Separation of components from impulses in reassigned spectrograms

Sean A. Fulop[a)]
Department of Linguistics, California State University, Fresno, California 93740-8001

Kelly Fitz[b)]
School of Electrical Engineering and Computer Science, Washington State University, Pullman, Washington 99164-2752

(Received 13 September 2006; revised 1 December 2006; accepted 6 December 2006)

Two computational methods for pruning a reassigned spectrogram to show only quasisinusoidal components, or only impulses, or both, are presented mathematically and provided with step-by-step algorithms. Both methods compute the second-order mixed partial derivative of the short-time Fourier transform phase, and rely on the conditions that components and impulses are each well-represented by reassigned spectrographic points possessing particular values of this derivative. This use of the mixed second-order derivative was introduced by Nelson [J. Acoust. Soc. Am. **110**, 2575–2592 (2001)] but here our goals are to completely describe the computation of this derivative in a way that highlights the relations to the two most influential methods of computing a reassigned spectrogram, and also to demonstrate the utility of this technique for plotting spectrograms showing line components or impulses while excluding most other points. When applied to speech signals, vocal tract resonances (formants) or glottal pulsations can be effectively isolated in expanded views of the phonation process. [DOI: 10.1121/1.2431329]

PACS number(s): 43.60.Hj, 43.72.Ar [EJS] Pages: 1510–1518

I. INTRODUCTION

A. Overview

This paper presents signal processing techniques which extend previous work[2] on the reassigned (or *time-corrected instantaneous frequency*) spectrogram. The procedures discussed herein allow a kind of spectrogram (i.e., time-frequency analysis) to be computed which isolates either the line components or the impulses comprising a signal (or both together). Implementable step-by-step algorithms for applying these methods are presented here for the first time, together with examples of fruitful application to speech signals.

Immediately in the sequel we first recap previous descriptions of the reassigned spectrogram with examples. We then in Sec. II describe the theory behind using higher-order derivatives of the short-time Fourier transform phase (complex argument) to "prune" reassigned spectrograms and thereby show only those points meeting certain physical conditions. Numerous example images complement the presentation of these ideas, which were first set down by Nelson.[3] In Sec. III, two possible methods for computing the higher-order short-time Fourier transform (STFT) phase derivatives are presented with algorithms that invite the reader's implementation. The mathematical derivation of an exact transform method for computing the second-order mixed partial derivative of the STFT phase is included in this section. The paper concludes with Sec. IV which discusses the physical interpretation of this second-order derivative, and relates it to a previously published notion[4] of *consensus* among reassigned instantaneous frequencies.

B. Digression on the reassigned spectrogram

We here briefly review the reassigned spectrogram, which is discussed more fully in previous work.[2] We assume our signal $f(t)$ can be modeled as the sum of general AM/FM components:

$$f(t) = \sum_n A_n(t) e^{i[\Omega_n(t)+\phi_n]} \tag{1}$$

and that the STFT is defined in the following way:

$$\mathrm{STFT}_h(\omega, T) = \int_{-\infty}^{\infty} f(t+T) h(-t) e^{-i\omega t} dt. \tag{2}$$

This form of the transform is equivalent modulo a phase factor $e^{i\omega t}$ to the more prevalent form in which the window is time-translated with the signal held to a fixed time [see Eq. (9)].

The *channelized instantaneous frequency* (CIF) of a signal as a function of time and frequency is

$$\mathrm{CIF}(\omega, T) = \frac{\partial}{\partial T} \arg(\mathrm{STFT}_h(\omega, T)), \tag{3}$$

where STFT_h is the short-time Fourier transform using window function h. If there is just one AM/FM component dominant in the neighborhood of a frequency bin, then the CIF spectrum will show the instantaneous frequency of

a)Author to whom correspondence should be addressed. Department of Linguistics PB92, California State University, 5245 N. Backer Ave., Fresno, CA 93740-8001. Phone: (559) 278-4896. Electronic mail: sfulop@csufresno.edu
b)Present address: Starkey Hearing Research Center, 2150 Shattuck Ave., Suite 408, Berkeley, CA 94704.

