Time-frequency and advanced frequency estimation techniques for the investigation of bat echolocation calls

Yannis Kopsinis

Institute for Digital Communications,

School of Engineering and Electronics,

the University of Edinburgh Alexander Graham Bell Bldg,

King's Buildings,

EH9 3JL,

Edinburgh,

UK,

email: kopsinis@ieee.org

Elias Aboutanios

School of Electrical Engineering and Telecommunications,

The University of New South Wales Sydney,

NSW 2052,

Australia,

email:elias@ieee.org

Dean A. Waters

Institute of Integrative and Comparative Biology,

Faculty of Biological Sciences University of Leeds,

Leeds LS2 9JT,

UK,

email:D.A.Waters@leeds.ac.uk

Steve McLaughlin

Institute for Digital Communications,

School of Engineering and Electronics,

the University of Edinburgh Alexander Graham Bell Bldg,

King's Buildings,

EH9 3JL,

Edinburgh,

UK,

email:sml@ee.ed.ac.uk[a)]

(Dated: July 31, 2009)

Abstract

In this paper, techniques for time-frequency analysis and investigation of bat echolocation calls are studied. Particularly, enhanced resolution techniques are developed and/or used in this specific context for the first time. When compared to traditional time-frequency representation methods, the proposed techniques are more capable of showing previously unseen features in the structure of bat echolocation calls. It should be emphasised that although the study is focused on bat echolocation recordings, the results are more general and applicable to many other types of signal.

PACS numbers: 43.60.Hj, 43.80.Ka

I. INTRODUCTION

Since the first description of the ultrasonic calls made by bats[1] and evidence that they are used for echolocation[2], various analytical techniques have been applied in order to characterise and investigate them. While the temporal characteristics of echolocation calls could be established readily even on simple oscilloscopes, the frequency characteristics proved more problematic for early technology. The fact that some bats used frequency-modulated signals going from high to low frequencies was recognised early on[3,4], and was subsequently followed by the discovery of constant frequency calls, such as those used by Rhinolophid bats[5,6]. High-speed tape recorders and zero-crossing techniques (e.g.[7]) allowed basic frequency parameters such as maximum and minimum frequencies to be determined, and permitted the classification of echolocation calls into different types[8]. Analogue techniques, such as the "Sound Spectrograph"[9], which was developed commercially as the Kay Electric Company sound spectrograph and first released in 1951, were then used to analyse signals in the joint time-frequency domain (e.g.[10]). Detailed descriptions of call structures, however, had to wait until technology was developed to acquire signals for digital analysis along with the development of more computationally efficient techniques. Fourier techniques became more widely available with the development of the Cooley-Tukey algorithm[11] allowing the use of the computationally more efficient Fast Fourier Transform (FFT). Such techniques could then be applied to the analysis of bat echolocation calls using digital computers for the purposes of understanding the structure of the call from the point of view of biosonar[12] as well as the relationship between call structure and foraging task[13].

Signal decomposition techniques are applied to bat echolocation calls for a number of reasons. Firstly, they allow the analysis and interpretation of echolocation call structure with reference to species identification[14], taxonomy[15], task[16], environment[17] and target[18]. This is usually done by the visual representation of the signal, either as a power spectrum

[a)]This work was performed as part of the BIAS consortium under a grant funded by the EPSRC under their Basic Technology Programme.

or sonogram via an STFT. Secondly, the parameters extracted from the signal via decomposition may be used to understand the mechanisms by which bats may determine target range[19], characteristics[20], and azimuth and elevation[21].

Fourier spectrogram based on the short time Fourier Transform (STFT) is the most widely used technique for analysing time-varying signals. With respect to the inspection of echolocation bat calls particularly, apart from limited exceptions (see for example[22,23]), it is the technique of choice. However, it is not the best method for time frequency (TF) analysis for the investigation of the frequency content of biologically produced signals.

The aim of this paper is two-fold. It first serves as a guide for those involved in the analysis of bat echolocation calls to facilitate the correct interpretation of the frequency content of such signals. Second, traditional techniques for the analysis of bat echolocation calls are compared with some less well known methods offering certain advantages. It should be emphasised that although the study is focused on bat echolocation types of signal, the results are more general and applicable to other type of signals.

II. DATA COLLECTION METHOD

The echolocation calls analysed and discussed in this paper are field recordings collected during 2007 and 2008. The recordings of *Pipistrellus pipistrellus*, *Pipistrellus pygmaeus* and *Myotis daubentonii* took place in the UK, while those of *Pipistrellis capensis*, *Pipistrellus kuhli* and *Tadarida aegyptiaca* were made in South Africa. In all cases, the bats were recorded in free flight while foraging with the microphone mounted on a pole with a maximum height of 3m to better align it with the foraging bat. Calls were obtained from bats flying free from clutter to remove interfering echoes, and with the heterodyned audio output of the bat detector switched off to eliminate any low-frequency feedback through the microphone and consequent low-frequency artifacts. In all cases, a Pettersson D-1000x bat detector with a solid dielectric microphone was used and the calls were sampled at 750 kHz and digitised with 16 bit resolution, streamed to a compact flash card in the device and saved in

the .wav file format standard. These calls were then sorted visually based on the sonograms generated in Batsound Pro (Pettersson Electronic, Sweden). Species identification was made with reference to distributions and call information in the literature and in the event of any ambiguity in identification, or any evidence of clipping, the affected call was not selected for analysis.

III. SPECTROGRAM

The spectrogram, which is essentially the squared modulus of the short time Fourier transform (STFT)[24,25], is one of the most widespread tools for time-frequency analysis of time-varying signals.

The STFT of a signal $x(t)$ is given by:

$$S_x^{(h)}(t,f) = \int x(u)h^*(u-t)e^{-j2\pi fu}du, \tag{1}$$

where $x(u)$ is the signal analysed and $h^*(t)$ is the complex conjugate of an analysis window localised around $t = 0$ and $f = 0$. The superscript $^{(h)}$ indicates the dependence of the STFT analysis on the characteristics, especially the length, of the analysis window used. If $h(t)$ were omitted, or equivalently set to 1 $\forall t$, then the STFT reduces to the Fourier transform. Otherwise, $h(u-t)$ suppresses the signal outside an area around the time instance t. In simple words, allowing the parameter t to take on all the values within the time interval where the signal is evolving centers the analysis window on t, and permits us to estimate via the FT the frequency content of the signal in the neighbourhood of t delimited by $h(t)$. Finally, the spectrogram, $SPEC_x^{(h)}(t,f) = |S_x^{(h)}(t,f)|^2$, can be loosely interpreted as the energy distribution of the signal over the time-frequency (TF) plane.

A. Spectrogram time-frequency resolution

According to Eq. (1) if a short analysis window, $h(u)$, is used, then only a short interval of the signal centered on t enters into the calculation of the STFT and the result of $S_x^{(h)}(t,f)$

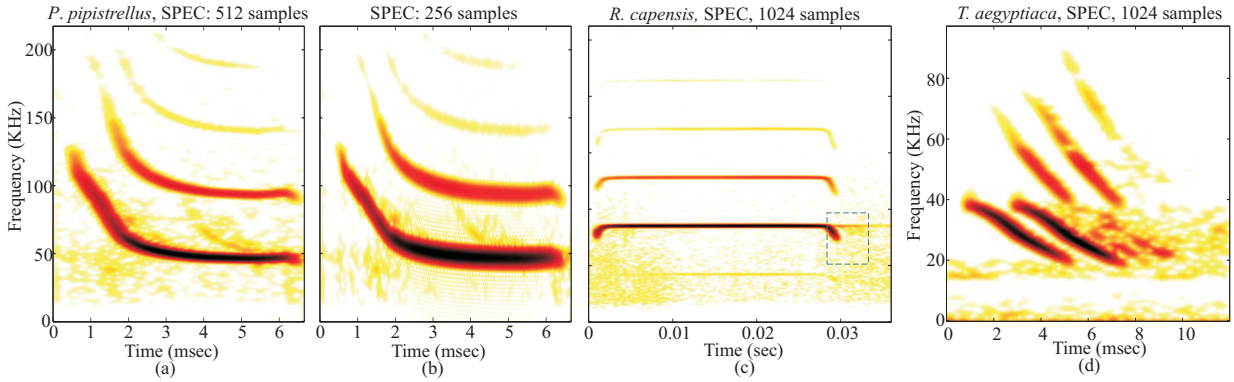

FIG. 1. (color online) (a-b) Spectrograms of echolocation calls produced by a *P. pipistrellus* bat using 512 and 256 samples long Gaussian analysis windows respectively. (c-d) Spectrograms of a *R. capensis* and a *T. aegyptiaca* bat call respectively.

is well localised around t. Otherwise, if $h(u)$ is large, then the outcome of the STFT is influenced by a larger portion of the signal, i.e. a longer time interval around the analysis time point t, leading to low time resolution in the sense that signal components having time separation shorter than the analysis window duration will contribute simultaneously to the estimate of the signal energy at the associated TF plane points, rendering them indistinguishable.

