obtained in decreasing order of frequency, are shown in Fig. 2.

The MM effect clearly occurs for both components. The highest frequency component appears in IMFs 1 and 2, whereas the lowest frequency component appears within IMFs 2 and 3.
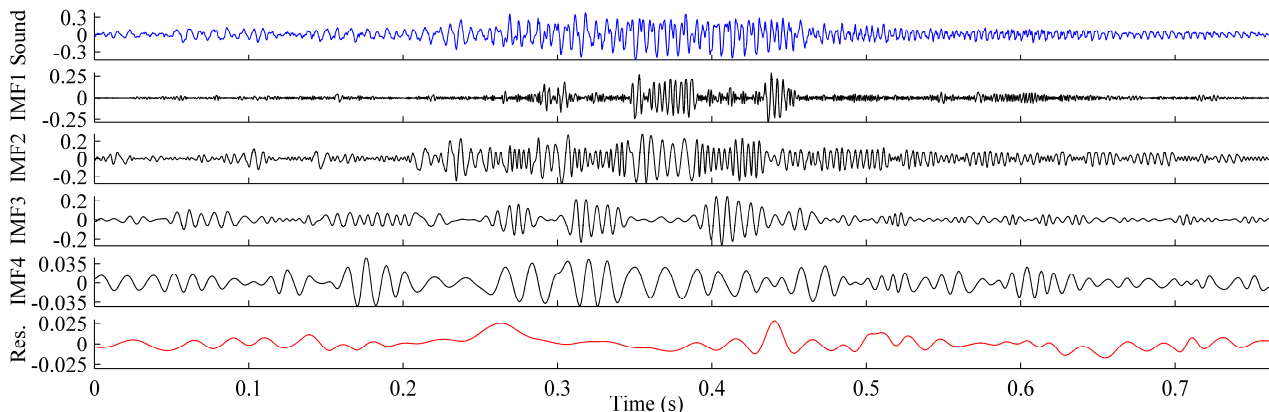
We calculated the power spectral density (PSD) of each IMF using Welch's periodogram with a Hanning window of 80 ms, 40 ms overlap, and 1,024 points for the fast Fourier transform. In order to make different PSDs comparable, we divided them by their respective maximum value (Fig. 3).

As shown in Fig. 3, MM is evident because IMF 2 includes two widely separated frequency components, which correspond to the CAS components. Moreover, the CAS component at around 140 Hz is included within both IMF 2 and IMF 3, whose PSDs overlap to a great extent. Due to this MM effect, the obtained IMFs do not ensure that the application of the Hilbert transform would yield physically meaningful IF estimates.

4.2. EEMD and NA-MEMD as solutions for mode mixing

Over the past few years, many studies have focused on solving the MM effect of EMD. Although several solutions have been proposed, EEMD and NA-MEMD are the most well established and widely used methods. These methods are examples of noise-assisted techniques, which use the benefits of noise in data analysis.

The MM effect occurs when some frequency scales are missing in the original signal. In this case, envelopes calculated during the sifting process are influenced by the extrema of widely different frequency components. However, when applied to white noise, which has scales uniformly distributed across the entire time-frequency plane, EMD acts as an adaptive dyadic filter bank [52,53]. Accordingly, when white noise is added to a multicomponent signal, all signal components with different frequencies are automatically separated by the reference scales set by white noise.
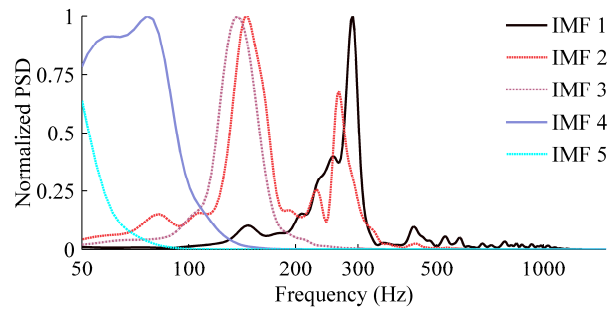
Fig. 3. The MM effect in the frequency domain. PSDs of IMFs shown in Fig. 2.

Fig. 2. The MM effect of EMD. IMFs obtained by means of EMD on an RS signal with polyphonic CAS.

Based on the aforementioned principle, EEMD was first proposed by Wu and Huang in 2009 [44]. The method consists of the iterative application of the original EMD to a signal plus multiple realizations of white noise. The final IMFs are calculated as the mean of those resulting from each iteration. Although the resulting IMFs contain a residual noise level, it can be almost totally cancelled using an appropriate number of iterations.

The MEMD method was later proposed by Rehman and Mandic [54], initially as an extension of EMD for multivariate signals. Like EMD, MEMD has a dyadic filter bank property on white noise [50]. Based on this property, NA-MEMD was proposed to avoid MM in multivariate signals. The idea behind NA-MEMD consists of adding extra channels containing different realizations of white noise to the original signal, and then decomposing the resulting multivariate signal by means of MEMD. This method can also be applied to univariate signals. In this case, only those IMFs obtained for the first channel (original signal) are retrieved.

To provide an example of the performance of noise-assisted techniques in avoiding the MM effect, we have applied the EEMD method to the polyphonic CAS signal shown in the previous section, 4.1, using 100 iterations and noise added at an SNR of 0 dB. The resulting IMFs are shown in Fig. 4.

As shown in Fig. 4, the EEMD method manages to reduce MM, as different frequency components are separated in different IMFs. This separation of frequency components is better observed in the PSDs of the resulting IMFs, shown in Fig. 5.

Comparing these PSDs with those shown in Fig. 3, it is clear that noise has forced the frequency components to be uniformly distributed along the whole frequency range, thus separating widely different frequency components into different IMFs.

Although both EEMD and NA-MEMD manage to reduce MM, there are some differences in the performance of the two methods, which are analyzed in depth in the following section.

4.3. Performance assessment of EMD, EEMD, and NA-MEMD in recorded RS signals

With the aim of comparing the performance of EMD, EEMD, and NA-MEMD in RS signal decomposition, we applied these methods to the 636 RS signals recorded as described in section 2.1, which included normal RS and

CAS.

The choice of the EEMD and NA-MEMD parameters highly depends on the type of signal to be analyzed. Therefore, we followed some basic instructions, as described in [44] and [55], to choose the analysis parameters of each method. Input parameters for EEMD include the number of iterations and the SNR for the added noise. As explained in section 4.2, the residual noise level (n_{res}) of the obtained IMFs can be reduced by increasing the number of iterations. Usually, a few hundred iterations are enough to significantly reduce n_{res}. In fact, n_{res} decreases following the rule $n_{res} = n/\sqrt{N}$, where n is the amplitude of the added noise and N is the number of iterations [44]. Based on this rule, we decided to use square numbers for the number of iterations and a wide range of SNRs.

With regard to NA-MEMD, input parameters include the number of noisy extra channels, the amplitude of the added noise, and the number of directions used in the MEMD process. At least two noisy extra channels should be used and, as a rule of thumb, the minimum number of directions should be twice the number of data channels [55]. Therefore, for EEMD and NA-MEMD, all possible combinations of the following parameters were tested for each RS signal:

- EEMD → number of iterations: 1, 2, 4, 16, 25, 36, 64, 100, 225, and 400; SNRs: -9, -6, -3, 0, 3, 6, 9, 15, and 21.
- NA-MEMD → number of extra channels: 2, 3, and 4; number of directions: 8, 16, 32, and 64; SNRs: -9, -6, -3, 0, 3, 6, 9, 15, and 21.

We programmed the EEMD algorithm using the original EMD algorithm reported by Rilling and Flandrin [56,57].
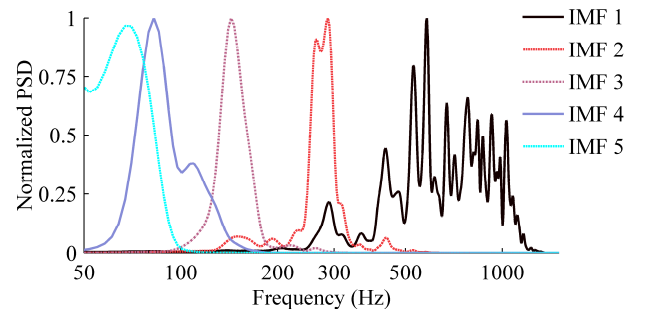

Fig. 5. **Separation of frequency components by EEMD.** PSDs of IMFs shown in Fig. 4.
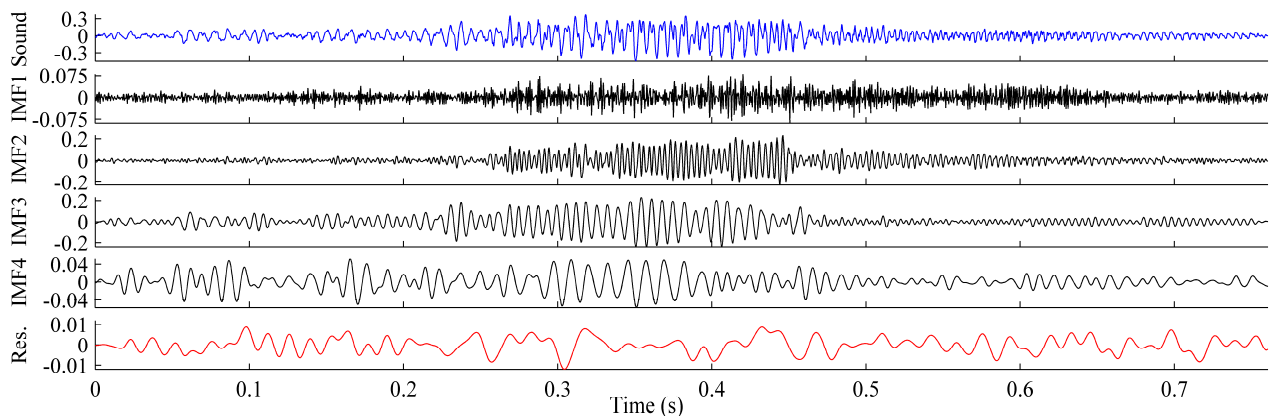

Fig. 4. **Avoiding MM by EEMD.** IMFs obtained by EEMD of the recorded polyphonic CAS signal shown in Fig. 2.

For the NA-MEMD method, we used the Matlab code provided by Mandic [55], which applies the MEMD algorithm reported by Rehman and Mandic [54].

The results were evaluated by means of six quantitative parameters that allowed us to choose the most suitable method for RS signal decomposition. The proposed parameters were divided into three groups depending on the measured feature.

4.3.1. Size and processing time

The first parameter used to compare the performance of EMD, EEMD and NA-MEMD was the total number of IMFs (N_{IMF}) resulting from the decomposition of each RS signal. Moreover, since the frequency range of interest for RS analysis goes from 70 Hz onwards, we also calculated the number of IMFs whose central frequency (f_c), measured from the PSD, was greater than 70 Hz (N_{IMF-FR}).

The results shown in Fig. 6 indicate that the EEMD method (solid lines in Fig. 6-a) provides lower IMFs (about 8-10 IMFs) than the NA-MEMD method, which produces between 14 and 16 IMFs (solid lines in Fig. 6-b). Similar to EEMD, the mean N_{IMF} of the original EMD for all RS signals was 9.1 IMFs.