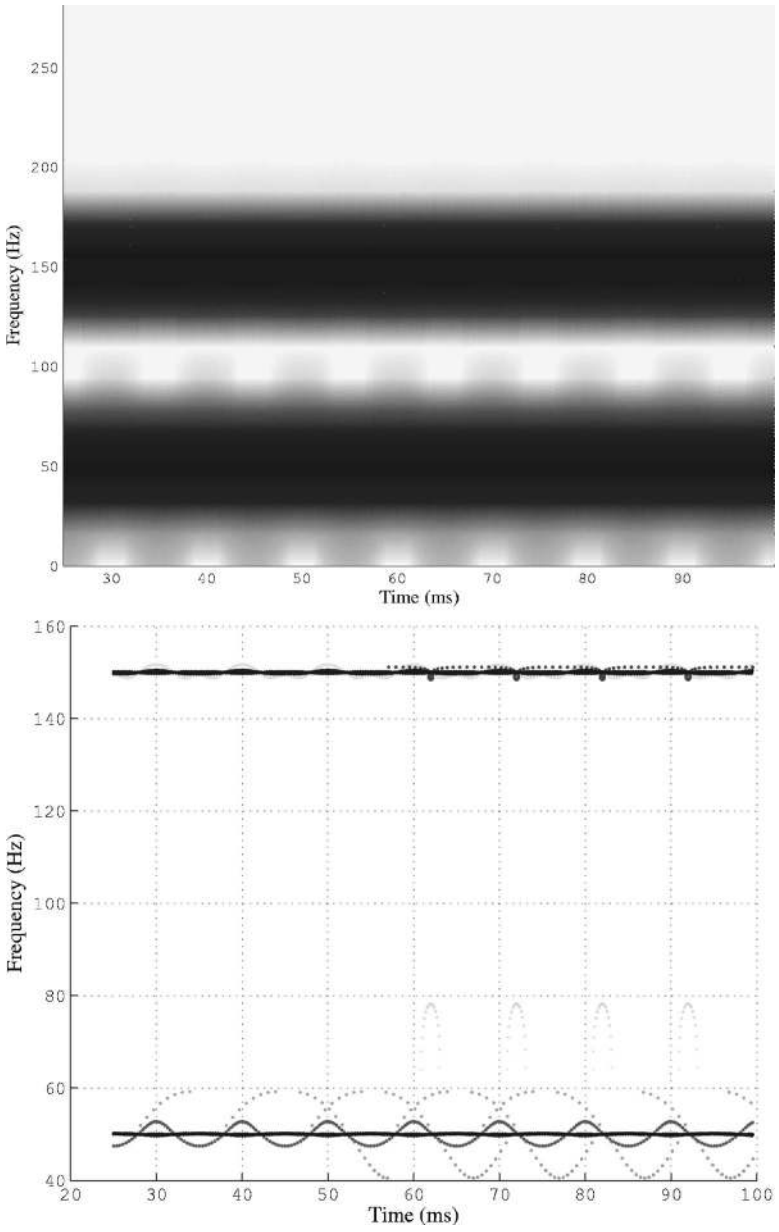


FIG. 1. The figure contrasts the conventional spectrogram (upper panel) with the reassigned spectrogram of a signal composed from a 50 Hz and 150 Hz sine wave for 1/8 second, sampled at 32 kHz. Some interference artifacts can be noted. Both images computed using 1600 point frames and 10 point frame advance.

that component with arbitrary precision (i.e., not quantized by the discrete time-frequency grid).

The *local group delay* (LGD) of a signal is given by

$$\mathrm{LGD}(\omega,T) = -\frac{\partial}{\partial \omega}\arg\left(\mathrm{STFT}_h(\omega,T)\right). \tag{4}$$

The reassigned spectrogram plots each STFT magnitude at the new location of its computed CIF, and at a time point corrected by its LGD. The LGD at a time-frequency point (ω_0, T_0) on the STFT matrix may be interpreted as the average true time of $\mathrm{STFT}_h(\omega_0, T_0)$; this is an estimate of the "time correction" to the maximum energy point of the dominant AM/FM signal component observed at (ω_0, T_0).

The new time-frequency coordinate (computed using the CIF and LGD) is a more meaningful location for the STFT magnitude, and corresponds to the local mean of the complex time-frequency energy distribution of the signal.[5] The resulting increase in spectrographic imaging precision is illustrated in Figs. 1–3.

II. REASSIGNED SPECTROGRAMS WITH PARTIAL DERIVATIVE THRESHOLDING

Despite the obvious gains in clarity with which the location and frequency modulation of line components can be shown in these reassigned spectrograms, as well as the improved time localization of impulsive events, the images can be disappointingly noisy. In the figures thus far, one can note random speckle and interference not clearly associated with either a line component or an impulse.

This is owing to numerous factors, but in rough sketch, the algorithm employed to locate the AM/FM components in the signal has a meaningful output only in the neighborhood of a component. Where there is no component of significant amplitude, the time-frequency locations of the points to be plotted can become random. There can also be a variety of other computation artifacts resulting from failure of the separability condition, among other circumstances. We next detail a method which has the potential to "denoise" our spectro-