On the other hand, the bandwidth of the analysis window is responsible for the frequency resolution achieved. Specifically, the smaller the bandwidth of the analysis window is, the better the frequency localisation becomes providing improved frequency resolution.

Unfortunately, analysis windows having arbitrarily short length and small bandwidth at the same time cannot be constructed, due to the Heisenberg uncertainty principle, leading to a fundamental resolution trade off: Using a window with a short duration gives a high time resolution at the expense of a compromised frequency resolution, and vice versa.

B. Spectrograms of echolocation bat calls

Fig. 1(a-b) shows the spectrograms[26] of echolocation calls belonging to *Pipistrellus pipistrellus* bat. We observe that this is a multiharmonic signal with the fundamental frequency

component evolving from about 140 KHz to 50 KHz. Moreover, it has a characteristic nonlinear chirp shape consisting of two distinct parts. The first one is a fast time-varying down-chirp (from 0.5 msec to 2 msec) and the second one, corresponding to the rest of the signal, has a nearly constant frequency profile. In Fig. 1(a) the STFT uses a Gaussian analysis window that is 512 samples long, corresponding to about 0.68 msec time duration. This is a relatively long window that provides high frequency but low time resolution. The result is a good concentration of the spectrogram energy around the frequencies that correspond to the constant frequency part of the signal which appears thin compared to the down-chirp part. In fact, when dealing with fast time-varying signals (or signal sections), the adoption of long analysis windows can be problematic particularly if the signal changes significantly during the window duration. Indeed, improvement in the down-chirp frequency concentration, both in the fundamental component and the harmonics, can be achieved using a 256 samples long analysis window as is shown in Fig. 1(b). Inevitably, the shorter window leads to poor frequency localisation and consequently to a wide spectrogram around the actual frequencies of the constant frequency part.

The second echolocation call examined here belongs to a *R. capensis* bat and is a multiharmonic constant frequency signal having characteristic short downchirp tails. Its spectrogram, using a 1024 samples long analysis window, is shown in Fig. 1(c). In this case, a long enough analysis window guarantees enhanced frequency localisation but at the expense of a compromised time resolution. As a result, the exact time instances that the signal appears and disappears cannot be accurately determined by the spectrogram. In a similar fashion, the nonlinear tails slightly diverge from their actual shape.

The ability of STFT to analyse a signal that is corrupted by noise is investigated next using an echolocation call signal from *Tadarida aegyptiaca* (illustrated in Fig. 1(d)). The actual recording used here presents an additional challenge as it contains the echolocation calls of two bats emitted with a time difference of just 1 msec. The energy concentration of the specific signal can not be improved by using a different window. This is due to the slope that the chirps have.

Based on these examples, we can infer that the spectrogram via the STFT cannot serve as a generalised analysis tool for all types of bat calls. Moreover, as we saw in the case of the nonlinear chirp, the results are limited since a particular analysis window is not appropriate for the entire signal. However, there are signal shapes that lend themselves to an accurate interpretation of their spectrograms provided a suitable window length is used.

IV. A BROADER CLASS OF QUADRATIC TIME-FREQUENCY REPRESENTATIONS

The spectrogram, discussed above, and the scalogram belong to the class of energy TF distributions, with both being given by the square modulus of their respective linear transformations, namely the STFT and CWT. An alternative road to TF representations, which provides greater flexibility and in many cases better performance, is to directly estimate the signal energy on the TF plane[27,28]. A prominent example of this is the Wigner-Ville distribution (WVD):

$$WV_x(t,f) = \int x\left(t+\frac{u}{2}\right) x^*\left(t-\frac{u}{2}\right) e^{-j2\pi fu} du. \tag{2}$$

For the sake of clarity and simplicity, the WVD study is initially based on two artificial signals shown in Figs. 2(a) and (c). The first one is a combination of a constant frequency and a linear chirp signal whereas the second one is a nonlinear chirp signal. The WVD estimate corresponding to Fig. 2(a) is shown in Fig. 2(b). Two immediate observations can be made. First, the energy around each of the actual signal components, indicated by ST1, ST2 and ST3, is extremely well concentrated both in time and frequency, significantly outperforming the spectrogram (or scalogram). Indeed, the WVD offers exact localization of linear chirps[30]. Second, there are areas, clustered into three groups IT1, IT2 and IT3, that falsely show high energy. These are the result of interference terms that the WVD inherently produces, and are misleading with respect to the readability of the TF representation. It is pivotal to comprehend the generation mechanism and characteristics of the WVD artifacts, both for enhancing our ability to correctly interpret the TF representation of an unknown

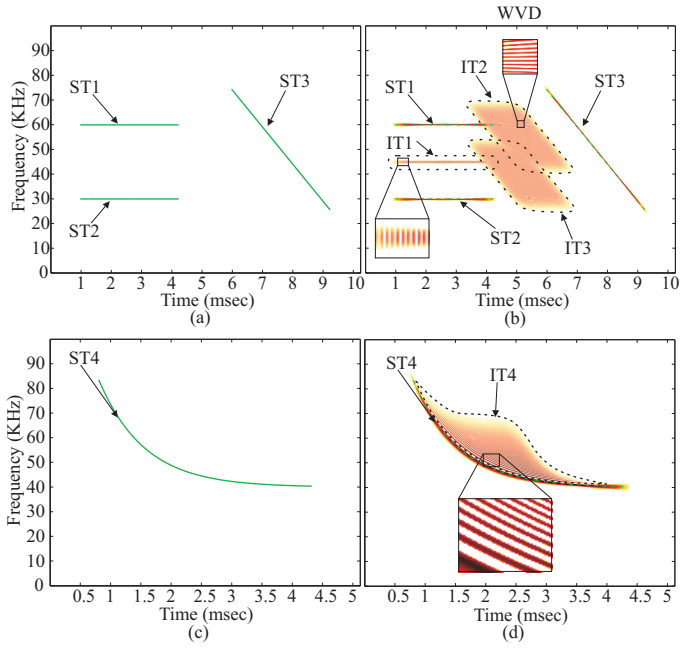

FIG. 2. (color online) (a) Test signal comprising two constant frequency harmonics and a linear chirp. (b) The corresponding WVD. (c-d) Nonlinear chirp test signal and the corresponding WVD respectively.

signal and for introducing techniques for interference reduction. As an example consider a signal x that is the sum of two distinct components x_1 and x_2, having energy concentrations around the points (t_1, f_1) and (t_2, f_2) respectively. Then, it is straightforward to show that $WV_x(t,f) = WV_{x_1}(t,f) + WV_{x_2}(t,f) + 2\mathrm{Re}\{WV_{x_1x_2}(t,f)\}$, where the operator Re indicates the real part. The first two terms are the signal terms (STs) that accurately depict the energies of the two signal components on the TF plane, whereas the third term is a cross-product interference term (IT) that appears as an additional, false, component. The IT is due to the interaction of the two signal components and is located exactly at the centre of the line segment connecting the points (t_1, f_1) and (t_2, f_2).