Figure 6 also shows that the mean N_{IMF-FR} for EEMD was around 5 IMFs at most (dotted lines in Fig. 6-a), which means that EEMD produces between 3 and 5 IMFs (difference between N_{IMF} and N_{IMF-FR}) outside the frequency range of interest for RS analysis. Nevertheless, the mean N_{IMF-FR} for NA-MEMD was around 7 IMFs (dotted lines in Fig. 6-b), which indicates that this method produces more redundant IMFs (around 7-9 IMFs) at low frequencies, which are irrelevant for RS analysis.

As for the high frequencies, although CAS may appear at up to 1,000 Hz, the frequency range of normal RS barely exceeds 250 Hz [2], so having many IMFs covering high frequencies generates redundancy. In this sense, we analyzed f_c of the first four IMFs generated by EEMD and NA-MEMD along all RS signals (Fig. 7).

As shown, NA-MEMD usually produces 4 IMFs above 250 Hz, whereas EEMD only produces 2 IMFs in the same frequency range. Therefore, NA-MEMD also produces more redundant IMFs at high frequencies than EEMD.

In addition to N_{IMF} and N_{IMF-FR}, we also calculated the mean decomposition computing time (D_{CT}) along all RS signals for each method. In other words, we looked at how long it takes for each method to decompose an RS signal into IMFs. All simulations were run in a server with Windows Server 2008 R2 Enterprise installed, an Intel® Xeon® processor E7340 at 2.40 GHz with 4 kernels, and 88 GB of usable RAM. Results shown in Fig. 8 indicate that EEMD is much faster than NA-MEMD.

Despite the fact that D_{CT} increases exponentially with the number of iterations in EEMD, a few seconds is long enough to decompose RS signals using a few hundred iterations, which are sufficient to reduce MM, as explained in the next section, 4.3.2. However, NA-MEMD is a time-consuming method. While the amount of MM decreases with an increase in number of directions (see section 4.3.2), D_{CT} increases in the same way. Therefore, D_{CT} required to substantially reduce the MM effect is too high in comparison with EEMD.

4.3.2. Reduction of MM

In section 4.1 we showed that MM causes frequency overlap between PSDs of different IMFs. Based on this fact, we propose the following parameter to measure the amount of MM, based on frequency overlap (FO) between pairs of IMFs:

$$FO_{i,j}(\%)$$
$$= 100 \frac{max([f_{c2i}, f_{c8i}] \cap [f_{c2j}, f_{c8j}]) - min([f_{c2i}, f_{c8i}] \cap [f_{c2j}, f_{c8j}])}{min\{IDR60_i, IDR60_j\}}$$
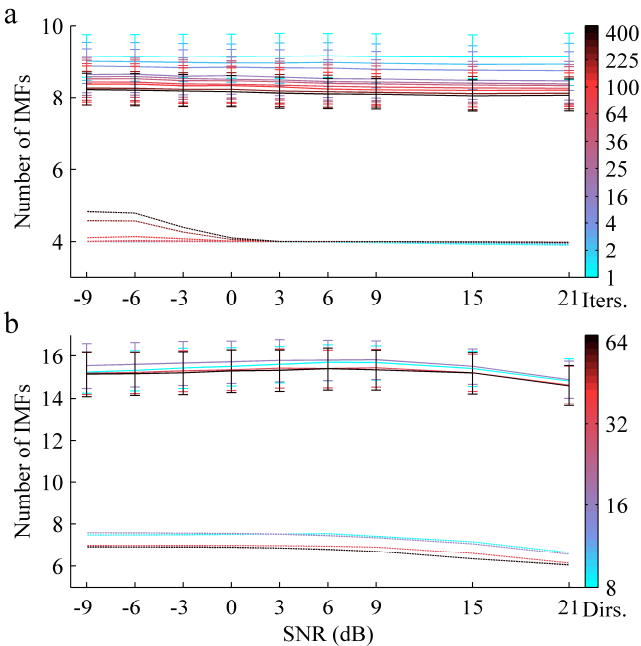
$$(8)$$

Fig. 6. **Size of EEMD and NA-MEMD.** N_{IMF} (solid lines) and N_{IMF-FR} (dotted lines) for EEMD (a) and NA-MEMD (b). All values are the mean and standard deviation along all RS signals. For NA-MEMD (b), all values are also averaged along the number of extra channels.
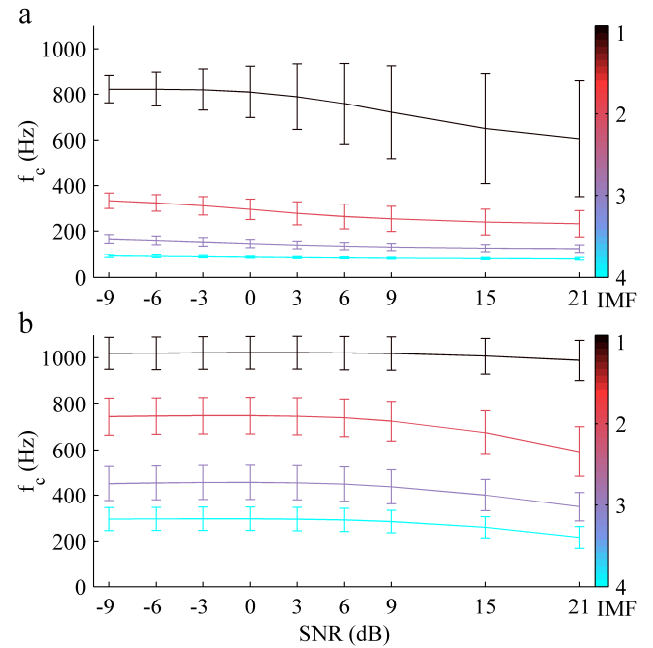

Fig. 7. **Tracking of high frequencies by EEMD and NA-MEMD.** Central frequency (f_c) of IMFs 1-4 for EEMD (a) and NA-MEMD (b). All values are the mean and standard deviation along all RS signals. For EEMD (a), all values are also averaged along the number of iterations. For NA-MEMD (b), all values are also averaged along the number of extra channels and the number of directions.

where f_{c2i}, f_{c8i}, and $IDR60_i$ are frequency parameters measured from the PSD of the *i-th* IMF. Specifically, f_{c2} and f_{c8} are the frequencies at which 20% and 80% of the energy of an IMF are reached, respectively. The *IDR60* parameter is calculated as the difference between f_{c8} and f_{c2}.

We calculated the mean $FO_{i,j}$ along all RS signals, for EMD, EEMD, and NA-MEMD, and for different pairs of IMFs. The mean $FO_{i,j}$ for the original EMD was 22.5% between IMFs 1-2, 22.0% between IMFs 2-3, and 21.3% between IMFs 3-4. The results for EEMD are shown in Fig. 9, which illustrates that $FO_{i,j}$, and hence the amount
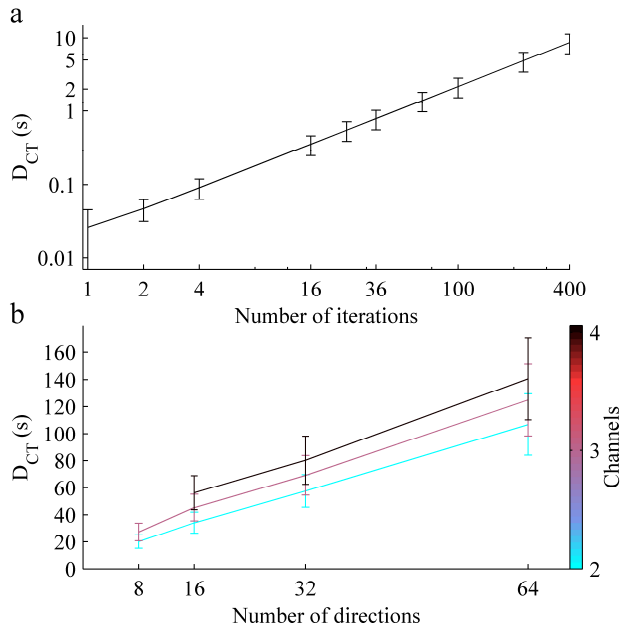
of MM, mainly depends on the SNR when the number of iterations is greater than or equal to 16.

Since $FO_{i,j}$ decreases as the SNR decreases, it would seem that using the lowest SNR is the best solution for reducing MM. However, as explained in the next section, 4.3.3, using very low SNRs increases the residual noise level.

Figure 10 shows the results for NA-MEMD. Since this method produces more IMFs than the EEMD method, we calculated $FO_{i,j}$ between pairs of IMFs from IMF 1 to IMF 5.

As the figure shows, $FO_{i,j}$ depends on both the SNR and the number of directions. In general, the amount of MM decreases with a decrease in SNR and an increase in the number of directions (Figs. 10-b,c,d). However, $FO_{1,2}$ increases with a decrease in SNR, especially for a low number of directions (Fig. 10-a). As explained in section 4.3.1, NA-MEMD tends to generate too many IMFs at high frequencies. If the SNR is too low, noise components cause IMFs to be uniformly distributed along all frequencies, which forces IMFs covering high frequencies to be cramped, and $FO_{1,2}$ increases. On the contrary, if the SNR is high, the effect of noise is negligible, which allows IMFs 1 and 2 to be more widely separated. In any case, NA-MEMD needs a high number of directions to achieve results similar to those of EEMD, and this greatly increases the D_{CT} (see section 4.3.1).

4.3.3. Residual noise level

A major challenge when working with EEMD and NA-MEMD is minimizing the residual noise level in the resulting IMFs. In order to quantify this residual noise level, we propose the following parameters:

Fig. 8. **Processing time, D_{CT}, for EEMD (a) and NA-MEMD (b).** All values are the mean and standard deviation along all RS signals and SNRs.
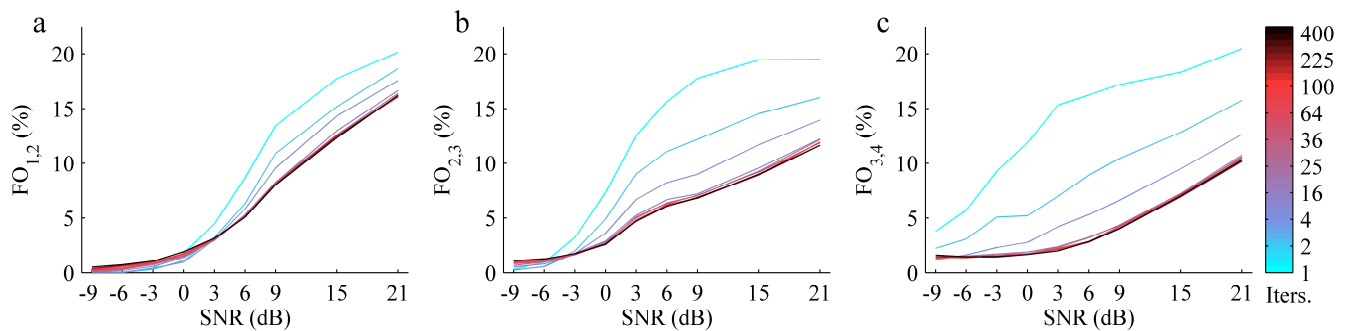

Fig. 9. **MM reduction by EEMD.** Frequency overlap (*FO*) between IMFs 1-2 (a), IMFs 2-3 (b), and IMFs 3-4 (c) for the EEMD method. All values are the mean along all RS signals.
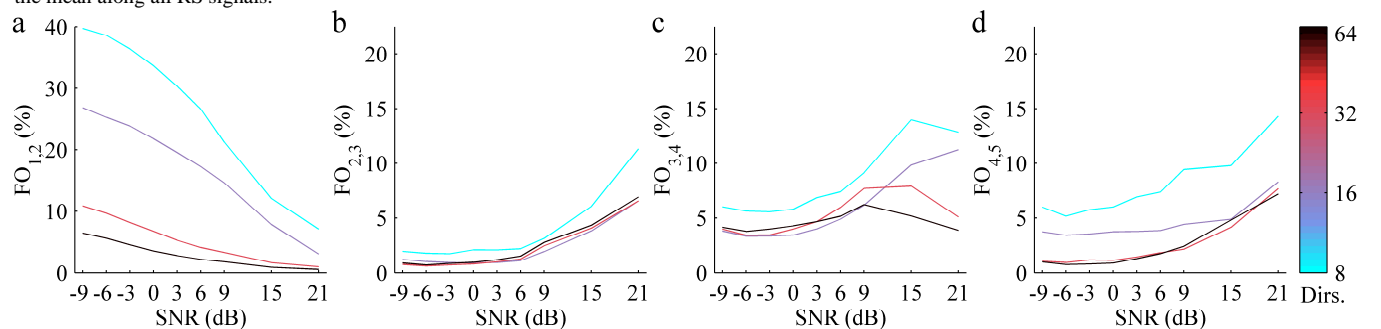

Fig. 10. **MM reduction by NA-MEMD.** Frequency overlap (*FO*) between IMFs 1-2 (a), IMFs 2-3 (b), IMFs 3-4 (c), and IMFs 4-5 (d) for the NA-MEMD method. All values are the mean along all RS signals and the number of extra channels.