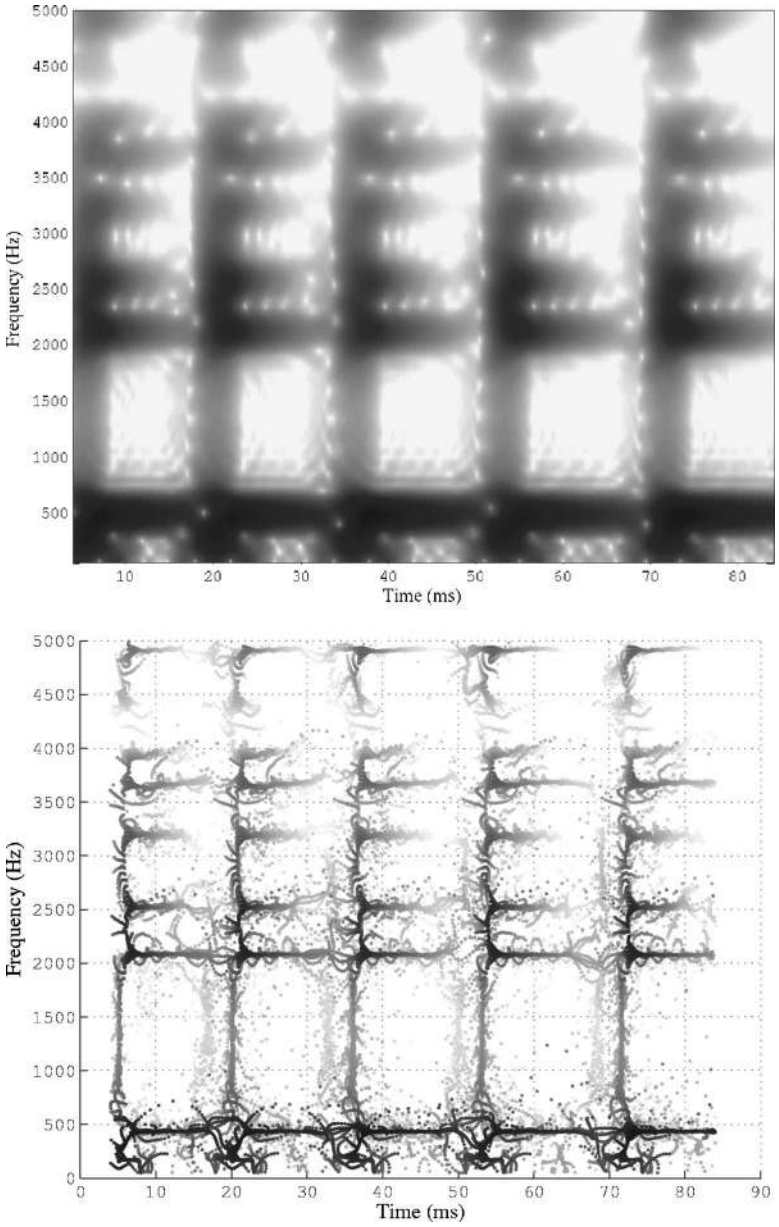


FIG. 2. The conventional spectrogram (upper panel) shows a portion of the vowel [e] (day) pronounced with "creaky voice" for low airflow; the reassigned spectrogram of the lower panel shows the same signal, in which the individual pulsations of the vocal cords are rendered clearly visible along with the mouth resonances (formants) which they excite at each impulse. All speech examples in this paper are sampled at 51.2 kHz. Both images computed using 400 point frames and 4 point frame advance.

grams by isolating quasistationary (low chirp rate) components in a display, or alternatively to permit highly time-localized events (impulses) to be isolated.

A. Interpretation of the mixed partial derivative

The technique to be expounded was first theoretically outlined by Nelson,[3] and involves the computation of the second-order mixed partial derivative of the STFT phase. This is equivalent to either the frequency derivative of the CIF, or to the time derivative of the LGD, since it is a fundamental theorem of calculus that the mixed partial derivative can be taken in either order.

Nelson (op. cit.) argued that the nearly stationary AM/FM components of a signal $x(T)$ satisfy

$$\frac{\partial^2}{\partial\omega\,\partial T}\arg(\mathrm{STFT}_h(\omega,T)) = \frac{\partial}{\partial\omega}\mathrm{CIF}_x(\omega,T) \approx 0. \qquad (5)$$

Further explication of this fact is deferred to a later discussion. By plotting just those points in a reassigned spectrogram meeting this condition on the second-order mixed partial phase derivative to within a threshold, a spectrogram showing just the line components can be drawn.

Nelson (op. cit.) further asserted that the impulses in a signal $x(T)$ satisfy

$$\frac{\partial^2}{\partial T\,\partial\omega}\arg(\mathrm{STFT}_h(\omega,T)) = \frac{\partial}{\partial T}\mathrm{LGD}_x(\omega,T) \approx 1. \qquad (6)$$

By plotting just those points meeting this condition to within a threshold, a spectrogram showing just the impulsive events in a signal can alternatively be drawn. Plotting all points meeting the disjunction of the above conditions results in a "de-noised" spectrogram showing quasisinusoidal components and impulses together, to the exclusion of most everything else.

B. Application to the reassigned spectrogram

To selectively plot components meeting the condition of a quasisinusoid, one keeps only those points having a value