When the signal under consideration is more complex, such as in Fig. 2(a), then all of the signal components interact with one another to produce numerous ITs. Consequently, the interference term IT1 is the result of the interaction between the CF harmonics ST1 and ST2, and is found midway between these two signal terms. Similarly, IT2 is the result of

the interaction between ST3 and ST1. The width of the areas of these ITs is related to the relative distance between the points of ST3 and ST1. For example, the top-right (top-left) part of IT2 is due to the interaction of the rightmost (leftmost) part of ST1 with the topmost part of ST3. Similarly, IT3 is generated from the interactions between ST2 and ST3.

Fig. 2(d) shows the WVD of the nonlinear chirp of 2(c). The localisation of the nonlinear chirp is worse than that of the linear chirp but is superior to the spectrogram. Also, the observed ITs are not the result of interactions between signal parts belonging to different components, but are caused by different parts of the same signal component. For this reason they are often referred to as *inner interferences* to differentiate them from the *outer interferences* of Fig. 2(b). However, a clear distinction between the two types of interference terms does not exist since they are both generated by the same mechanism.

In the case of digital signals (which is true here), interferences between the signal components lying at the negative and possitive frequencies might appears. For that reason, the WVD is applied on the analytic signal version of the signal under consideration[30].

A. WVD Interference reduction

An important feature of the ITs is their sinusoidal oscillatory structure, evident in the magnified insets shown in Fig. 2(b),(d). We observe that the direction of the oscillations is perpendicular to the line connecting the interacting components (e.g., ST1 and ST2 or ST1 and ST3 in Fig. 2(b)). Moreover, the frequency of the oscillations is distance dependent and the further apart the signal components are, the faster the oscillations become. This can been seen in the magnified inset in Fig. 2(d). Indeed, we see that the frequency of the oscillations is changing as the ITs move away from the signal term, with the reason being that the distance between the signal parts that produce the distant ITs is greater.

The oscillatory nature of ITs can be exploited in order to eliminate them by means of smoothing. According to this framework the ITs smoothing is realised along the time and the frequency axis separately using two analysis window functions $g(t)$ and $h(t)$ respectively. The

resulting TF representation is referred to as smoothed pseudo-WV distribution (SPWVD) which is given by:

$$SPWV_x(t,f) = \int du \int d\tau \left\{ h^{'}(u) g(\tau - t) \right. \\ \left. x\left(\tau + \frac{u}{2}\right) x^*\left(\tau - \frac{u}{2}\right) e^{-j2\pi f u} \right\},$$

where $h^{'}(\tau) = h\left(\frac{1}{2}\tau\right) h^*\left(-\frac{1}{2}\tau\right)$.

The window $g(t)$ can reduce ITs evolving parallel to the time axis. The required length depends on the oscillation frequency and should cover at least a full oscillation period. In other words, the further apart two interacting components are the shorter $g(t)$ is allowed to be. On the other hand, $h(t)$ is capable of reducing ITs evolving parallel to the frequency axis. In contrast to $g(t)$, the shorter $h(t)$ is, the more intense the smoothing along frequency axis becomes. The reason is that $h(t)$ does not contribute directly to the overall smoothing function but via the Fourier transform of $h^{'}(t)$.

Unfortunately, apart from ITs elimination, the smoothing operation reduces the concentration of the actual signal components. As a result, there is always a trade-off between the time-frequency resolution achieved and the amount of ITs that will be eliminated through smoothing. In any case, the shorter $g(t)$ and the longer $h(t)$, that effectively reduces the ITs, should be used.

Fig. 3 shows the TF representation of the SPWVD of bat signals analysed in previous sections. In the case of *P. pipistrellus* echolocation call (Fig. 3(a)) the SPWVD yields a well-concentrated TF representation of the nonlinear chirp, especially in the down-chirp parts where the signal energy is strong enough. At the same time, the low amplitude signal harmonics are missed. The window functions $g(t)$ and $h(t)$ are Gaussian with lengths 201 and 551 samples respectively. Observe that there are some ITs between the harmonic components. These could be completely removed by adopting a longer window $g(t)$, say 301 samples long. However, this would lead to reduced concentration of the actual signal components.

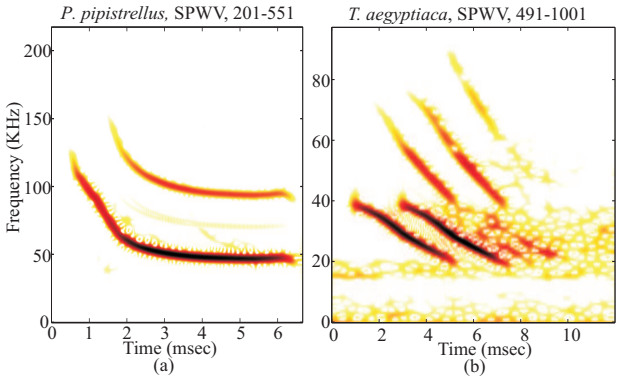

FIG. 3. (color online) SPWVD TF representations of echolocation calls produced by *P. pipistrellus* and *T. aegyptiaca* bats.

Fig. 3(b) shows the SPWVD distribution of *T. aegyptiaca* using $g(t)$ and $h(t)$ 491 and 1001 long respectively. We observe that the resulting TF-representation is significantly more concentrated around the actual frequencies than the spectrogram depicted in Fig. 1(d)). Moreover, both methods have similar sensitivity to noise but in the case of the SPWVD the noise is shown in a less compact way, forming cell-like structures. This, in some cases, can favor better readability.

With respect to the *R. capensis* echolocation call which spectrogram discussed in Fig. 1(c), the corresponding SPWVD do not offer any advantage so it is not shown.

V. HIGH RESOLUTION SPECTRUM TECHNIQUES

In this section we present an alternative adaptive filterbank approach[31] called the minimum variance estimator (MVE). The adaptation of a bank of filters to the signalleads to a significantly higher resolution compared to the STFT, albeit at a higher computational load.

Assuming that the signal is stationary or almost stationary over a segment of length L then the MVE aims to estimate the energy contribution that a specific f requency f makes to that signal. This is done by designing a finite impulse response (FIR) filter of length M that passes the signal component at f undistorted while attenuating the rest of the frequencies

as much as possible. Let us denote by $\mathbf{h}_f$ the coefficients vector of such an FIR filter, then the output of the convolution of the filter with the L samples long signal segment yields $K = L - M + 1$ samples given by

$$y_k(f) = \mathbf{h}_f^H(\omega)\mathbf{x}_k, \quad \text{for} \quad k = 1 \dots K, \tag{3}$$

where $\mathbf{x}_k = [x_k, x_{k+1}, \dots, x_{k+M-1}]^T$. The mean power over the K output samples is given by

$$\bar{P}(f) = \frac{1}{K}\sum_{k=1}^{K} |y_k(f)|^2 = \mathbf{h}_f^H \mathbf{R} \mathbf{h}_f \tag{4}$$

where $\mathbf{R} = \frac{1}{K}\sum_{k=1}^{k} \mathbf{x}_k \mathbf{x}_k^H$ is the $M \times M$ sample covariance matrix of the input data. The design problem can be expressed as the following constrained minimisation:

$$\min_{\mathbf{h}_f} \mathbf{h}_f^H \mathbf{R} \mathbf{h}_f, \quad \text{s.t. } \mathbf{h}_f^H \mathbf{s}(f) = 1, \tag{5}$$

with $\mathbf{s}(f) = [1, e^{j2\pi f}, e^{j4\pi f}, \dots, e^{j2(M-1)\pi f}]^T$ being the template at the frequency of interest. It can be shown that the solution to this problem is a filter of the form[31]:

$$\mathbf{h}_f = \frac{\mathbf{R}^{-1}\mathbf{s}(f)}{\mathbf{s}^H(f)\mathbf{R}^{-1}\mathbf{s}(f)}. \tag{6}$$

The MVE filter then strikes the best balance between the attenuation and the main lobe width[32]. Using a bank of these filters, with each one tuned to a different frequency on a specified grid, we can obtain a sampling of the overall frequency content of the signal on that grid.