- *CC*: cross-correlation at zero lag between the PSD of the original signal (PSD_o) and the PSD of the reconstructed signal (PSD_{rec}).

$$CC = \hat{R}_{PSD_o PSD_{rec}}(0) = \frac{\sum_n PSD_o(n)PSD_{rec}(n)}{\sqrt{\sum_n PSD_o(n)^2 \sum_n PSD_{rec}(n)^2}} \quad (9)$$

- *PSDR*: ratio of the absolute error between PSD_{rec} and PSD_o versus PSD_o.

$$PSDR(\%) = 100 \frac{\sum_n |PSD_{rec}(n) - PSD_o(n)|}{\sum_n PSD_o(n)} \quad (10)$$

where n is the number of points used for the fast Fourier transform. The reconstructed signals were calculated as the direct sum of the corresponding IMFs and residues.

Ideally, *PSDR* and *CC* would be 0 and 1, respectively, if the reconstructed signal were exactly equal to the original signal. This is the case of EMD and NA-MEMD, which provide a perfect reconstruction of the original signal. However, EEMD causes a slight error in the reconstructed signal due to the use of white Gaussian noise in the decomposition process.
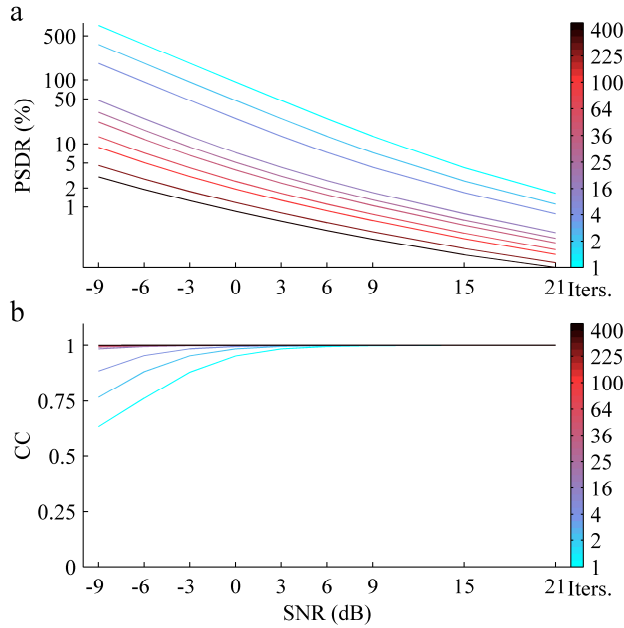

a

b

This error depends on both the amplitude of the white noise and the number of iterations, as shown in Fig. 11.

As shown, *CC* reaches its maximum using 16 iterations or more, independently of the SNR. However, *PSDR* highly depends on the two parameters. Nevertheless, by applying a few hundred iterations, we obtained an acceptable *PSDR* (below 3%) for a wide range of SNRs. Therefore, we can assume that the residual noise level is not a major drawback of EEMD in RS signals provided that the SNR and number of iterations are correctly chosen.

4.4. Selection of parameters in EEMD

Based on the previous results, we decided that EEMD was better than NA-MEMD for RS signal decomposition, since EEMD produced fewer redundant IMFs, managed to reduce the MM effect to a greater extent, and was faster than NA-MEMD. However, in order for EEMD to perform at its fullest potential, we had to fix its parameters (SNR and number of iterations) so that MM was reduced as much as possible and the residual noise level was not significant. To this end, we analyzed the $FO_{i,j}$ and the *PSDR* parameters together for different pairs of IMFs, as shown in Fig. 12.

Values of the intersection points between both parameters for each number of iterations are shown in Table 1.

Assuming 3% as an acceptable upper limit for $FO_{i,j}$ and *PSDR*, an SNR below 1 dB and more than 64 iterations should be used for EEMD. We decided to fix the SNR at 0 dB and use 100 iterations to decompose RS signals by means of EEMD. The $FO_{i,j}$ did not decrease significantly, neither by using more iterations nor by decreasing the SNR. However, increasing the number of iterations or decreasing the SNR greatly increased the D_{CT} and the *PSDR*, respectively.

5. IF-IA estimation and the HS in RS signals

Having decomposed an RS signal into IMFs, the next step in calculating the HS is IF-IA estimation. The use of HHT involves estimating IF as the phase derivative of the analytic signal of each IMF. It is the most intuitive and direct way to define IF of a real signal [17]. For the case of a length-M IMF, the analytic signal, $z_i(n)$, is defined as follows:

Fig. 11. **Measures of residual noise level.** *PSDR* (a) and *CC* (b) for EEMD, as a function of SNR and number of iterations. All values are the mean along all RS signals.
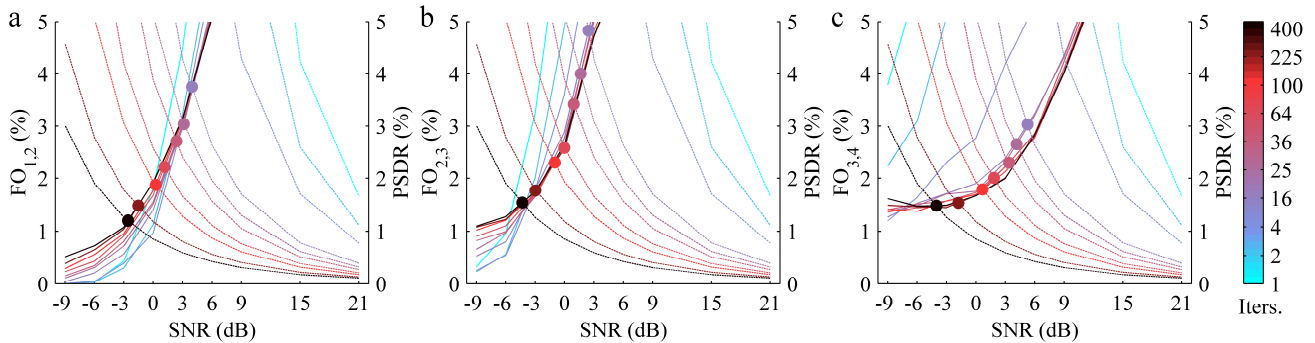

a

b

c

Fig. 12. **Performance parameters of EEMD.** Combination of measures of the amount of MM (*FO*) (solid lines) and residual noise level (*PSDR*) (dotted lines) for IMFs 1-2 (a), 2-3 (b), and 3-4 (c) obtained by EEMD. All values are the mean along all RS signals.

Table 1. Performance of EEMD for RS signal decomposition as a function of the number of iterations and SNR.

	Iterations	1	2	4	16	25	36	64	100	225	400
IMFs 1-2	SNR (dB)	8.7	7.7	6.3	3.6	2.8	2.1	1.0	0.1	-1.7	-2.7
	$FO_{1,2}$/PSDR (%)	14.1	9.8	7.0	3.9	3.2	2.9	2.3	1.9	1.5	1.2
IMFs 2-3	SNR (dB)	8.0	6.7	5.4	2.2	1.3	0.6	-0.5	-1.6	-3.7	-5.2
	$FO_{2,3}$/PSDR (%)	17.1	11.7	8.4	5.1	4.2	3.6	2.8	2.5	2.0	1.7
IMFs 3-4	SNR (dB)	7.9	7.5	6.9	4.7	3.5	2.7	1.1	-0.3	-2.8	-5.0
	$FO_{3,4}$/PSDR (%)	17.4	10.2	6.4	3.3	2.9	2.5	2.2	2.0	1.7	1.7

FO: frequency overlap parameter; *PSDR*: residual noise level parameter.

$$z_i(n) = IMF_i(n) + jH[IMF_i(n)] = a_i(n)exp[j\Phi_i(n)], n = 1..M \tag{11}$$

where *H[ ]* is the Hilbert transform, $a_i(n)$ is the IA, and $\Phi_i(n)$ is the phase of $z_i(n)$. Having calculated $\Phi_i(n)$ for each *i-th* IMF, the next issue is how to address the phase derivative in discrete time. The most common approach is to use finite impulse response derivative filters. This is the case of our first IF estimation method, which is a five-point least squares polynomial derivative (LSPD) approximation [58]:

$$f_i(n) = \frac{f_m}{2\pi}\sum_{k=0}^{4} b_k\Phi_i(n-k), b = [b_0, b_1, b_2, b_3, b_4] = \frac{1}{10}[2,1,0,-1,-2], n = 1..M \tag{12}$$

where f_m is the sample frequency. A major problem of this estimator is that it has very high variance. However, low variance estimators are preferable for calculating the HS, since it ought to be an accurate time-frequency representation in which the signal energy is as concentrated as possible around the IF.

In order to reduce variance, Kay proposed a weighted phase difference estimator [17]. This method consists of calculating the IF estimate by a weighted averaging of a sequence of phase difference measurements, as follows:

$$f_i\left(n + \left\lfloor\frac{N}{2}\right\rfloor\right) = \frac{f_m}{2\pi}\sum_{k=0}^{N-2} w(k)[\Phi_i(n+k+2) - \Phi_i(n+k+1)], n = 1..M - N \tag{13}$$

$$w(k) = \frac{\frac{3}{2}N}{N^2-1}\left(1 - \left(\frac{k-\left(\frac{N}{2}-1\right)}{N/2}\right)^2\right) \tag{14}$$

where *w(k)* is the length-*N* averaging window. The larger the window size, the smaller the variance will be. After testing different window sizes, we propose an averaging window of 32 samples.

Together with the aforementioned methods, in this study we also tested an alternative approach for IF estimation based on the Teager energy operator (TEO) [59]. This method has very low computational complexity and is very straightforward, as IF and IA are directly calculated from the IMF signal as follows:

$$\psi[IMF(n)] = IMF^2(n) - IMF(n-1)IMF(n+1) \tag{15}$$

$$f_i(n) \approx \frac{f_m}{2\pi}\sqrt{\frac{\psi[IMF_i(n+2)-IMF_i(n+1)]}{\psi[IMF_i(n+2)]}}, n = 1..M-3 \tag{16}$$

$$|a_i(n)| \approx \frac{\psi[IMF_i(n+2)]}{\sqrt{\psi[IMF_i(n+2)-IMF_i(n+1)]}}, n = 1..M-3 \tag{17}$$

where $\psi[ ]$ is the TEO.