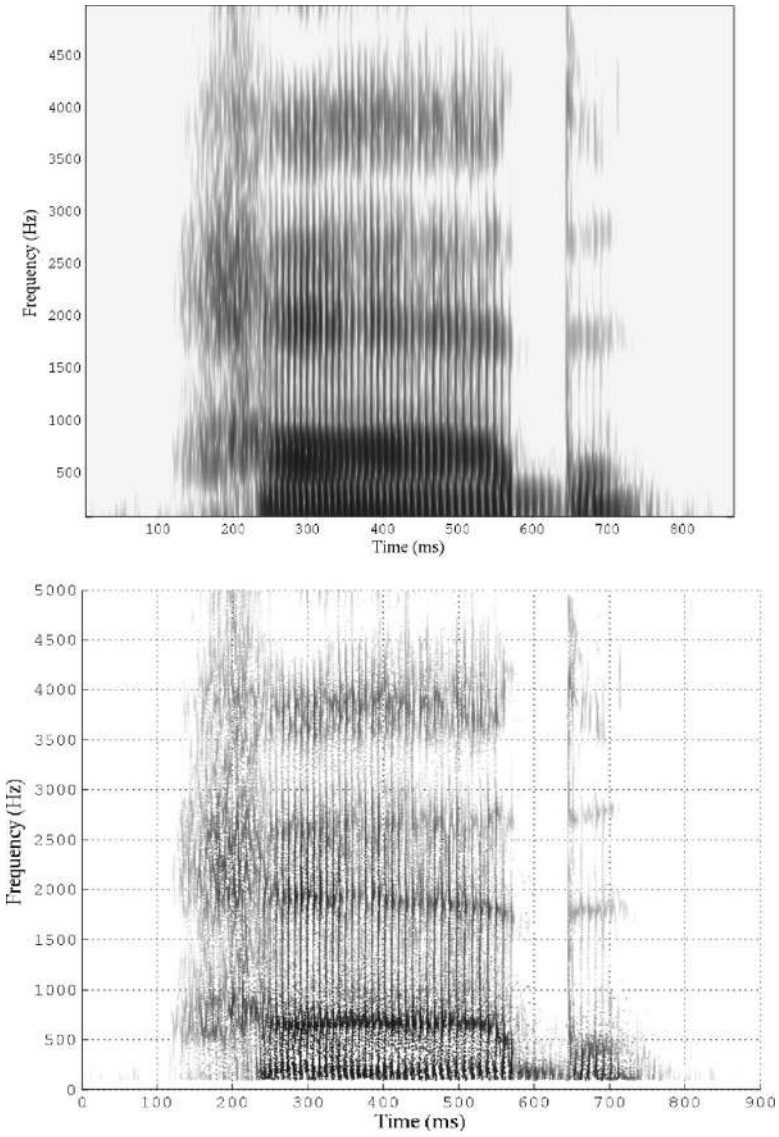


FIG. 3. This figure shows a conventional spectrogram of the English word *had* (upper panel, computed using 300 point frames and 20 point frame advance), along with a reassigned spectrogram below (computed with 40 point frame advance).

of $\frac{\partial^2}{\partial T \partial \omega} \arg(\mathrm{STFT}_h(\omega, T))$ sufficiently near 0, when one computes the derivative using Nelson's definition. The precise threshold can be empirically determined, and will in practice depend on the degree of deviation from a pure sinusoid that is tolerable in the application at hand. This means that greater tolerance in this threshold will be required where line components having high chirp rates are expected—for speech signals an absolute value of the derivative on the order of 0.2 is often a reasonable threshold (see Figs. 4 and 5). On the other hand, a numerical derivative threshold value which is several orders of magnitude smaller can be used to eliminate nearly every point that does not represent a pure sinusoid with no frequency modulation, as is illustrated in Fig. 6.

By Nelson's definition of the short-time Fourier transform, points whose second-order mixed partial phase derivative is near 1 are likely to be associated with impulsive events. For display purposes it is appropriate to be quite tolerant in this threshold, depending on what sort of signal content we desire to regard as "impulsive." A derivative value between 0.75 and 1.25 usually yields good results for speech signals, without straying too far from identifiably impulse-like events (see Figs. 7 and 8).

III. COMPUTING THE HIGHER-ORDER MIXED PARTIAL DERIVATIVE

A. Cross-spectral method

The mathematical theory behind cross-spectral expressions for all higher-order partial derivatives of the STFT phase is completely presented in prior literature,[3] and we have relied on this in developing the particular algorithm for the second-order mixed partial derivative that is presented below. The steps in the computational method will be based on the "Nelson method" algorithm for the reassigned spectrogram published by Fulop and Fitz.[2] Readers are invited to refer to that algorithm to complement that presented below. It is important to note that, just as with the cross-spectral method for computing the first-order STFT phase derivatives (and thereby the reassigned spectrogram) discussed in previous work,[1,2] the method presented here will compute an *approximation* of the second-order mixed partial STFT phase

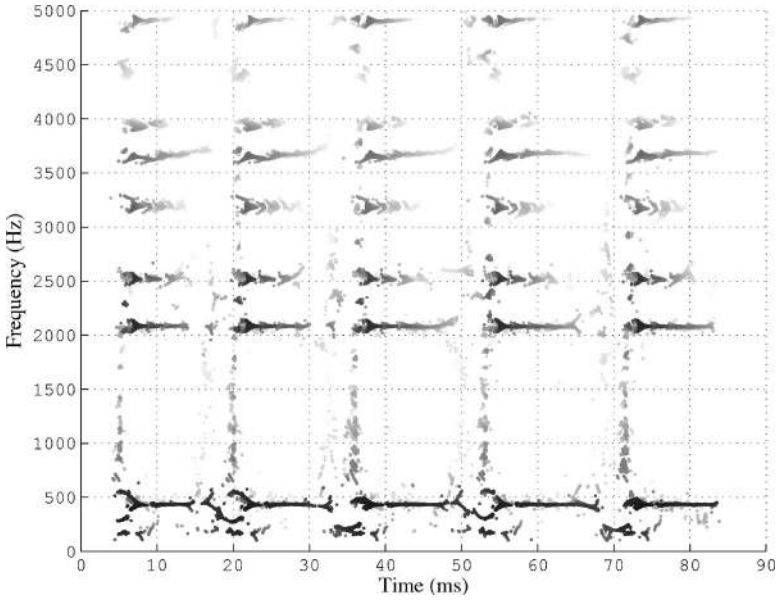


FIG. 4. Another view of the creaky voiced vowel shown in Fig. 2, showing only those points whose absolute value mixed partial derivative of the STFT phase is less than 0.1.

derivative. This approximation is generally so close that it might not matter for practical purposes.