The frequency resolution achieved by the MVE is determined by the filter length M. A longer filter would exhibit a narrower main lobe and higher sidelobe attenuation, yielding a higher frequency resolution. On the other hand, the reliability of the sample covariance matrix of the input data and consequently the $\bar{P}(f)$ estimates is given by the amount of averaging used (which is set by K). Therefore, for a fixed L, there is an inherent compromise between the resolution and reliability of the spectral estimate. It should be emphasized that, in order to ensure the matrix $\mathbf{R}$ is invertible, the minimum allowable value of K is $M + 1$. Equivalently, this translates to a total data segment $L \geq 2M$. Note that since $L = M+K+1$, a large M and/or K requires the availability of a long enough signal segment.

A. Implementation of the MVE and Performance Tradeoffs

Consider the general case where we have available N data samples of a non-stationary signal, and we seek to estimate its TF spectral distribution. Then, similarly to the STFT, the signal can be partitioned into a number of segments of length L, over which the signal can be assumed to be stationary or nearly stationary, and the MVE applied to each of those signal segments.

As pointed in the previous section, the implementation of the MVE spectrum estimator presents a number of compromises in setting the various inter-related parameters. At this point we introduce an extra parameter that facilitates the discussion. Let α specify the number of samples that L exceeds its minimum allowable value by, that is $L = 2M + \alpha$. Equivalently, we have $K = M + 1 + \alpha$. In order to specify the MVE algorithm, any two parameters among L, M, K and α need be set. Here we propose fixing the parameters M and α.

The larger that α is, the smoother the TF representation looks. Thus, α should be large enough in order for the MVE estimates to be reliable. We have found that a value in the interval $[30, 100]$ is acceptable. With respect to M, it is desirable to make it as large as possible in order to achieve high frequency resolution. However, this necessitates a large L which may violate the stationarity requirement. In fact, L must be kept small enough for the stationarity assumption to be, at least approximately, maintained. In the case of a chirp for instance, the faster the rate of change of the chirp frequency, the more severe the limit on L (and therefore on M) is. In other words, although a longer filter would lead to improved spectral resolution, the signal non-stationarity would negate this advantage as it blurs the spectrum itself. Having set M and α, the resulting length of the signal segments can be computed from $L = 2M + \alpha$.

When segmenting a data record of N samples into the blocks of length L, the same approach to the STFT can be used. That is, one can either use non-overlapping blocks or alternatively allow them to overlap. In the first case, the total number of segments will

be $N_t = \lfloor \frac{N}{L} \rfloor$. Thus, N_t determines the time points of the TF plane where the spectral estimates are computed. In practice, however, the total signal length N is usually not very large and the readability of the TF representation is enhanced by overlapping the blocks. If the segments overlap by Δ samples, the number of blocks N_t becomes

$$N_t = \left\lfloor \frac{N - L}{\Delta} \right\rfloor . \tag{7}$$

From this, the value of the overlap, Δ, can then be calculated to give the desired number of time points N_t. Note that whereas the sampling of the time axis is set by N_t, that of the frequency axis can be made arbitrarily fine by choosing a denser frequency grid. This is akin to the zero padding of the DFT and comes at the expense of an increased computational load.

B. MVE application on actual echolocation calls

A number of examples of MVE-based TF representations are shown in Fig 4. For comparison purposes, the spectrogram of the first echolocation call studied with MVE is given in Fig 4(a). This signal consists of 3 distinct components that originated from different sources. First, present is a call denoted by (A) together with one relatively weak harmonic belonging to a *P. capensis*. Second, another bat call is denoted by (B). And finally, the return of the strong harmonic of call (A) is designated by (C). Moreover, we observe that the down chirp part of (A) exhibits a frequency null (D) that is not well localised by the spectrogram in either the time or frequency. Figs. 4(b) and (c) present the outputs of the MVE with two filters of lengths M equal to 30 and 100 respectively with $\alpha = 30$ in both cases. Note that the filter with $M = 30$ is not long enough to adequately attenuate the interference. As a result the TF representation looks spiky with many disturbing interference lines between strong signal components. However, the frequency null is accurately reproduced in its full extent. Setting M to 100 enhances the frequency resolution and leads to a much clearer and well concentrated TF representation. Although there are still some artifacts especially between the two strong components (A) and (B) the accurate interpretation of the frequency

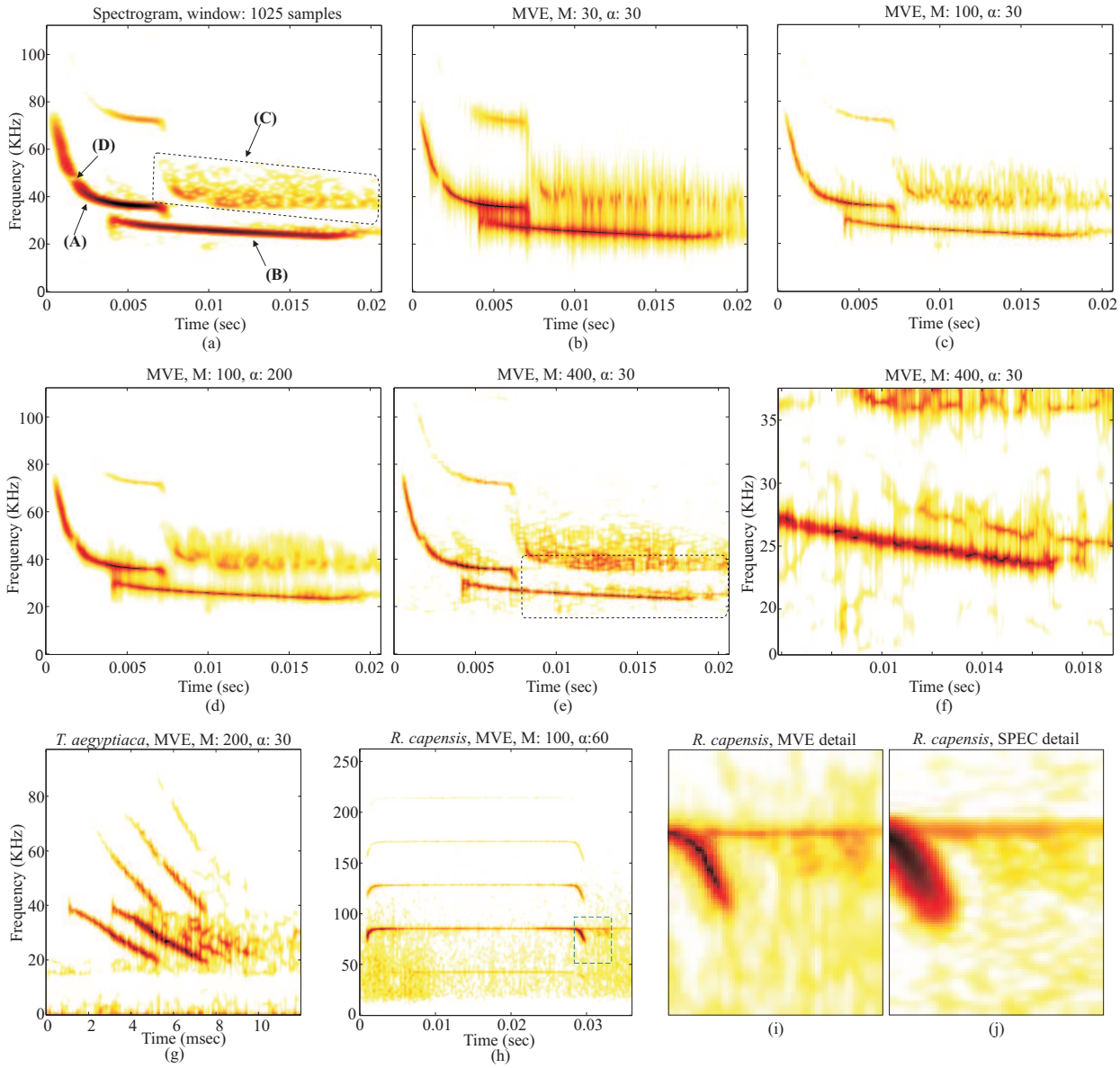

FIG. 4. (color online) Study of MVE on several echolocation calls.