After calculating IF and IA for each IMF, the HS can be directly obtained by constructing a two-dimensional array with the accumulation of all of the values of the IA sequences at the positions determined by the corresponding IF values and time instants. Since time instants can be determined within the resolution of the sampling period, and IFs can be precise at any number below the Nyquist frequency, the HS can have high time and high frequency resolutions. Both resolutions depend on the bin size selected for each dimension. In this study, we defined time and frequency bin sizes of $1/f_m$ seconds (0.32 ms for $f_m = 3125$ Hz) and 0.5 Hz, respectively. The resulting HS consisted of a matrix of 3125 rows and *M* columns, where *M* was the number of samples of the analyzed RS signal. For display reasons, a smoothing filter is usually applied to the HS. In this study, we used a 20-sample Gaussian filter with a standard deviation of five samples.

Figure 13 shows the spectrogram and three HSs, whose IFs were calculated by means of the aforementioned IF estimators, of a synthetic polyphonic CAS, $c_{3,3}(t)$, added to a normal RS signal at an SNR of 6 dB (section 2.2). The spectrogram was calculated using a 250-sample length Hanning window, with 240 overlapping samples, and 2,048 points for the fast Fourier transform.

It is clearly noticeable that the five-point LSPD approximation (Fig. 13-c) and the TEO-based method (Fig. 13-d) have very high variance. However, Kay's method (Fig. 13-b) greatly reduces variance and provides an accurate HS in which the energy of the signal is highly concentrated around its IF. Furthermore, this EEMD-Kay-based HS has higher energy concentration and resolution than spectrogram. The performance of these two representations for CAS characterization is evaluated in the next section using a larger dataset.
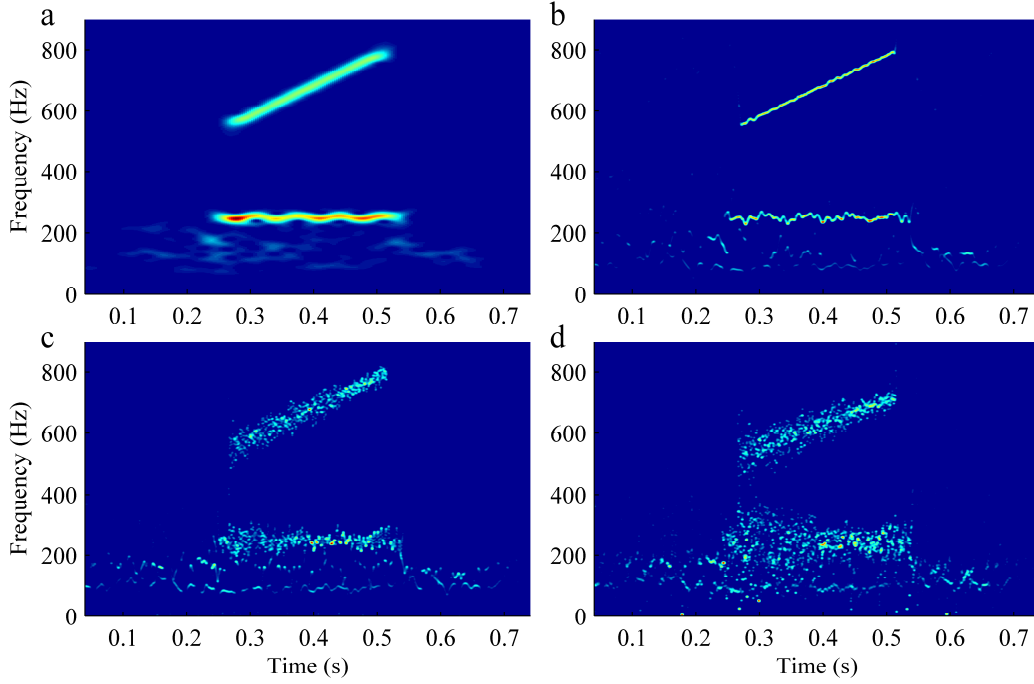
Fig. 13. **Energy concentration of different TFDs.** Spectrogram (a) and the HSs with IFs calculated using Kay's method (b), five-point LSPD approximation (c), and TEO-based method (d) of a polyphonic CAS.

6. Performance assessment of the HS for CAS characterization

6.1. Processing of the HS for CAS segmentation

In this section, we propose a new method for the automatic segmentation and characterization of CAS based on HS processing. The proposed algorithm is based on the region growing methodology and consists of three parts: detection of analysis areas in the HS, selection of seed points for local region growing, and region linking.

In light of the results from the previous sections 4 and 5, we decided to calculate the HS using EEMD with 100 iterations and 0 dB for the SNR, Kay's method with a 32-sample length window for IF estimation, time and frequency bin sizes of 0.32 ms and 0.5 Hz, respectively, and a 20-sample Gaussian smoothing filter with a standard deviation of five samples.

6.1.1. Detection of analysis areas in the HS

The first part of the CAS segmentation algorithm consisted of the CAS detection algorithm proposed in our previous study [43]. That algorithm detected the segments within an RS signal that were more likely to contain CAS based on the hypothesis that the IF dispersion markedly decreases when CAS appear in an RS signal. For that purpose, IF and IA sequences were calculated from the IMFs of an RS signal that had been previously decomposed by EEMD. In that previous study, IF was calculated using the five-point LSPD approximation, which had high variance and emphasized the differences in IF dispersion between those segments of an RS signal containing CAS and those containing normal RS. So, since IF and IA sequences allowed us to work independently in either a time-frequency or a time-energy domain, we were able to use simple dispersion-based

criteria on IF sequences to delimit RS signal segments with a lower IF dispersion. Each delimited RS signal segment was characterized by means of a specific set of features extracted from the IF and IA, including the mean and standard deviation IF, among other features. Those features were used to classify each delimited RS signal segment as containing CAS or normal RS using a support vector machine classifier.

In the present study, we used the outputs from the aforementioned algorithm (classification, mean IF, and standard deviation IF) to mark out an analysis area in the HS for each delimited segment of an RS signal. Each analysis area was centered on the mean IF, had a frequency width of twice the standard deviation IF, and was delimited by the first and the last time instant of the delimited RS signal segment. Figure 14 shows some examples of analysis areas detected from two different synthetic CAS signals. Analysis areas with green edges (CAS areas) correspond to segments classified as containing CAS, whereas analysis areas with red edges (normal areas) correspond to segments classified as containing normal RS.

Although some areas of an RS signal may be misclassified, CAS areas are more likely to contain either an entire or a part of a CAS component than normal areas. Moreover, we considered CAS areas not overlapping with normal RS to be more likely to actually contain CAS components than CAS areas overlapping with normal RS. According to these criteria, three types of analysis areas were defined:

- High-pitched CAS areas: CAS areas whose mean IF was above 250 Hz.
- Low-pitched CAS areas: CAS areas whose mean IF was below 250 Hz.
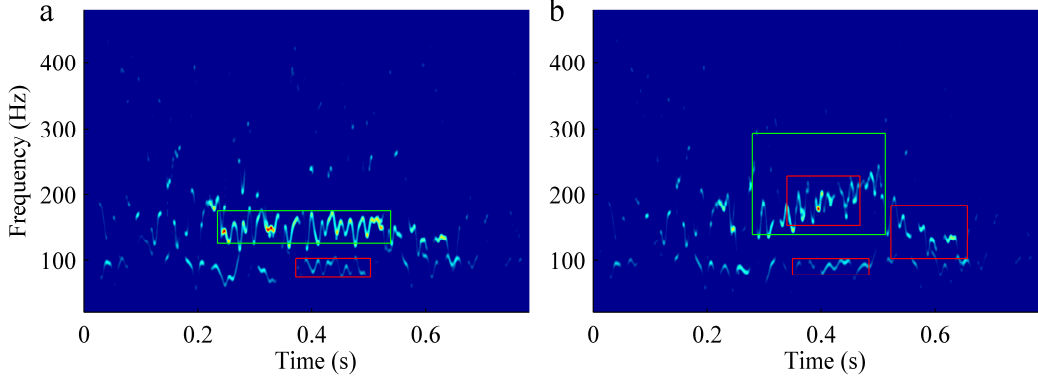- Normal areas

Fig. 14. **Detection of analysis areas in the HS.** Detected analysis areas for synthetic CAS signals $c_{1,2}(t)$ (a) and $c_{2,2}(t)$ (b) added to a normal RS signal at an SNR of -2 dB and -4 dB, respectively, as described in section 2.2.

This classification of analysis areas was used in section 6.1.3 for applying different thresholds when linking regions detected from each type of area. Previously, all areas were considered for seed point searching and local region growing, as explained in the next section.

6.1.2. Seed point search and local region growing

Ideally, CAS components are represented in the HS as ridges describing the IF where signal energy concentrates (see Fig. 13-b and Fig. 14). These ridges are composed of several linked regions, which, in turn, are formed by a set of connected points. Therefore, having detected an analysis area in the HS of an RS signal, the next step was to detect regions with a high concentration of energy around this area.

A determining parameter in this CAS segmentation algorithm was the point amplitude, which was associated with the signal energy at a certain IF. Since different RS signals had different signal energies, we first normalized the HS by dividing it by its maximum. Then, we determined an amplitude threshold to reject those points corresponding to background noise or having very low amplitude. After analyzing histograms from several HSs, we decided to consider only those points with an amplitude exceeding 0.05.

A region growing algorithm was applied to each detected analysis area, in which regions were grown from seed points by adding neighboring points that met a particular inclusion criterion. The first seed point was the point with the highest amplitude inside the analysis area. The first region was then grown by adding neighboring points that met the following criterion:

$$|f_{np} - \bar{f}_{CAS}| \leq \beta_1 \sigma_{f,area}$$

where f_{np} was the frequency of the neighboring point, $\bar{f}_{CAS}$ was the mean frequency along all points already included in the first region, $\sigma_{f,area}$ was the standard deviation frequency along all points of the HS within the analysis area, and β_1 was a scale factor, which was empirically set to 3. When the growth of the first region stopped, another seed point was sought and a new region was grown. Each new seed point was the point with the highest amplitude not yet belonging to any region and sought between the following frequency boundaries:

$$\max\{\bar{f}_{CAS} - \beta_2 \sigma_{f,CAS}, f_{min,area}\} \leq f_{np} \leq \min\{\bar{f}_{CAS} + \beta_2 \sigma_{f,CAS}, f_{max,area}\}$$

where $\sigma_{f,CAS}$ was the standard deviation frequency along all points already included in a region, $f_{min,area}$ and $f_{max,area}$ were the minimum and the maximum frequencies of the analysis area, respectively, and β_2 was a scale factor. This factor was first set to 3 when searching for new seed points between the temporal boundaries of the analysis area. When no new seed points were found within these temporal boundaries, β_2 was set to 2 to search for new seed points outside the temporal boundaries of the analysis area.

Each new region was grown by adding neighboring points that met either the following criterion:

$$|f_{np} - \bar{f}_{CAS}| \leq \beta_1 \sigma_{f,CAS}$$

when the analysis area corresponded to a segment extracted from either IMF 1 or IMF 2, or the following criterion:

$$|f_{np} - \bar{f}_{CAS}| \leq \max\{\beta_1 \sigma_{f,CAS}, \beta_1 \sigma_{f,area}\}$$

when the analysis area corresponded to a segment extracted from either IMF 3 or IMF 4. The region growing process was continued until no new seed points were found.