(1) First, one builds two matrices `S` and `Sdel` of windowed signal frames of length `win_size` (user-supplied to the procedure) time samples, with `Sdel` having frames that are delayed by one sample with respect to `S`. For the present purposes a standard Hann window function will suffice, but other windows may be more appropriate for other applications. The windowed signal frames overlap by the same user-input number of points in each of the matrices.

(2) One next computes four short-time Fourier transform matrices; each column is an `fftn`-length-Fourier transform of a signal frame, computed with a fast Fourier transform function called `fft`. The length value `fftn` is supplied by the user. The difference between `fftn` and `win_size` is zero-padded up to `fftn` for the computation.

`STFTdel=fft(Sdel)`

`STFT=fft(S)`

`STFTfreqdel` is just `STFT` rotated by one frequency bin—this can be accomplished by shifting the rows in `STFT` up by one step and moving the former last row to the new first row.

`STFTfrtimedel` is `STFTdel` similarly rotated by one frequency bin.

(3) Next compute a cross-spectral surface by applying Nelson's[3] theory:

$$\texttt{MixCIF} = \texttt{STFT} \times \texttt{STFTdel}^{*} \times (\texttt{STFTfreqdel} \times \texttt{STFTfrtimedel}^{*})^{*}, \tag{7}$$

where the notation X^{*} for complex X indicates the complex conjugate (pointwise if X is a matrix of complex numbers). The notation $A \times B$ for matrices A, B denotes a point-by-point product, not a matrix multiplication.

(4) Now the partial frequency derivative of the channelized instantaneous frequency can be computed:

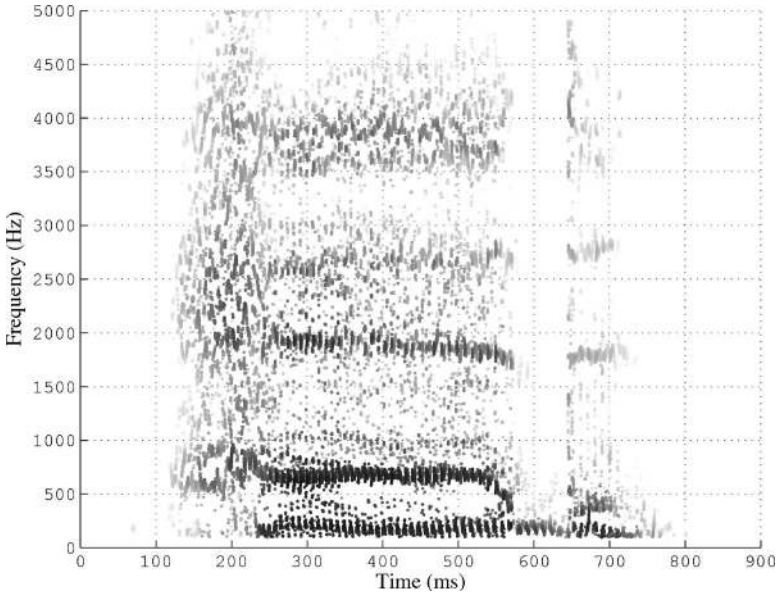


FIG. 5. Another view of the word *had* from Fig. 3, this time showing only points having $|\frac{\partial^2}{\partial\omega\partial T}\arg(\mathrm{STFT}_h(\omega,T))| < 0.1$.

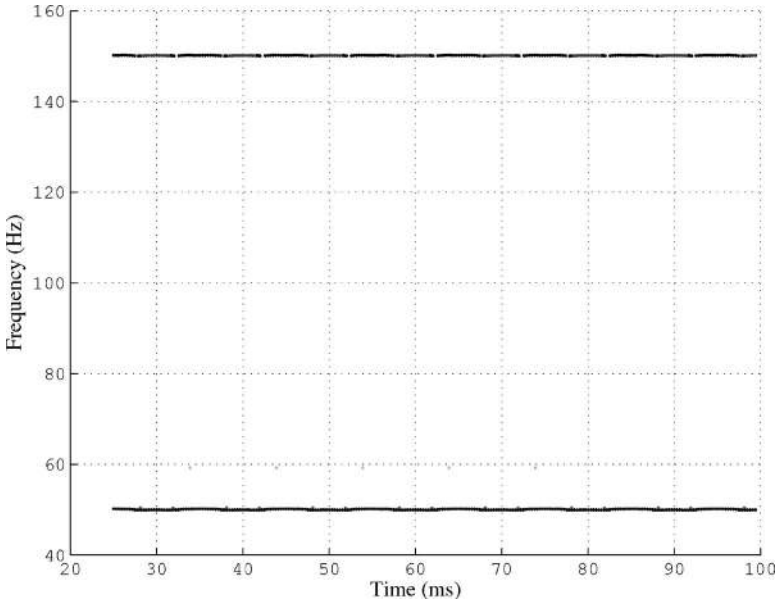


FIG. 6. Another view of the double sine wave shown in Fig. 1, this time showing only those points whose absolute value second-order mixed partial derivative of the STFT phase is less than 10^{-4}.

$$\texttt{CIFderiv} = \frac{\texttt{fftn} \cdot \texttt{Fs}}{2\pi \cdot \texttt{win\_size}} \times \arg(\texttt{MixCIF}) \times \arg(\texttt{MixCIF}), \tag{8}$$

where the arg(·) function is valued on the range $(0, 2\pi)$, and `Fs` is the sampling rate (in Hz) of the signal.