content of the signal is not difficult. Additionally, the echolocation call return (C) is shown with high detail. However, the visibility of the frequency null is somewhat degraded by the longer filter used which then spans the duration of the null. Now recall that α determines the reliability (smoothness) of the TF representation. Using $\alpha = 30$ guaranties enough reliability (small variance of the magnitude estimates). However, if a still smoother TF representation is desired, α can be increased, to say 200 as shown in Fig. 4(d). Increasing the filter length to $M = 400$ a much greater frequency resolution can be achieved as is

evident in Fig. 4(e). The consequences, however, are threefold. First, the ability of the filter to attenuate interfering frequencies is enhanced leading to a clear distinction between strong neighboring signal components, e.g. (A) and (B). Also, components that are much closer together can now be distinguished. An example of this is given in the area marked by the dotted rectangle which is magnified and displayed in Fig. 4(f). An additional signal component slightly above component (B) is now visible. In fact, this is a delayed version of (B) resulting from its reflection by an object. Its delay with respect to (B) is similar to the delay of the return call (C) from its originating signal (A). This leads us to infer that the two bats were likely close together and their calls reflected from the same object. A third, negative consequence of the use of a longer filter is a relative loss of time resolution especially in the return (C) leading to fairly elongated signal parts in the low amplitude components. However, this does not have a significant effect on the strong components (such as the downchirp part of (A)) apart from the frequency null area which seems quite distorted. For completeness, the MVE TF representations of some of the echolocation calls investigated before are also given in Figs. 4(g) and (h). The MVE TF distribution of the *T. aegyptiaca* echolocation call appears somewhat sharper than those obtained from the spectrogram and SPWVD. In the *R. capensis* call case, we see a more concentrated localization of the signal components at the expense of a relatively poor accuracy in the power evaluation of all the harmonics but the strongest one. This is a disadvantage of MVE since it is known[33] that the MVE is a biased estimator of the amplitude. However, the high resolution achieved can be helpful in interpreting and extracting valuable information from the TF representation. Indeed, the detail revealed by the close up, shown in Fig. 4(i)), of the echolocation call "tail" (marked by the dotted rectangle in Fig. 4(h)) indicates that a return of the strong harmonic has been captured in the specific recording. Although this conclusion could be drawn from the corresponding spectrogram shown in Fig. 4(j), the MVE provides more details, allowing more accurate statements to be made about the nature and number of reflecting objects.

VI. SPECTRAL ANALYSIS BASED ON INSTANTANEOUS FREQUENCY ESTIMATES

Although the notion of Instantaneous Frequency (IF) was introduced decades ago (see e.g.[27]) there is still a lot of controversy regarding a proper definition as well as a physical interpretation for it[34–36]. Perhaps, the most widely accepted definition of the IF is as the derivative of the phase of the signal. More specifically, for a signal $x(t) = a(t)\cos(\phi(t))$ the IF in hertz is given by:

$$IF = \frac{1}{2\pi}\frac{d\phi}{dt}, \tag{8}$$

where, $\phi(t)$ is the phase and $a(t)$ the instantaneous amplitude (IA) of the signal. Now it is impossible to compute two unknowns, namely the amplitude and phase, from the observed real signal $x(t)$[37]. Thus, there have been numerous attempts to solve this issue and for a review and comparison of these corresponding methods we refer the interested reader to[38–42].

In this study, we adopt the analytic signal approach[37] according to which the signal, $x(t)$, under examination can be written in complex form as:

$$w(t) = x(t) + i\mathcal{H}[x(t)] = x(t) + iy(t), \tag{9}$$

where $\mathcal{H}[x(t)]$ is the Hilbert transform[43] of $x(t)$. The specific complex representation $w(t)$ is analytic in the sense that it contains the full spectral content of $x(t)$ solely in the positive frequencies. Accordingly, the IA and phase can unambiguously be derived as

$$a(t) = \sqrt{x^2(t) + y^2(t)} \tag{10}$$

and

$$\phi(t) = \arctan\left(\frac{y(t)}{x(t)}\right). \tag{11}$$

In order to efficiently use the IF for the study of echolocation calls we need first to understand its properties when dealing with different signals. It should be stated that IF estimation *does not offer an alternative technique for constructing a TF representation of a signal*. In other words, there is only a limited number of cases that the IF coincides with

the actual frequency content of a signal, and in general, the IF at a time t assumes values that do not correspond to any of the frequencies of the signal components revealed by the TF representation to constitute the signal at t. Nonetheless, the IF can be a useful tool for echolocation call study provided that the ability to correctly interpret it is first developed.

A. Interpretation and properties of IF

The behaviour of the IF of a signal depends strongly on whether the signal is monocomponent or multicomponent. By monocomponent we refer to a signal that, at any time instant, contains only a single frequency. In a broader, more relaxed sense, a signal can be characterised as monocomponent if it is highly narrow-band locally. With respect to echolocation calls, a monocomponent signal does not contain harmonics. In contrast, an echolocation call comprised of a number of harmonics is a multicomponent signal since it contains at any time instant more than one frequency. Moreover, multiple overlapped echolocation calls would make up a multicomponent signal regardless of whether they contain harmonics or not. Note, that the characterisation of a signal as multicomponent or monocomponent applies locally in time and does not refer to full length of the data record. In other words a signal can be monocomponent or multicomponent at various times as components of different frequencies appear and/or disappear.

Once the IF and IA of a signal have been computed, they can be combined in order to form a spectrogram-like plot $H(t, f)$ which, for any time instant t, is the IA, $a(t)$, of the signal at the frequency f which is active, determined by IF(t), and zero at the rest of the frequencies. $H(t, f)$ is usually refered to as Hilbert spectrum[44].

B. IF of monocomponent and multicomponent signals

When applied to a noiseless monocomponent signal, the IF provides exact estimation of the frequency content of the signal. This is a very useful property in cases of echolocation calls that have no, or extremely attenuated, harmonics. Unfortunately, the situation is much

more complicated in the case of multicomponent signals.

The behaviour of the IF of multicomponent signals is investigated first using the two-component signal depicted in Fig: 5(a). In particular, subfigures 1 and 2 show these two components individually while subfigure 3 displays their sum. They are chosen so that the amplitude of the first component decreases with time whereas that of the second component increases. The Hilbert spectrum of the signal is shown in Fig. 5(b1), together with the actual frequency tracks of the two components (indicated by dashed lines). The tracks represent parallel linear chirps, the lower of which belongs to the first component (shown in 5(a1)). The resulting IF plot exhibits a spiky form that clearly does not match the actual frequency content of the signal. That is to be expected since, at any time instant, the rate of change of the phase (which gives the IF) includes the effects of both instantaneous frequencies and it is not possible to decouple those two effects (in order to resolve the two tracks) using a single function estimating the IF. We will now briefly discuss some properties of the IF that are useful for the types of signals that are of interest to us in this study. Proofs and/or further details can be found in the following references[34,45–48]. The following are a number of properties of the IF that can be observed in Fig. 5(c):

1. The IF is a non-symmetric oscillation exhibiting spikes that point downwards or upwards. More specifically, these spikes point towards the component with the larger amplitude. Therefore, in the first half of the plot, the lower chirp has the larger amplitude (see Fig. 5(a1)) and consequently the spikes are directed down. The opposite is true for the second half of the signal.