6.1.3. Region linking

The last step in the CAS segmentation algorithm was to retain only those regions that guaranteed the temporal and frequency continuity of the CAS component. We first rejected any CAS component not containing at least one region longer than 20 ms for high-pitched CAS areas, 50 ms for low-pitched CAS areas, or 80 ms for normal areas. In this way, we prevented the detection of false CAS components that might result from linking many short regions corresponding to background noise. Then, assuming that the longest region inside the analysis area truly belonged to the CAS component, adjacent regions were progressively checked from the nearest to the farthest in both directions along the time axis. Three parameters were calculated for each *i-th* region to measure its proximity to regions already retained as part of the CAS component:

- $t_{i\text{-}cas}$ and $f_{i\text{-}cas}$: temporal and frequency distances between the two nearest points among those of the *i-th* region and those of all regions already retained.
- $\Delta fmean_{i\text{-}cas}$: difference between the mean frequency of the *i-th* region and the mean frequency of the nearest 20-millisecond length segment along the regions already retained.

We considered the nearest regions to be more likely to belong to the CAS component than the farthest regions. Accordingly, we defined three ranges for $t_{i\text{-}cas}$ whose boundaries were determined by thresholds *th1*, *th2*, and *th3* as follows:

$$\text{Range j: } tht_{j-1} < t_{i-cas} \leq tht_j \, , j = 1..3, tht0 = 0$$

Regions belonging to ranges 3 and 1 had the most and the least restrictive conditions, respectively, for retention. The *i-th* region belonging to the *j-th* range was retained if it met one of the following criteria:

$$d_i \geq thdur_j \, \& \, \Delta fmean_{i-cas} \leq thf \, \& \, f_{i-cas} \leq thf \quad (18)$$

$$t_{i-cas} \leq 100 \, ms \, \& \, d_i \geq 125 \, ms \quad\quad\quad (19)$$

where threshold *thdurj* was the required minimum length of the *i-th* region (d_i) belonging to the *j-th* range, and threshold *thf* was the limit for the frequency parameters $f_{i\text{-}cas}$ and $\Delta fmean_{i\text{-}cas}$ and guaranteed the frequency continuity of the CAS component. We made all the thresholds dependent on the type of analysis area (see section 6.1.1) and whether the *i-th* region was inside or outside the analysis area. In this study, thresholds *tht1*, *tht2*, *tht3*, *thdur1*, *thdur2*, *thdur3*, and *thf* were empirically set to the values shown in Table 2. These values were fixed after analyzing many HSs from recorded RS signals.

All regions not belonging to some range and not meeting any of the criteria defined in (18) and (19) were rejected. After checking all of the regions, the regions retained formed the segmented CAS component. Finally, according to the definition of CAS [1,2], we rejected any CAS component shorter than 100 ms.

Each segmented CAS component was characterized by means of the most relevant parameters from a clinical point of view:

- Duration (D): difference between the last and the first point along the time axis.
- Weighted mean frequency (Fmean) of all points, whose amplitudes in the HS were the weights.

6.2. Characterization of simulated CAS signals

In this section, we applied the proposed CAS segmentation algorithm to the HS and spectrogram of the 109 synthetic CAS signals described in section 2.2. Since our algorithm was designed to be applied to the proposed EEMD-Kay-based HS, we had to adjust some parameters for its use with spectrogram, which was calculated using a 250-sample length Hanning window, with 240 overlapping samples, and 2,048 points for the fast Fourier transform. Specifically, we increased the amplitude threshold from 0.05 to 0.1 and β_1 (see section 6.1.2) from 3 to 4.

Together with D and F_{mean}, we calculated the following two parameters for each segmented CAS component as a means of measuring the concentration of both TFDs:

- σ_F: average value of the point by point weighted standard deviation frequency.
- $\sigma_{F\text{-}FI}$: average value of the point by point weighted standard deviation frequency in relation to real IF values, which were defined in (2) and (4) (section 2.2).

These parameters allowed us to compare the performance of the proposed HS and spectrogram for RS analysis, at different SNRs. Figure 15 shows the absolute value of the differences between D (Err_D) and F_{mean} (Err_{Fmean}) calculated using the CAS segmentation algorithm and their real values, and the concentration measures (σ_F and $\sigma_{F\text{-}FI}$) for the HS and spectrogram. Statistical differences between these parameters of the two TFDs were evaluated using a one-sided Wilcoxon signed rank test at the 5% significance level.

As shown in Fig. 15-a, there were two SNR ranges in which Err_D was similar for both TFDs. One of those SNR ranges included synthetic CAS signals added to a normal RS signal at SNRs greater than or equal to 0 dB, which represented situations where, due to their high amplitude, CAS components could be easily detected by the two methods. Similarly, the performance of both TFDs was also similar within the SNR range from -8 dB to -6 dB, which included only synthetic CAS signals containing components above 200 Hz (see section 2.2). Despite the low SNR, those CAS components did not overlap with normal RS and, therefore, could be segmented more easily by both TFDs. The mean Err_D in all of those cases was 18.2 ± 31.6 ms for the HS and 19 ± 21.6 ms for the spectrogram. Compared to the duration of synthetic CAS signals (either 250 ms or 300 ms), both mean Err_D were acceptable, as they represented between 6% and 8% of synthetic CAS duration.

Table 2. Thresholds for region linking

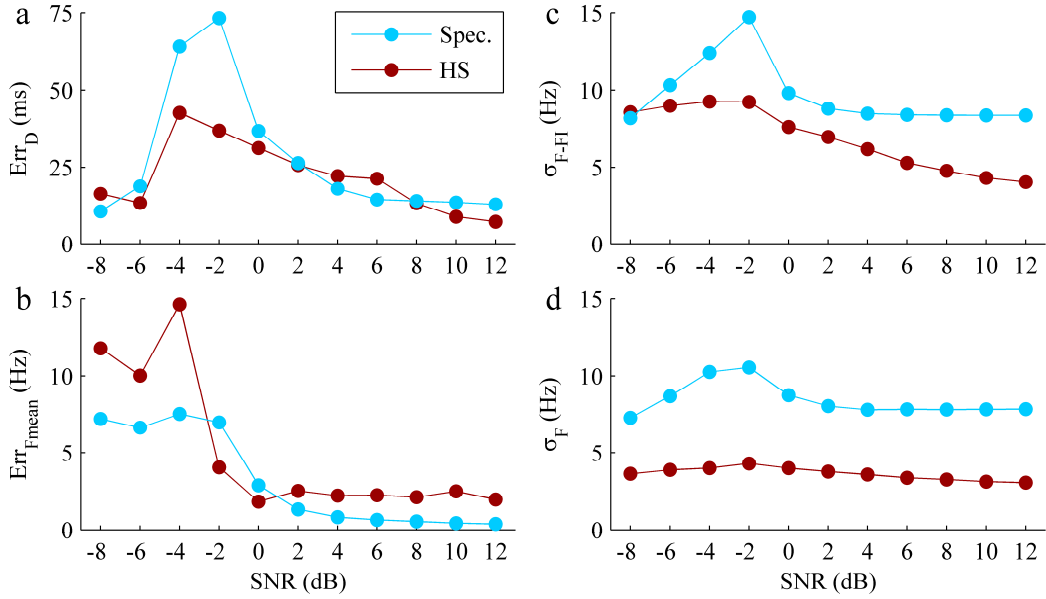| Type of analysis area	tht1 (ms)	thdur1 (ms)	tht2 (ms)	thdur2 (ms)	tht3 (ms)	thdur3 (ms)	thf (Hz)
High-pitched CAS areas, inside	20	none	30	10	100	50	25
High-pitched CAS areas, outside	17.5	5	30	25	100	50	25
Low-pitched CAS areas, inside	17.5	5	30	25	50	100	25
Low-pitched CAS areas, outside	10	15	22.5	35	50	100	25
Normal areas, inside	10	15	22.5	35	22.5	none	20
Normal areas, outside	5	20	10	55	10	none	20

Fig. 15. **Characterization of synthetic CAS signals.** Absolute value of the differences between calculated and real D (Err_D) (a) and F_{mean} (Err_{Fmean}) (b), $\sigma_{F\text{-}FI}$ (c), and σ_F (d), for the HS and spectrogram, as a function of SNR. All values are the mean along RS signals of each SNR value.

However, the advantages of the HS became clear in the SNR range from -4 dB to -2 dB, which represented an unfavorable scenario for synthetic CAS signals containing components below 200 Hz, due to their overlap with normal RS. In those cases, thanks to the high temporal resolution and energy concentration of the HS, it allowed the temporal boundaries of CAS components to be more accurately determined ($Err_D = 39.7 \pm 46.5$ ms). However, the poor resolution and the scattered energy of the spectrogram prevented it from delimiting CAS accurately ($Err_D = 68.9 \pm 84.6$ ms). These differences between Err_D of the two TFDs were statistically significant (p = 0.0018), which indicated that Err_D was higher for the spectrogram than for the HS. Relative to synthetic CAS duration, mean Err_D was between 13% and 16% for the HS and between 23% and 28% for the spectrogram.

As an example, Figs. 16 and 17 show the CAS segmentation of two synthetic monophonic CAS signals added to a normal RS signal at an SNR of -2 dB and -4 dB, respectively, as described in section 2.2.

White rectangles in the figures above (Figs. 16-c,d and 17-c,d) show the boundaries of the theoretical IFs. As shown, it is more difficult to distinguish between CAS components and normal RS in the spectrogram than in the HS, where the boundaries of the CAS component can be detected more accurately. However, some normal RS components are detected as part of the CAS components in the spectrogram. The performance of spectrogram could be improved by increasing the amplitude threshold. However, it would be more difficult to detect weak CAS components, as explained in the next section 6.3.

With regard to Err_{Fmean} (Fig. 15-b), although it was higher in the HS (4.1 ± 9.7 Hz) than in the spectrogram (2.6 ± 6.1 Hz), both Err_{Fmean} were low in comparison with Err_D, which is more critical. However, there were clear and significant differences (p << 0.0001) between the frequency dispersion (σ_F and $\sigma_{F\text{-}FI}$) of both TFDs, as shown in Figs. 15-c and 15-d, which means that the HS

has higher energy concentration than spectrogram.

6.3. Characterization of recorded CAS signals

In this section, we applied the proposed CAS segmentation algorithm to the HS and spectrogram of the 283 CAS signals recorded as described in section 2.1. In this case, as quantitative measurements of Err_D and Err_{Fmean} could not be obtained, we calculated the mean and standard deviation of D, F_{mean}, and σ_F along all segmented CAS components (Table 3).

There were no relevant differences between D and F_{mean} measured from the two TFDs. Even so, these parameters were absolute measures and were not representative of the performance of the two TFDs in recorded CAS segmentation, unlike Err_D and Err_{Fmean} described in the previous section 6.2. Nevertheless, there were significant differences between the σ_F of both TFDs (p << 0.0001), which again demonstrates that the HS has higher energy concentration than spectrogram.

The advantages of the HS over spectrogram in recorded CAS segmentation, especially weak CAS with low energy, are clearly illustrated in Figs. 18, 19, and 20.