The final quantity computed by the above algorithm is equivalent to the partial time derivative of the local group delay, and either of these represents the (unique) second-order mixed partial derivative of the STFT phase.

B. Exact transform method

This subsection presents the mathematical theory behind an alternative approach to Nelson's cross-spectral theory, which can be called the "exact transform approach." While the originators of this technique for defining a reassigned spectrogram should be recognized as Auger and Flandrin,[6] in fact the method it employs for computing the phase derivative of the STFT was used to compute the group delay in a number of earlier papers by Yegnanarayana and collaborators,[7] and they in turn credit the basis for the mathematics to Oppenheim and Shafer's digital signal processing textbook.[8] This approach has not heretofore been extended to compute higher-order partial derivatives as below. To align with the literature on the exact transform approach to the reassigned spectrogram, the short-time Fourier transform is now defined as a complex function of continuous time t and radian frequency ω by

$$X(t,\omega) = \int x(\tau) h^*(t-\tau) e^{-j\omega\tau} d\tau \tag{9}$$

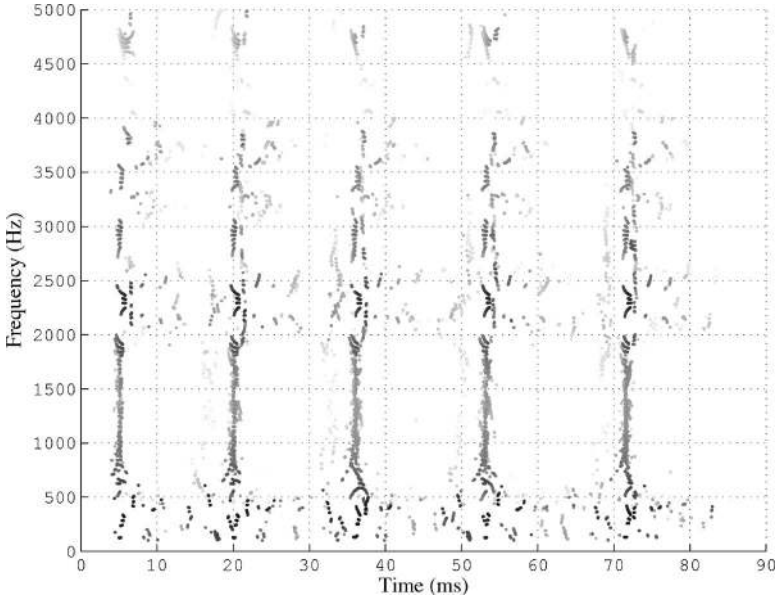
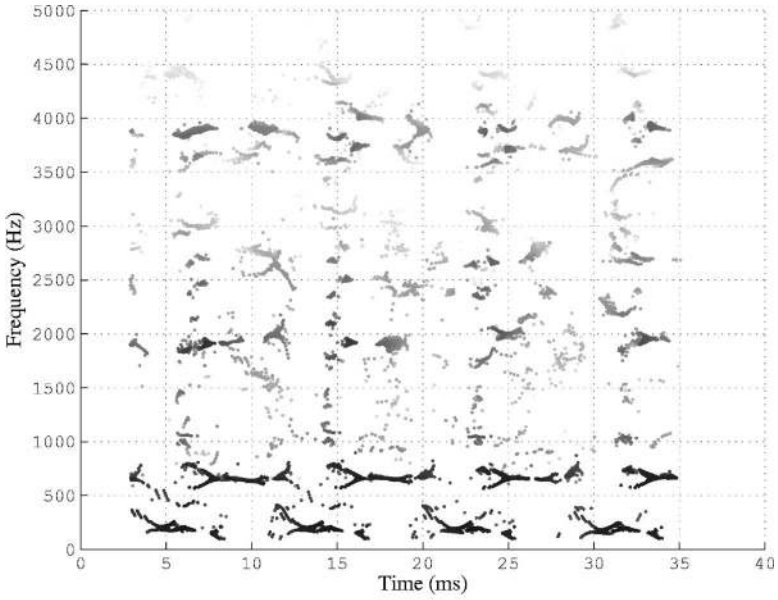


FIG. 7. This reassigned spectrogram shows the same vowel as in Figs. 2 and 5, but now a partial derivative threshold of time-correction characteristic of impulses has been applied, so only those points having $\left|\frac{\partial^2}{\partial\omega\partial T}\arg(\mathrm{STFT}_h(\omega,T)) - 1\right| < 0.25$ are plotted.