2. Smoothing of the IF by applying moving average with a long enough analysis frame T, allows improved estimates of the frequency of the component with the larger amplitude to be obtained. Let the unweighted average[49] of the IF resulting from the smoothing procedure be donoted by IF_u. Then

$$IF_u(t) = \frac{1}{T}\int_{t-\frac{T}{2}}^{t+\frac{T}{2}} IF(u)du \tag{12}$$

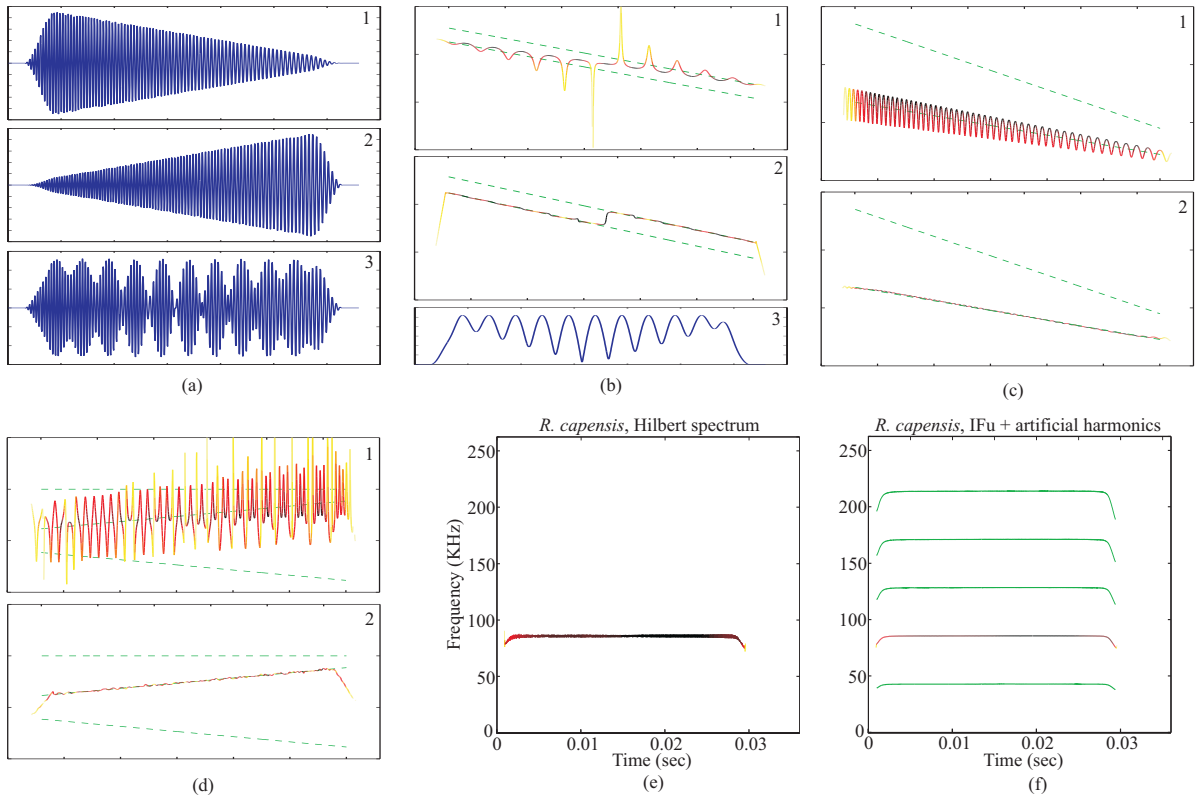

FIG. 5. (color online) (a-d) Analysis of multicomponent signals based IF estimates. (e,f) IF-based analysis of an R. capensis echolocation call

The frequency estimate, $IF_u(t)$, is shown in Fig. 5(b2). As expected, it locks on the first component during the first half of the signal duration (since that component has the larger amplitude) and then picks up the higher frequency component during the second half (as it now stronger). This property is quite useful and enables the estimation of the fundamental frequency of the echolocation bat calls as it usually is the stronger component of the signal.

3. The strength of the IF oscillations (that is the height of the spikes) depends on the relative amplitudes of the components. The closer the amplitudes are to each other, the larger the spikes become. On the other hand, when one component dominates the other (see for instance the beginning and end sections of the signal), then the spikes are small.

Besides the properties of the IF listed above, we mention that the instantaneous amplitude estimate, eq. (10), matches the envelope of the signal as shown in Fig. 5(b3).

Some additional properties of the IF can also be gleaned from another example, shown in Fig. 5(c1), of a multicomponent signal consisting of a linear down-chirp with a lower amplitude harmonic. Unlike the previous example, the amplitudes of both the fundamental and harmonic are constant in time whereas the difference (or separation) between their frequencies is time-varying, which is due to the fact that they are harmonics. Now we make the following observations:

4. The period of the oscillations of the IF depends on the difference between the frequencies of the two components. In fact, the larger this difference is, the faster the oscillations become.

5. As pointed out in property 4, the amplitude of the oscillations depends on the relative amplitudes of the components. Additionally, in the current example we see that the magnitude of the spikes depends also on the difference in the frequencies, becoming larger as this difference increases.

When more than two signal components come into play, a more complicated IF plot is generated as shown in Fig. 5(d1). Once again the actual frequencies of the components are represented by the dashed lines, and their amplitudes are constant and equal to 0.4, 1 and 0.5 respectively starting from the higher to the lower frequency. Most of the properties discussed above can be adapted to the case of three or more components. Note that the unweighted average is still valid and can be directly applied as shown in Fig. 5(d2). Indeed, the middle component is the dominant one here. However, we should emphasise that as the number of components increases, the dominant component of the signal has to be much stronger relative to the rest of the components in order for the IF_u estimate to be accurate.

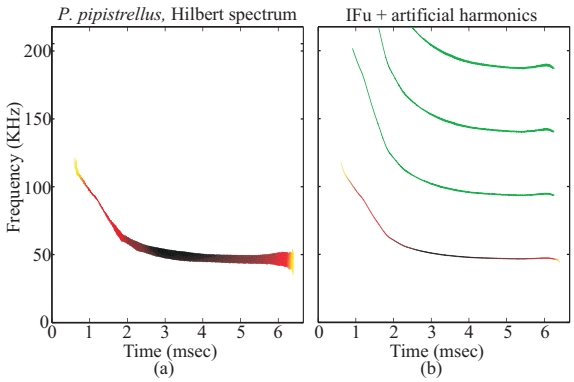

FIG. 6. (color online) IF-based analysis of a *P. pipistrellus* echolocation call

C. Echolocation Data Study

The first real echolocation call that we study with the IF method is an *R. capensis* signal. The spectrogram and MVE TF representations of this signal were presented in Figs. 1(d), 4(h) respectively, and the Hilbert spectrum was given in Fig. 5(e). The best the IF can achieve is to estimate a single frequency component. The result is a well-concentrated estimate of the highest amplitude component, which is the second lowest frequency component. In fact, the width of the curve is the result of the oscillations emanating from the presence of the other harmonics. However, as the amplitudes of the remaining components of the signal are very low relatively to the dominant one, the resulting estimate is quite accurate. This is a common scenario in bat echolocation calls. When the unweighted average is applied, the frequency estimate is forced to lock accurately on the dominant frequency component as shown by the second lowest component in Fig. 5(f). Moreover, if it were known that the estimated component is the second harmonic, we can use the unweighted average estimate to obtain the frequency tracks of other potential harmonics that may or may not be contained in the signal.