The figures above show that the spectrogram failed to entirely extract some CAS components with low amplitude, especially in CAS signals that contained several CAS components with quite different amplitudes. This performance was contrary to that shown in the examples with synthetic CAS (Fig. 16 and Fig. 17), in which the spectrogram detected some normal RS components as CAS components. However, the same algorithm and thresholds were used in both cases. This means that spectrogram is more dependent on the amplitude threshold. However, since the HS achieves more energy concentration along the CAS components, it is less dependent on this parameter of the CAS segmentation algorithm.
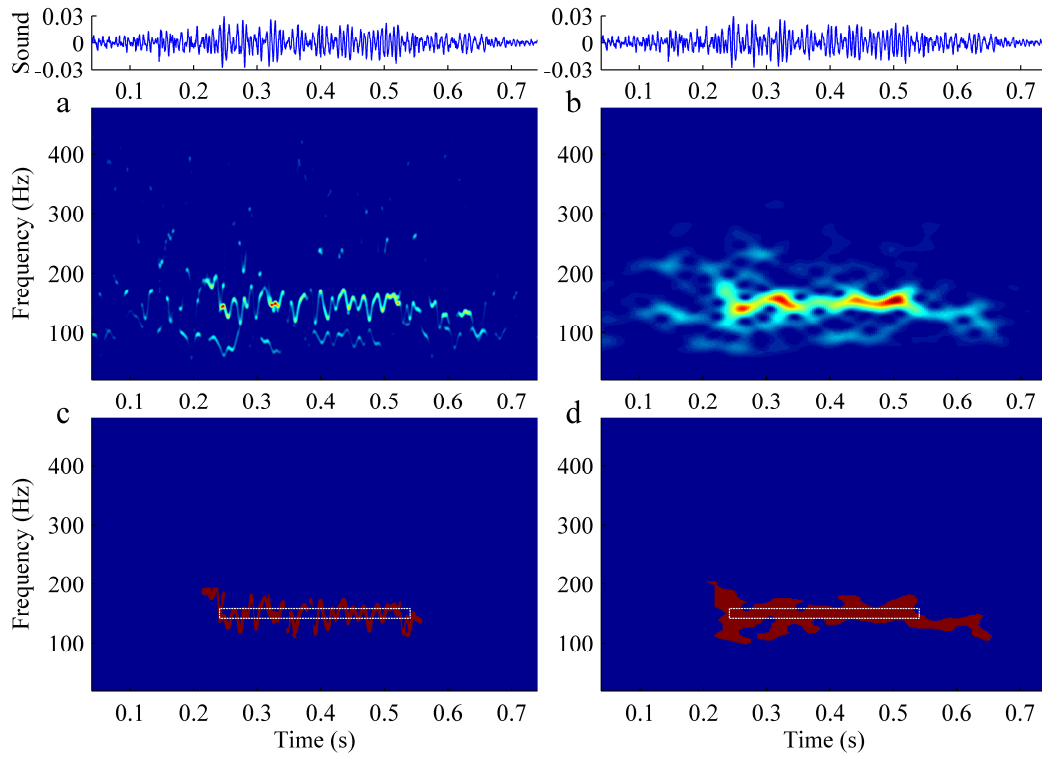
Fig. 16. **Segmentation of synthetic CAS signals (I).** HS (a) and spectrogram (b) of a synthetic monophonic CAS signal. Segmented CAS component using the HS (c) and spectrogram (d).
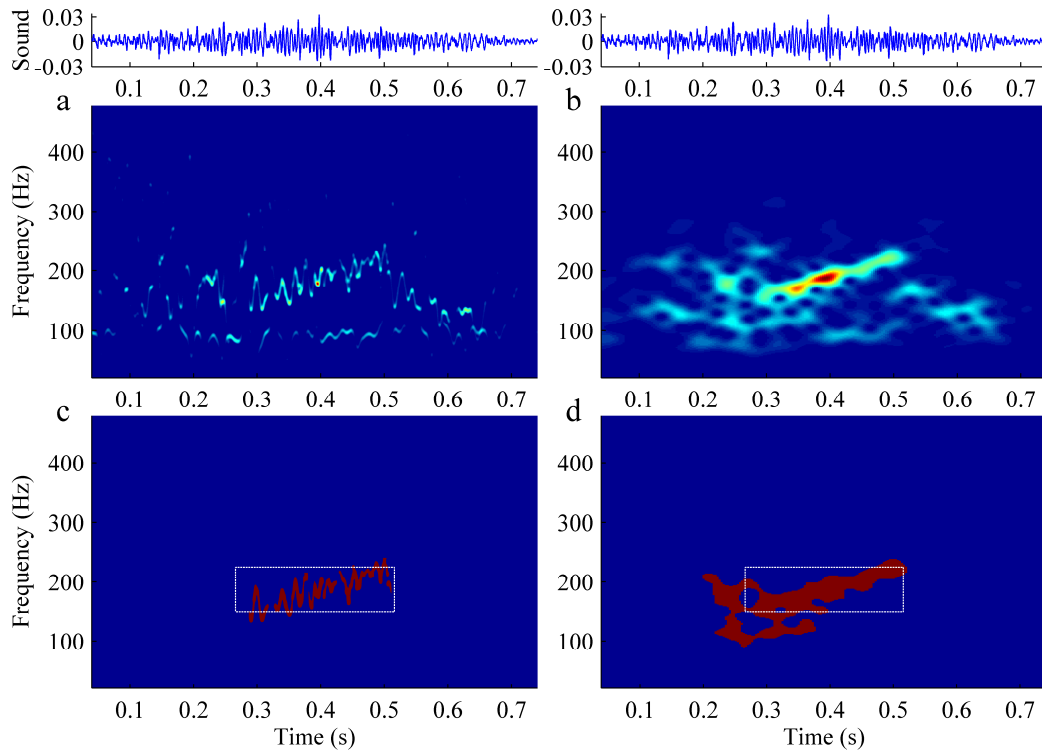

Fig. 17. **Segmentation of synthetic CAS signals (II).** HS (a) and spectrogram (b) of a synthetic monophonic CAS signal with frequency sweeping. Segmented CAS component using the HS (c) and spectrogram (d).

7. Discussion and conclusions

In this study, we conducted a comprehensive evaluation of the performance of the HHT for RS analysis, which led us to propose an EEMD-Kay-based HS that performed very well for CAS characterization. In comparison with spectrogram, which is the most widely used technique for CAS analysis, the HS detected CAS components more precisely, especially those at low SNR that overlap with normal RS.

The most critical stage of the HHT is EMD, due to its MM effect, which causes poor separation of frequency scales. We have gone into detail about the MM effect of EMD in RS signals and the performance of EEMD and NA-MEMD to solve MM. We propose a number of parameters to quantify the size, reduction of MM, and residual noise level of each method. The results after applying EEMD and NA-MEMD to recorded RS signals showed that EEMD is more concise than NA-MEMD, as EEMD produces fewer redundant IMFs and is faster than NA-MEMD. Moreover, EEMD reduces the MM effect more effectively than NA-MEMD.

Table 3. Characteristic parameters of recorded CAS signals

	Spectrogram	HS
D (ms)	299.7 ± 132.6	302.8 ± 122.9
F_{mean} (Hz)	287.0 ± 136.0	285.6 ± 135.0
σ_F (Hz)	7.2 ± 2.6	4.2 ± 1.2

D: duration; F_{mean}: mean frequency; σ_F: frequency dispersion

Although EEMD has already been used in previous studies as part of the HHT [60-62], the procedure for the correct choice of its parameters (amplitude of the added noise and number of iterations) is still unclear, and they must be adjusted to the characteristics of different signals. In this regard, our proposed parameters could be used to assess MM and select EEMD parameters in other applications.

The IF is estimated in the HHT by means of the phase derivative of the analytic signal of IMFs. However, a major drawback of this IF estimation method is the high variance of the IF estimates. Kay's IF estimator proved to be a direct and straightforward method that significantly reduced that variance, which is a desirable property for the purpose of obtaining an accurate HS with high concentrations of signal energy. In fact, we propose the use of an EEMD-Kay-based HS as an alternative and precise time-frequency representation of RS signals.

The main advantage of the proposed HHT-based approach over other TFD-based approaches for RS analysis is the high temporal and high frequency resolution of the HS. Since IF is calculated by differentiation, time resolution can be as high as that determined by the sampling rate. Moreover, the frequency resolution in the HS does not depend on the data length as Fourier-based or wavelet-based techniques, but rather it is determined by the bin size selected. Furthermore, the EEMD allows us to separate different signal components prior to IF and IA calculation without having to process an entire TFD.
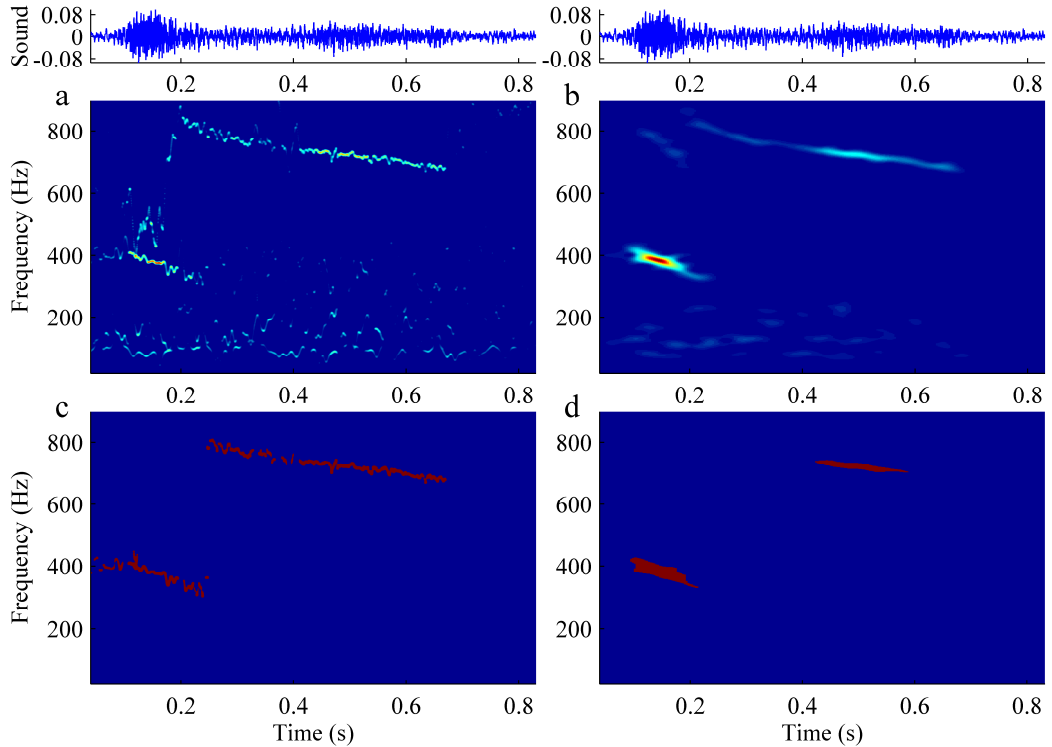