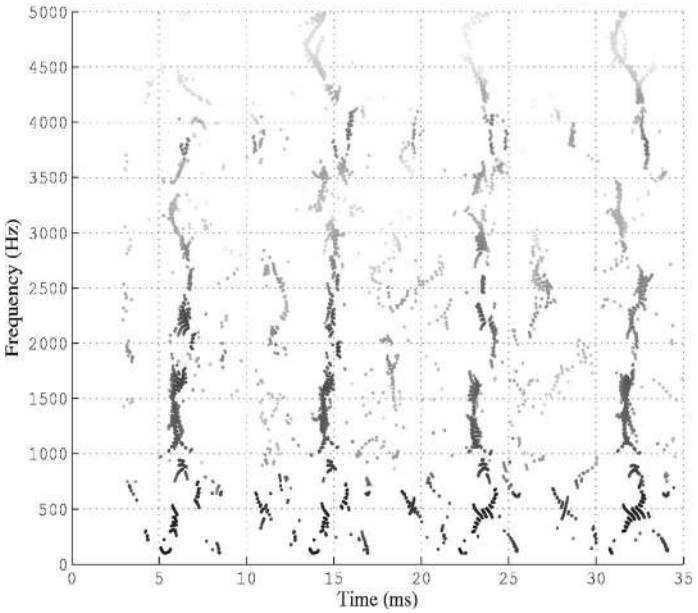


FIG. 8. These spectrograms zoom in on a few vocal cord pulsations from the vowel [æ] in the utterance of *had* shown in Figs. 3 and 6. The upper panel shows just the line components, while the lower panel focuses instead on the impulsive events. Both images computed with 300 point frames and 4 point frame advance.

$$= M(t,\omega)e^{j\phi(t,\omega)} \tag{10}$$

where $h(t)$ is a finite-length, real-valued window function, $M(t,\omega)$ is the magnitude of the short-time Fourier transform, and $\phi(t,\omega)$ is its phase.

Since

$$\frac{\partial X(t,\omega)}{\partial t} = \frac{\partial M(t,\omega)}{\partial t}e^{j\phi(t,\omega)} + j\frac{\partial \phi(t,\omega)}{\partial t}X(t,\omega), \tag{11}$$

the mixed partial derivative of the short-time Fourier transform with respect to time t and frequency ω can be expressed as a sum of two terms:

$$\frac{\partial^2 X(t,\omega)}{\partial t\,\partial\,\omega} = \frac{\partial}{\partial \omega}\left[\frac{\partial M(t,\omega)}{\partial t}e^{j\phi(t,\omega)}\right] + j\frac{\partial}{\partial \omega}\left[\frac{\partial \phi(t,\omega)}{\partial t}X(t,\omega)\right]. \tag{12}$$

The first term on the right-hand side of Eq. (12) can be written

$$\begin{aligned}\frac{\partial}{\partial \omega}\left[\frac{\partial M(t,\omega)}{\partial t}e^{j\phi(t,\omega)}\right] &= \frac{\partial^2 M(t,\omega)}{\partial t\,\partial\,\omega}e^{j\phi(t,\omega)} \\ &+ jX(t,\omega)\Re\left\{\frac{X_{\mathcal{D}h}(t,\omega)}{X(t,\omega)}\right\}\Re\left\{\frac{X_{\mathcal{T}h}(t,\omega)}{X(t,\omega)}\right\} \\ &- jtX(t,\omega)\Re\left\{\frac{X_{\mathcal{D}h}(t,\omega)}{X(t,\omega)}\right\}\end{aligned} \tag{13}$$

and the second term on the right-hand side of Eq. (12) can be written

$$j\frac{\partial}{\partial\omega}\left[\frac{\partial\phi(t,\omega)}{\partial t}X(t,\omega)\right]=tX(t,\omega)\Im\left\{\frac{X_{\mathcal{D}h}(t,\omega)}{X(t,\omega)}\right\}-X_{\mathcal{T}h}(t,\omega)\Im\left\{\frac{X_{\mathcal{D}h}(t,\omega)}{X(t,\omega)}\right\}+jX(t,\omega)\frac{\partial^2\phi(t,\omega)}{\partial t\,\partial\omega}. \quad (14)$$

Here, $X_{\mathcal{T}h}(t,\omega)$ is the short-time Fourier transform computed using a time-weighted anlaysis window $h_{\mathcal{T}}(t)=t\cdot h(t)$ and $X_{\mathcal{D}h}(t,\omega)$ is the short-time Fourier transform computed using a time-derivative analysis window $h_{\mathcal{D}}(t)=\frac{d}{dt}h(t)$.

The complicated expression that we obtain by substituting Eqs. (13) and (14) into Eq. (12) is useful because one of its terms contains the mixed partial derivative of the short-time Fourier transform phase, $\frac{\partial^2\phi(t,\omega)}{\partial t\partial\omega}$, which is the quantity of interest. A much simpler expression for the mixed partial derivative of the short-time Fourier transform with respect to time t and frequency ω is

$$\frac{\partial^2X(t,\omega)}{\partial t\,\partial\omega}=-jtX_{\mathcal{D}h}(t,\omega)+jX_{\mathcal{TD}h}(t,\omega), \quad (15)$$

where $X_{\mathcal{TD}h}(t,\omega)$ is the short-time Fourier transform computed using the analysis window

$$h_{\mathcal{TD}}(t)=t\frac{d}{dt}h(t)=th_{\mathcal{D}}(t). \quad (16)$$

Equating the right-hand sides of Eqs. (12) and (15), using the substitutions given by Eqs. (13) and (14), and taking advantage of the fact that the short-time Fourier transform phase is a real quantity, so its mixed partial derivative must also be real, we can obtain the simplified expression for the mixed partial derivative of short-time Fourier transform phase:

$$\frac{\partial^2\phi(t,\omega)}{\partial t\,\partial\omega}=\Re\left\{\frac{X_{\mathcal{TD}h}(t,\omega)}{X(t,\omega)}\right\}-\Re\left\{\frac{X_{\mathcal{T}h}(t,\omega)X_{\mathcal{D}h}(t,\omega)}{X(t,\omega)^2}\right\}. \quad (17)$$

Equation (17) gives a method of computing the mixed partial derivative of short-time Fourier transform phase at discrete time-frequency coordinates using only discrete short-time Fourier transform data, with no approximations of partial derivatives. The different initial definition of the short-time Fourier transform in Eq. (9) has the effect of shifting the relevant values of the mixed partial phase derivatives, so that line components and impulses are now characterized by derivatives near -1 and 0, respectively.