Next, we consider the application of the IF method to the *P. pipistrelus* echolocation call studied with previous methods. The resulting Hilbert spectrum is illustrated in Fig. 6(a). We observe that the frequency track is well concentrated in the down-chirp part but exhibits some very fast oscillation in the region of constant frequency. Again we stress that

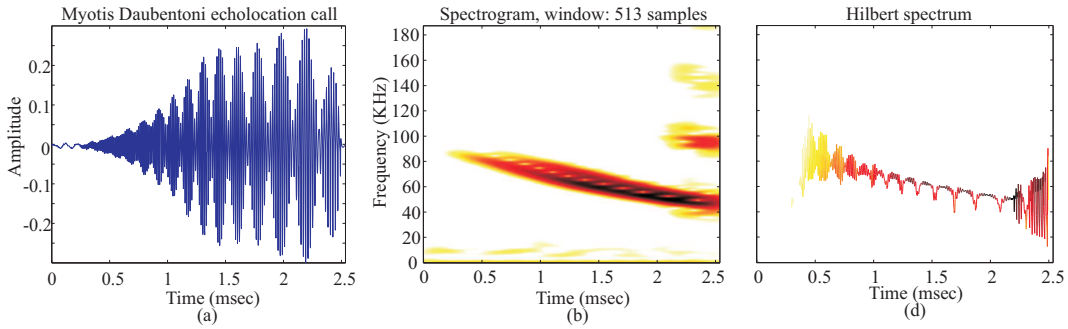

FIG. 7. (color online) IF-based anaysisis of *Myotis Daubentoni* echolocation call.

the degree of concentration of the IF estimates around a particular frequency is independent of the time- frequency resolution of the method. Instead, it depends only on the amplitudes of other signal components present at each time instant. Therefore, we infer from the observed frequency trajectory that the harmonics are largely attenuated during the down-chirp part, something that is confirmed by the related spectrogram shown in Fig. 1. This implies that the down-chirp part of the signal is effectively monocomponent. In the section of constant frequency, on the other hand, the harmonics are stronger leading to a coarser IF estimate. Again a more accurate estimate of the highest amplitude signal component can be computed using the unweighted average (especially as the harmonics are significantly weaker than the strongest component). This result, together with its potential harmonics, is depicted in Fig. 6(b).

Apart from harmonics estimation, IF-based analysis can also prove useful in cases that other TF representation methods do not provide results that are easy to interpret. Take as an example an echolocation call produced by a *Myotis Daubentoni* bat shown in Fig. 7(a). These specific bat calls often appear to have a special amplitude modulation. The spectrogram of this call is displayed in Fig. 7(b), and consists of a slightly nonlinear chirp that looks somewhat different than usual; It is wider than expected and has energy nulls along its length. With the aid of the Hilbert spectrum, useful information about both the spectrogram shape and the amplitude modulation can be drawn. As can be seen in Fig. 7(c) the IF exhibits spikes similar to those appearing when two signal components having

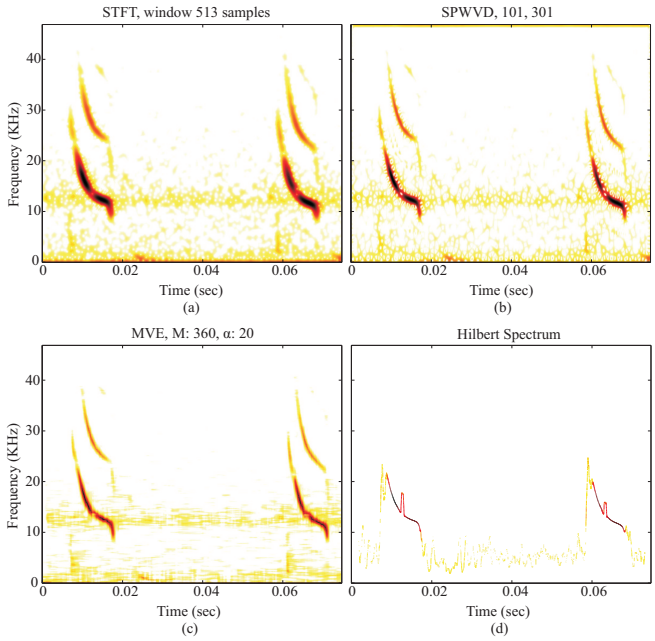

FIG. 8. (color online) Comparison of different methods when analysing two successive echolocation calls of a *P. pipistrellus* feeding-buzz.

closely spaced frequencies are active simultaneously (e.g. Fig. 5(b)). Moreover, due to the fact that the spikes point downwards, we can conclude that the signal component having the higher amplitude also has the lower frequency. This combination of two closed in frequency components is in agreement to the amplitude modulation that the signal exhibits.

VII. GENERAL DISCUSSION

Based on the above study, we can conclude that despite having a performance that is usually worse than its competitors, the STFT is often a reliable tool for analysis and investigation of bat echolocation calls especially in cases that the bat signals do not contain both constant frequency (or nearly constant frequency) and fast down chirp parts. However, one must always be mindful of the trade-off in the time-frequency resolution, which can make it hard to accurately determine the actual extent along the time or the frequency axis of the signal being investigated. Also a clear advantage of the STFT is its low computational cost.

The SPWVD is more flexible in dealing with signals comprising different types of echolo-

cation calls. With a proper selection of the smoothing windows, better performance than the STFT can be attained, albeit at a higher computational cost. A deep understanding of the generation mechanism of the interference terms is essential for two reasons. First, it assists in the proper tuning of the smoothing filters, and second it allows for a correct interpretation of the resulting TF representation. It should be noted that if the interference terms can be recognized, and therefore ignored, less smoothing would be needed in order to produce much better concentration and time-frequency resolution than the STFT.

To the best of our knowledge, the MVE has never before been applied to the study of bat echolocation calls. It appears to be a useful technique capable of revealing details "missed" by other methods. However, MVE time resolution is limited when it is necessary to use quite long filters (large values of M parameter). This is the case when the studied signal contains components closely spaced in frequency. Otherwise, somewhat disturbing interferences appear between the neighboring (in frequency) components. Moreover, it comes at a significant computational cost that is comparable or even larger than that of the SPWVD.

A final comparison of the above methods is shown in Fig. 8(a-c) concerning two successive echolocation calls of a *P. pipistrellus* feeding-buzz. At the buzz stage the echolocation calls exhibits abrupt frequency changes that limit the flexibility of the analysis window selection for the STFT. Indeed, the other methods offer more accurate TF representations.

IF analysis and Hilbert spectrum can not replace the TF representation methods. However, they can often provide extra, useful information and insight into the signal under consideration. The Hilbert spectrum of the unweighted average of the IF corresponding to the feeding-buzz is depicted in Fig. 8(d). Based on the results of section VI we would expect the unweighted average of the IF to coincide with the actual frequency of the harmonic having the higher amplitude. However, Fig. 8(d) this not to hold in this specific case since the IF estimate exhibits an abrupt "jump". This can be explained by the fact that the higher amplitude harmonic (the lower frequency one) has a spectral null, which can be noted from the TF representations 8(a-c). The IF estimate at the point of the frequency null, where the amplitude of the harmonic fades abruptly, tends to track the higher frequency harmonic.

At the time instant of the frequency null, the signal component with the larger amplitude is no longer the lower frequency harmonic, but the higher frequency one. Thus, the IF-based analysis is capable of revealing the frequency null and its temporal location.

VIII. CONCLUSION

In this paper, advanced frequency-content estimation techniques were exposed and applied in the context of biological signal analysis. It was shown that the proposed techniques are capable of showing previously unseen features in bat echolocation call structure.

References

[1] G. W. Pierce and D. R. Griffin, "Experimental determination of supersonic notes emitted by bats," Journal of Mammalogy **19**, 454–455 (1938).

[2] D. R. Griffin and R. Galambos, "The sensory basis of obstacle avoidance by flying bats," Journal Of Experimental Biology **86**, 481–506 (1941).

[3] D. R. Griffin, "Measurements of the ultrasonic cries of bats," Journal Of The Acoustical Society Of America **22**, 247–255 (1950).

[4] D. R. Griffin, "Ultrasonic sounds of bats, their acoustical dimensions and biological significance," Anatomical Record **108**, 540–540 (1950).

[5] F. P. Mohres, "The ultrasound orientation of bats," The Science of Nature **39**, 273–279 (1952).

[6] F. P. Mohres, "On the ultrasonic orientation of the horseshoe nosed bats (chiroptera-rhinolophinae)," journal of Comparative Physiology **34**, 547–588 (1953).

[7] D. A. Cahlander, J. J. G. McCue, and F. A. Webster, "The determination of distance by echolocating bats," Nature **201**, 544–546 (1964).

[8] D. R. Griffin and A. Novick, "Acoustic orientation of neotropical bats," Journal Of Experimental Biology **130**, 251–300 (1955).

[9] W. Koenig, H. K. Dunn, and L. Y. Lacy, "Sound specrograph," Journal Of The Acoustical Society Of America **18**, 19–49 (1946).