Fig. 18. **Segmentation of recorded CAS signals (I).** HS (a) and spectrogram (b) of a recorded CAS signal that contains two CAS components. Segmented CAS components using the HS (c) and spectrogram (d).
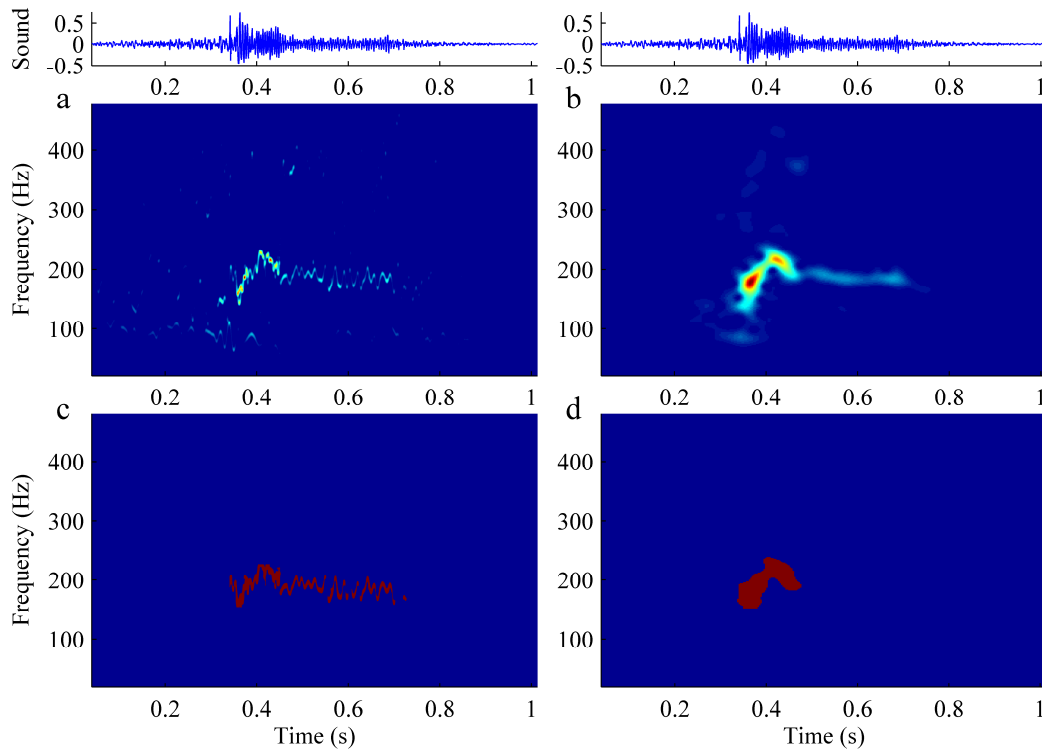
Fig. 19. **Segmentation of recorded CAS signals (II).** HS (a) and spectrogram (b) of a recorded monophonic CAS signal. Segmented CAS component using the HS (c) and spectrogram (d).
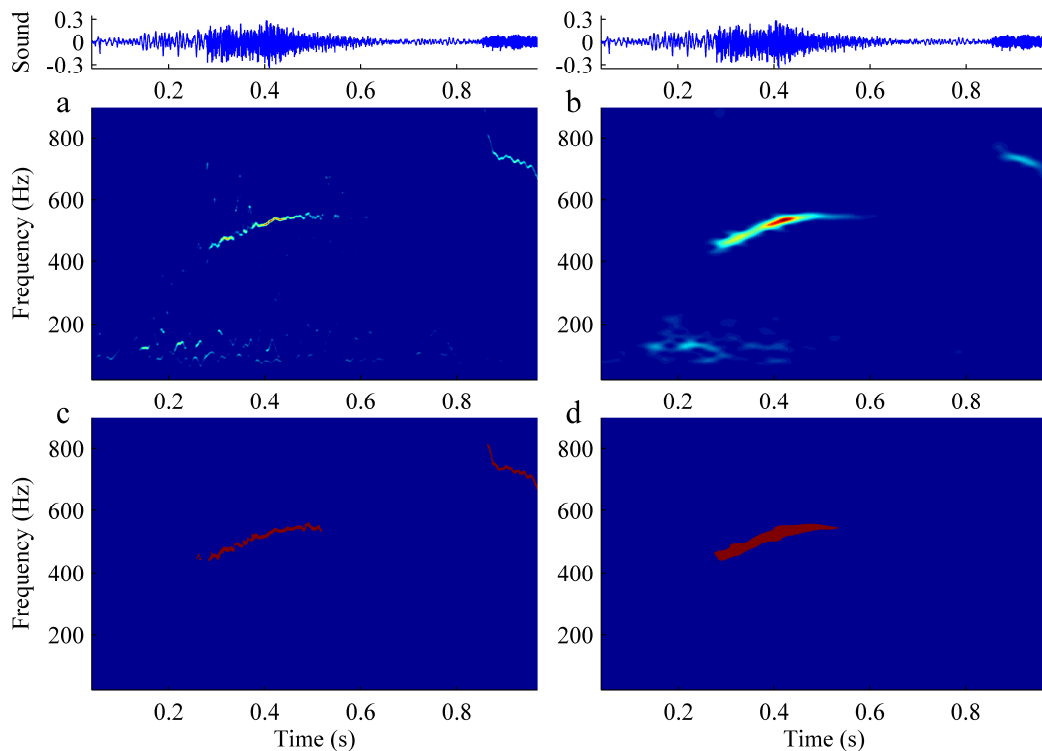

Fig. 20. **Segmentation of recorded CAS signals (III).** HS (a) and spectrogram (b) of a recorded CAS signal that contains two CAS components. Segmented CAS components using the HS (c) and spectrogram (d).

Accordingly, the HHT does not only allow working in a time-frequency plane, like other TFDs, but also analyzing IF and IA sequences separately.

Taking advantage of properties of the HS, we propose a new method for the automatic segmentation and characterization of CAS. A key point of our method is the

CAS detector (section 6.1.1) that we previously proposed in [43]. Using this CAS detector as the first step of our CAS segmentation algorithm, we can locate CAS components in the time-frequency plane, which facilitates their subsequent segmentation using a region growing methodology together with a set of region linking criteria. The main advantage of the CAS detector is that it is based solely on a number of IF criteria, which makes our CAS segmentation algorithm less dependent on amplitude criteria.

We applied the proposed CAS segmentation algorithm to the HS and spectrogram of two sets of synthetic and recorded CAS signals to compare the performance of the two TFDs. The results showed that the HS has some clear advantages over spectrogram. The resolution of spectrogram is limited due to the use of a finite length analysis window. Furthermore, the signal energy in spectrogram is more scattered across the entire time-frequency plane, which blurs the boundaries of some CAS components, especially in CAS components that overlap with normal RS at low SNRs. Therefore, the performance of spectrogram is more dependent on the amplitude thresholds. However, the HS achieves higher energy concentrations around the ridges described by CAS components, which makes the HS less dependent on amplitude criteria. This property, together with the high resolution in time and frequency domains, allows the duration (D) and mean frequency (F_{mean}) of CAS components to be more accurately determined using the HS than spectrogram.

If we consider that the mean D and mean F_{mean} obtained from recorded CAS signals (Table 3 in section 6.3) are representative of real CAS in asthmatic patients, the mean Err_D and mean Err_{Fmean} obtained from synthetic CAS signals (Fig. 15 in section 6.2) were quite acceptable. Both the HS and spectrogram had high performance in measuring F_{mean}, since their mean Err_{Fmean} (below 15 Hz) represented less than 5% of the mean F_{mean} of real CAS (around 290 Hz). With respect to D, the HS performed significantly better than spectrogram, especially for low SNRs, where their mean Err_D reached around 13% (39.7 ms) and 23% (68.9 ms), respectively, of the mean D of real CAS (around 300 ms). In addition to this difference in Err_D at low SNR, we showed some examples of weak CAS that were properly segmented by the HS but were either wrongly segmented or not detected by spectrogram. From a clinical point of view, spectrogram could lead to an underestimation of the presence of CAS and their features, which are directly related to the severity of airflow obstruction [1,2] in patients with obstructive pulmonary diseases.

This study implies a step forward in the analysis of CAS, as most previous approaches for CAS analysis [3-7,9] mainly focused on differentiating CAS from normal RS, but not on analyzing CAS features, such as duration and mean frequency, which are the most relevant clinical parameters. Moreover, some of those studies used spectrogram for CAS detection and extraction. However, our algorithm characterizes CAS more accurately in both time and frequency domains thanks to the properties of the HS, whose performance has been thoroughly tested on both synthetic and real CAS signals.

Our proposed version of the HHT based on EEMD and Kay's IF estimator is a promising tool for the analysis of RS signals. Due to its high resolution, the proposed HS is a suitable TFD to analyze not only CAS signals, but also shorter RS signal components, such as DAS. This methodology, including the CAS characterization algorithm could be included within a more complex RS analysis system that facilitates long-term monitoring and improves reliability in the diagnosis of obstructive pulmonary diseases.

Acknowledgements

References

[1] A. Bohadana, G. Izbicki, and S.S. Kraman, "Fundamentals of lung auscultation," *N. Engl. J. Med.*, vol. 370, pp. 744–751, Feb. 2014.

[2] A.R.A. Sovijärvi, L.P. Malmberg, G. Charbonneau, J. Vanderschoot, F. Dalmasso, C. Sacco, M. Rossi, and J.E. Earis, "Characteristics of breath sounds and adventitious respiratory sounds," *Eur. Respir. Rev.*, vol. 10, pp. 591–596, 2000.

[3] A. Homs-Corbera, J.A. Fiz, J. Morera, and R. Jané, "Time-frequency detection and analysis of wheezes during forced exhalation," *IEEE Trans. Biomed. Eng.*, vol. 51, pp. 182–186, Jan. 2004.

[4] B.S. Lin, B.S. Lin, H.D. Wu, F.C. Chong, and S.J. Chen, "Wheeze recognition based on 2D bilateral filtering of spectrogram," *Biomed. Eng. Appl. Basis Commun.*, vol. 18, pp. 128–137, Jun. 2006.

[5] S.A. Taplidou and L.J. Hadjileontiadis, "Wheeze detection based on time-frequency analysis of breath sounds," *Comput. Biol. Med.*, vol. 37, pp. 1073–1083, Aug. 2007.

[6] R.J. Riella, P. Nohama, and J.M. Maia, "Method for automatic detection of wheezing in lung sounds," *Braz. J. Med. Biol. Res.*, vol. 42, pp. 674–684, Jul. 2009.

[7] A. Oliveira, C. Pinho, J. Dinis, D. Oliveira, and A. Marques, "Automatic wheeze detection and lung function evaluation: a preliminary study," in *Proc. Int. Conf. Health Inform.*, Barcelona, 2013, pp. 323–326.

[8] S.A. Taplidou and L.J. Hadjileontiadis, "Analysis of wheezes using wavelet higher order spectral features," *IEEE Trans. Biomed. Eng.*, vol. 57, pp. 1596–1610, Jul. 2010.

[9] F. Jin, S.S. Krishnan, and F. Sattar, "Adventitious sounds identification and extraction using temporal-spectral dominance-based features," *IEEE Trans. Biomed. Eng.*, vol. 58, pp. 3078–3087, Nov. 2011.

[10] L.J. Hadjileontiadis and S.M. Panas, "Separation of discontinuous adventitious sounds from vesicular sounds using a wavelet-based filter," *IEEE Trans. Biomed. Eng.*, vol. 44, pp. 1269–1281, 1997.

[11] F. Ayari, A.T. Alouani, and M. Ksouri, "Wavelets: an efficient tool for lung sounds analysis," in *IEEE Int. Conf. Comput. Syst. Appl.*, Doha, 2008, pp. 875–878.

[12] G. Serbes, C.O. Sakar, Y.P. Kahya, and N. Aydin, "Feature extraction using time-frequency/scale analysis and ensemble of feature sets for crackle detection," in *Conf. Proc. IEEE Eng. Med. Biol. Soc.*, Boston, 2011, pp. 3314–3317.