[10] A. Novick and J. R. Vaisnys, "Echolocation of flying insects by bat chilonycteris parnellii," Biological Bulletin **127**, 478–488 (1964).

[11] J. W. Cooley and J. W. Tukey, "An algorithm for machine calculation of complex fourier series," Mathematics Of Computation **19**, 297–301 (1965).

[12] J. A. Simmons and M. J. OÈFarrell, "Echolocation by long-eared bat, plecotus-phyllotis," Journal of Comparative Physiology **122**, 201–214 (1977).

[13] M. B. Fenton and G. P. Bell, "Echolocation and feeding-behavior in 4 species of myotis (chiroptera)," Canadian Journal of Zoology-Revue Canadienne De Zoologie **57**, 1271–1277 (1979).

[14] S. Parsons and G. Jones, "Acoustic identification of twelve species of echolocating bat by discriminant function analysis and artificial neural networks," Journal Of Experimental Biology **203**, 2641–2656 (2000).

[15] G. Jones, M. Morton, P. M. Hughes, and R. M. Budden, "Echolocation, flight morphology and foraging strategies of some west-african hipposiderid bats," Journal of Zoology **230**, 385–400 (1993).

[16] O. Berger-Tal, R. Berger-Tal, C. Korine, M. W. Holderied, and M. B. Fenton, "Echolocation calls produced by kuhl's pipistrelles in different flight situations," Journal Of Zoology **274**, 59–64 (2008).

[17] H. G. Broders, C. S. Findlay, and L. G. Zheng, "Effects of clutter on echolocation call structure of myotis septentrionalis and m-lucifugus," Journal of Mammalogy **85**, 273–281 (2004).

[18] J. H. Fullard, "Echolocation assemblages and their effects on moth auditory systems," Canadian Journal of Zoology-Revue Canadienne De Zoologie **60**, 2572–2576 (1982).

[19] I. Matsuo, K. Kunugiyama, and M. Yano, "An echolocation model for range discrimination of multiple closely spaced objects: Transformation of spectrogram into the reflected intensity distribution," Journal Of The Acoustical Society Of America **115**, 920–928

(2004).

[20] D. Genzel and L. Wiegrebe, "Time-variant spectral peak and notch detection in echolocation-call sequences in bats," Journal Of Experimental Biology **211**, 9–14 (2008).

[21] U. Firzlaff and G. Schuller, "Directionality of hearing in two cf/fm bats, pteronotus parnellii and rhinolophus rouxi," Hearing Research **197**, 74–86 (2004).

[22] R. A. Altes, "Wide-band, proportional-bandwidth wigner-ville analysis," IEEE Trans. Acoust., Speech, Signal Processing **38**, 1005–1012 (1990).

[23] R. A. Altes, "Signal processing for target recognition in biosonar," Neural Networks **8**, 1275–1295 (1995).

[24] S. N. Nawab and T. F. Quatieri, *Short-Time Fourier Transform* (Chapter in J.S Lim and A.V. Oppenheim, Prentice Hall, Englewood Cliffs, NJ) (1988).

[25] M. R. Portnoff, "Time-frequency representations of digital signals and systems based on short-time fourier analysis," IEEE Trans. Acoust., Speech, Signal Processing **28**, 55–69 (1980).

[26] The TF representations in the current section and section IV, are mainly produced with the aid of the Time-Frequency toolbox for Matlab[50].

[27] L. Cohen, *Time-Frequency Analysis: Theory and Applications* (Prentice Hall) (1995).

[28] F. Hlawatsch and G. F. Boudreaux-Bartels, "Linear and quadratic time-frequency signal representations," IEEE Signal Processing Mag. **9**, 21–67 (1992).

[29] L. Cohen, "Time-frequency distributions-a review," Proc. IEEE **77**, 941–981 (1989).

[30] P. Flandrin, *Time-Frequency/Time-scale analysis,* first edition (Academic Press) (1999).

[31] J. Capon, "High resolution frequency-wavenumber spectrum analysis," Proc. IEEE **57**, 1408–1418 (1969).

[32] An equivalent equation that directly gives the power estimate at the filter output bypassing the explicit filter impulse response computation is $\bar{P}_{min}(\omega) = \frac{1}{\mathbf{s}^H(\omega)\mathbf{R}^{-1}\mathbf{s}(\omega)}$.

[33] Y. Jiang, P. Stoica, and J. Li, "Array Signal Processing in the Known Waveform and Steering Vector Case," IEEE Trans. Signal Processing **52**, 23–35 (2004).

[34] L. Mandel, "Interpretation of instantaneous frequencies", American Journal of Physics

42, 840–846 (1973).

[35] B. Boashash, "Estimating and interpreting the instantaneous frequency of a signal - part 1: Fundamentals", Proc. IEEE **80**, 520–538 (1992).

[36] P. M. Oliveira and V. Barroso, "On the concept of instantaneous frequency", in *IEEE Int. Conf on Acoustics, Speech and Signal Processing, ICASSP 1998.* (1998).

[37] D. E. Vakman, *Signals, Oscillations, and Waves: A modern approach*, first edition (Artech House) (1998).

[38] B. Boashash, "Estimating and interpreting the instantaneous frequency of a signal - part 2: Algorithms and applications", Proc. IEEE **80**, 540–568 (1992).

[39] L. B. Fertig and J. H. McClellan, "Instantaneous frequency estimation using linear prediction with comparisons to the DESAs", IEEE Signal Processing Lett. **3**, 54–56 (1996).

[40] S. Mukhopadhyay and P. Sircar, "Parametric modelling of non-stationary signals: A unified approach", Signal Processing **60**, 135–152 (1997).

[41] P. Maragos, J. F. Kaiser, and T. F. Quatieri, "On amplitude and frequency demodulation using energy operators", IEEE Trans. Signal Processing **41**, 1532–1550 (1993).

[42] A. Potamianos and P. Maragos, "A comparison of the energy operator and the Hilbert transform approach to signal and speech demodulation", Signal Processing **37**, 95–120 (1994).

[43] S. Lawrence and J. Marple, "Computing the discrete-time analytic signal via fft", IEEE Trans. Signal Processing **47**, 2600–2603 (1999).

[44] It should be noted that Hilbert spectrum plots are usually related to a decomposition technique called empirical mode decomposition (EMD)[51] that aims to break a multicomponent signal into a number of monocomponent signals[52] in order to obtain accurate IF estimates. Unfortunately, EMD is rarely efficient for echolocation bat calls due to the fact that the high frequency harmonics are usually of low amplitude compared to the lower frequency components.

[45] A. Potamianos and P. Maragos, "Speech formant frequency and bandwidth traching using multiband energy demodulation", J. Acoust. Soc. Am. **99**, 3795–3806 (1996).

[46] P. J. Loughlin, "Time-varying frequencies of a signal", Proc. SPIE **3162**, 109–122 (1997).

[47] A. Potamianos and P. Maragos, "Speech analysis and synthesis using an am-fm modulation model", Signal Processing **28**, 195–209 (1999).

[48] L. Rossi and G. Girolami, "Instantaneous frequency and short term Fourier transforms: Application to piano sounds", J. Acoust. Soc. Am. **110**, 2412–2420 (2001).

[49] The term unweighted IF average is used to make a distinction with respect to the weighted IF average that appears in IF related literature (see for example[47]). Note that the weighted average is not suitable for the applications studied in this paper and is consequently not discussed here.

[50] F. Auger, P. Flandrin, P. Gonçalvès, and O. Lemoine, "Time-frequency toolbox, ver. 0.2", http://tftb.nongnu.org/ (date last viewed 31/07/09).

[51] N. E. Huang et. al., "The empirical mode decomposition and the hilbert spectrum for nonlinear and non-stationary time series analysis", Proc. R. Soc. Lond. A **454**, 903–995 (1998).

[52] Y. Kopsinis and S. McLauglin, "Investigation and performance enhancement of the empirical mode decomposition method based on a heuristic search optimization approach", IEEE Trans. Signal Processing, 1–13 (2008).

List of Figures