[13] I.T. Rekanos and L.J. Hadjileontiadis, "An iterative kurtosis-based technique for the detection of nonstationary bioacoustic signals," *Signal Process.*, vol. 86, pp. 3787–3795, 2006.

[14] L.J. Hadjileontiadis and I.T. Rekanos, "Detection of explosive lung and bowel sounds by means of fractal dimension," *IEEE Signal Process. Lett.*, vol. 10, pp. 311–314, 2003.

[15] S. Charleston-Villalobos, G. Dorantes-Méndez, R. González-Camarena, G. Chi-Lem, J.G. Carrillo, and T. Aljama-Corrales, "Acoustic thoracic image of crackle sounds using linear and nonlinear processing techniques," *Med. Biol. Eng. Comput.*, vol. 49, pp. 15–24, 2011.

[16] B. Boashash, "Time frequency signal analysis and processing. A comprehensive reference," Elsevier, First Edition, 2003, chapter 7, 770 pages. (9780080443355, 9780080543055).

[17] B. Boashash, "Estimating and interpreting the instantaneous frequency of a signal-part 2: algorithms and applications," *Proc. IEEE*, vol. 80, pp. 540–568, 1992.

[18] N.E. Huang, Z. Wu, S.R. Long, K.C. Arnold, X. Chen, and K. Blank, "On instantaneous frequency," *Adv. Adapt. Data Anal.*, vol. 1, pp. 177–229, Apr. 2009.

[19] P. O'Shea and B. Boashash, "Some robust instantaneous frequency estimation techniques with application to non-stationary transient detection," in *Proc. Eur. Conf. Signal Process.*, Barcelona, 1990, pp. 165–168.

[20] L.J. Griffiths, "Rapid measurement of digital instantaneous frequency," *IEEE Trans. Acoust.*, vol. 23, pp. 207–222, 1975.

[21] J.M. Cioffi and T. Kailath, "Fast, recursive-least-squares transversal filters for adaptive filtering," *IEEE Trans. Acoust.*, vol. 32, pp. 304–337, 1984.

[22] K.C. Sharman and B. Friedlander, "Time-varying autoregressive modeling of a class of nonstationary signals," in *IEEE Int. Conf. Acoust. Speech Signal Process.*, San Diego, 1984, pp. 227–230.

[23] L. Stankovic, I. Djurovic, S. Stankovic, M. Simeunovic, S. Djukanovic, and M. Dakovic, "Instantaneous frequency in time-frequency analysis: enhanced concepts and performance of estimation algorithms," *Dig. Signal Process.*, vol. 35, pp. 1–13, 2014.

[24] P. Flandrin and W. Martin, "A general class of estimators for the Wigner-Ville spectrum of non-stationary processes," in *Systems Analysis and Optimization of Systems, Lecture Notes in Control and Information Sciences*. Berlin, Vienna, New York Springer-Verlag, 1984, pp. 15–23.

[25] J. Jeong and W.J. Williams, "Kernel design for reduced interference distributions," *IEEE Trans. Signal Process.*, vol. 40, pp. 402–412, 1992.

[26] S.C. Sekhar and T.V. Sreenivas, "Adaptive spectrogram vs. adaptive pseudo-Wigner-Ville distribution for instantaneous frequency estimation," *Signal Process.*, vol. 83, pp. 1529–1543, 2003.

[27] J. Zhong and Y. Huang, "Time-frequency representation based on an adaptive short-time Fourier transform," *IEEE Trans. Signal Process.*, vol. 58, pp. 5118–5128, 2010.

[28] F. Auger and P. Flandrin, "Improving the readability of time-frequency and time-scale representations by the reassignment method," vol. 43, pp. 1068–1089, 1995.

[29] G.K. Nilsen, "Recursive time-frequency reassignment," *IEEE Trans. Signal Process.*, vol. 57, pp. 3283–3287, 2009.

[30] F. Auger, P. Flandrin, Y.-T. Lin, S. McLaughlin, S. Meignen, T. Oberlin, and H.-T. Wu, "Time-frequency reassignment and synchrosqueezing: an overview," *IEEE Signal Process. Mag.*, vol. 30, pp. 32–41, 2013.

[31] L. Rankine, M. Mesbah, and B. Boashash, "IF estimation for multicomponent signals using image processing techniques in the time-frequency domain," *Signal Process.*, vol. 87, pp. 1234–1250, 2007.

[32] P. Comon and C. Jutten, "Handbook of blind source separation. Independent component analysis and applications," Academic Press, 2010.

[33] N.E. Huang, Z. Shen, S.R. Long, M.C. Wu, H.H. Shih, Q. Zheng, N.C. Yen, C.C. Tung, and H.H. Liu, "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis," *Proc. R. Soc. A*, vol. 454, pp. 903–995, 1998.

[34] N.E. Huang, X. Chen, M.T. Lo, and Z. Wu, "On Hilbert spectral representation: a true time-frequency representation for nonlinear and nonstationary data," *Adv. Adapt. Data Anal.*, vol. 3, pp. 63–93, 2011.

[35] L. Mayeta, A. López, and A.F. Ruiz, "Evaluation of the Hilbert-Huang transform for myoelectric pattern classification: towards a method to detect movement intention," in *Panam. Health Care Exch.*, Medellin, 2013, pp. 1–6.

[36] A. Mert and A. Akan, "Hilbert-Huang transform based hierarchical clustering for EEG denoising," in *Proc. Eur. Conf. Signal Process.*, Marrakech, 2013, pp. 1–5.

[37] J. Yan and L. Lu, "Improved Hilbert-huang transform based weak signal detection methodology and its application on incipient fault diagnosis and ECG signal analysis," *Signal Process.*, vol. 98, pp. 74–87, 2014.

[38] Z. Zhidong, Z. Zhijin, and C. Yuquan, "Time-frequency analysis of heart sound based on HHT," in *Proc. IEEE Int. Conf. Commun. Circuits Syst.*, Hong Kong, 2005, pp. 926–929.

[39] L. Zhenzhen and D. Minghui, "HHT based lung sound crackle detection and classification," in *Proc. Int. Symp. Intell. Signal Process. Commun. Syst.*, 2005, pp. 385–388.

[40] B.A. Reyes, S. Charleston-Villalobos, R. Gonzalez-Camarena, and T. Aljama-Corrales, "Analysis of discontinuous adventitious lung sounds by Hilbert-Huang spectrum," in *Conf. Proc. IEEE Eng. Med. Biol. Soc.*, Vancouver, 2008, pp. 3620–3623.

[41] B.A. Reyes, S. Charleston-Villalobos, R. González-Camarena, and T. Aljama-Corrales, "Assessment of time-frequency representation techniques for thoracic sounds analysis," *Comput. Methods Programs Biomed.*, vol. 114, pp. 276–290, 2014.

[42] M. Lozano, J.A. Fiz, and R. Jané, "Estimation of instantaneous frequency from empirical mode decomposition on respiratory sounds analysis," in *Conf. Proc. IEEE Eng. Med. Biol. Soc.*, Osaka, 2013, pp. 981–984.

[43] M. Lozano, J.A. Fiz, and R. Jané, "Automatic differentiation of normal and continuous adventitious respiratory sounds using ensemble empirical mode decomposition and instantaneous frequency," *IEEE J. Biomed. Health Inform.*, 2015 [Epub ahead of print].

[44] Z. Wu and N.E. Huang, "Ensemble empirical mode decomposition: a noise-assisted data analysis method," *Adv. Adapt. Data Anal.*, vol. 1, pp. 1–41, Jan. 2009.

[45] L.J. Hadjileontiadis, "Empirical mode decomposition and fractal dimension filter. A novel technique for denoising explosive lung sounds," *IEEE Eng. Med. Biol. Mag.*, vol. 26, no. 1, pp. 30–39, 2007.

[46] S. Liu, R.X. Gao, D. John, J. Staudenmayer, and P. Freedson, "Tissue artifact removal from respiratory signals based on empirical mode decomposition," *Ann. Biomed. Eng.*, vol. 41, pp. 1003–1015, 2013.

[47] S. Charleston-Villalobos, R. Gonzalez-Camarena, G. Chi-Lem, and T. Aljama-Corrales, "Crackle sounds analysis by empirical mode decomposition," *IEEE Eng. Med. Biol. Mag.*, vol. 26, no. 1, pp. 40–47, 2007.

[48] S. Charleston-Villalobos, L.F. Dominguez-Robert, R. Gonzalez-Camarena, and A.T. Aljama-Corrales, "Heart sounds interference cancellation in lung sounds," in *Conf. Proc. IEEE Eng. Med. Biol. Soc.*, New York, 2006, pp. 1694–1697.

[49] J. Zhang, R. Yan, R.X. Gao, and Z. Feng, "Performance enhancement of ensemble empirical mode decomposition," *Mech. Syst. Signal Process.*, vol. 24, pp. 2104–2123, Oct. 2010.

[50] N. Rehman and D.P. Mandic, "Filter bank property of multivariate empirical mode decomposition," *IEEE Trans. Signal Process.*, vol. 59, no. 5, pp. 2421–2426, May 2011.

[51] L. ChingShun, W.A. Tanumihardja, and S. HongHui, "Lung-heart sound separation using noise assisted multivariate empirical mode decomposition," in *Int. Symp. Intell. Signal Process. Commun. Syst.*, Naha, 2013, pp. 726–730.

[52] Z. Wu and N.E. Huang, "A study of the characteristics of white noise using the empirical mode decomposition method," *Proc. R. Soc. Lond. A*, vol. 460, pp. 1597–1611, 2004.

[53] P. Flandrin, G. Rilling, and P. Gonçalvés, "Empirical mode decomposition as a filter bank," *IEEE Signal Process. Lett.*, vol. 11, pp. 112–114, 2004.

[54] N. Rehman and D.P. Mandic, "Multivariate empirical mode decomposition," *Proc. R. Soc. Lond. A*, vol. 466, pp. 1291–1302, 2010.

[55] [Online]. 2014 Available: http://www.commsp.ee.ic.ac.uk/~mandic/research/emd.htm.

[56] G. Rilling, P. Flandrin, and P. Gonçalvès, "On empirical mode decomposition and its algorithms," in *Proc. IEEE/EURASIP Workshop Nonlinear Signal Image Process.*, Grado, 2003.

[57] [Online]. 2012 Available: http://perso.ens-lyon.fr/patrick.flandrin/emd.html.

[58] W.J. Tompkins, "Biomedical Digital Signal Processing," Prentice-Hall, 1993, chapter 5, pp. 111–116.

[59] P. Maragos, J.F. Kaiser, and T.F. Quatieri, "Energy separation in signal modulations with application to speech analysis," *IEEE Trans. Signal Process.*, vol. 41, pp. 3024–3051, 1993.

[60] H. Li, Y. Wang, and Y. Ma, "Ensemble empirical mode decomposition and Hilbert-Huang transform applied to bearing fault diagnosis," in *IEEE Int. Congr. Image Signal Process.*, Yantai, 2010, pp. 3413–3417.

[61] W. Zhang, X. Zhang, and Y. Sun, "Based on EEMD-HHT marginal spectrum of speech emotion recognition," in *IEEE Int. Conf. Comput. Meas. Control Sens. Netw.*, Taiyuan, 2012, pp. 91–94.

[62] Y. Amirat, M. Benbouzid, T. Wang, and S. Turri, "Performance analysis of an EEMD-based Hilbert Huang transform as a bearing failure detector in wind turbines," in *Int. Conf. Green Energy*, Sfax, 2014, pp. 193–198.