The algorithm invoking the above considerations is as follows:

(1) A time ramp and frequency ramp are constructed for the modified window functions, and these depend in detail on whether there is an odd or even number of data points in each frame. Accordingly, the following algorithm should be used to obtain the ramps and the special windows:

1: **if** mod(`win_size`,2) **then**
2: Mw=(`win_size`−1)/2
3: `framp`= [(0:Mw), (−Mw:−1)] (using Matlab colon notation for a sequence of numbers stepping by one over the specified range)
4: `tramp` = (−Mw:Mw)
5: **else**
6: Mw=`win_size`/2
7: `framp` = [(0.5:Mw−0.5), (−Mw+0.5:−0.5)]
8: `tramp` = (−Mw+0.5:Mw−0.5)
9: **end if**
10: `tramp` =`tramp/Fs`, where `Fs` is the sampling rate (in Hz) of the signal
11: `framp` = (`Fs`/`win_size`)×`framp`
12: `Wt`= `tramp` ×`window`
13: `Wdt`=−`imag(ifft(framp`×`fft(window))`; (`ifft` is the inverse transform function to `fft`)
14: `Wtdt` = `tramp` ×`Wdt`

2. One next builds four matrices of windowed signal frames of length `win_size` time samples. The matrix `S` has its frames windowed by the nominal function `window`. The matrices `S_time`, `S_deriv`, and `S_td` have their frames windowed, respectively, by `Wt`, `Wdt`, and `Wtdt`. The windowed signal frames overlap by a user-supplied number of points.

3. One next computes four corresponding short-time Fourier transform matrices in the customary manner described in the previous algorithm:

```
STFT =fft(S)
STFT_time=fft(S_time)
STFT_deriv=fft(S_deriv)
STFT_td=fft(S_td)
```

4. `STFTeps = max( STFT, eps );` this adds the minimum floating point precision value defined by Matlab to any zero values.

5.

$$\texttt{MixPD}=2\pi\left(\Re\left(\frac{\texttt{STFT\_td}}{\texttt{STFTeps}}\right)-\Re\left(\frac{\texttt{STFT\_deriv}}{\texttt{STFTeps}}\right)\times\Re\left(\frac{\texttt{STFT\_time}}{\texttt{STFTeps}}\right)+\Im\left(\frac{\texttt{STFT\_deriv}}{\texttt{STFTeps}}\right)\times\Im\left(\frac{\texttt{STFT\_time}}{\texttt{STFTeps}}\right)\right). \quad (18)$$

Once again the notation $A\times B$ for matrices A,B denotes a point-by-point product.

Comparing the computational cost of the two algorithms, the chief difference comes from the exact method's requiring four fast Fourier transforms of various windowings of the entire signal matrix, while Nelson's method requires only two. These computations can take non-negligible amounts of time for a large number of signal frames, while on the other hand the remaining calculations to be performed in either method are simple pointwise multiplications and additions that will not differentiate the methods noticeably. In practice, it has been found that by far the most time-consuming aspect of obtaining a reassigned spectrogram by either method is the process of displaying all the points, a

process that depends entirely on the implementational details of the visualization component of one's software.

The images in this paper have been produced using Matlab software, and the form of the algorithms somewhat reflects the architecture of this language. For example, it is probably unwise in any computer language to literally construct the large matrices in the algorithms to perform the pointwise multiplications and other operations, and Matlab indeed handles matrices more efficiently than is implied by the simple representation. Readers may wish to refer to the Matlab implementations of these routines which are linked on the first author's web page.

IV. DISCUSSION

A. Justifying the interpretation of the phase derivative

In regions where the CIF is not changing with frequency, the spectrum is dominated by a single component that is highly concentrated in frequency (i.e., a sinusoid). In these regions, all nearby spectral data are mapped to the frequency of the dominant sinusoid, so that the variation (partial derivative) with respect to frequency is near zero. Similarly, in regions in which all spectral data are mapped to the time of a dominant component that is highly concentrated in time (i.e., an impulse), the variation (partial derivative) of the reassigned time with respect to time is near zero. Since the reassigned time is computed by adding the LGD to the nominal time t,

$$0 \simeq \frac{\partial}{\partial t}\left[t + \mathrm{LGD}(t,\omega)\right] = 1 + \frac{\partial}{\partial t}\mathrm{LGD}(t,\omega) \tag{19}$$

so

$$1 \simeq -\frac{\partial}{\partial t}\mathrm{LGD}(t,\omega). \tag{20}$$

That is, as the nominal time t *increases*, the time correction (LGD) for data in the neighborhood of a dominant impulse *decreases* proportionally.

B. The related notion of *consensus*

Figure 9 demonstrates the effect of frequency reassignment for a fragment of voiced speech. The upper plot shows the conventional (dashed lines) and reassigned (crosses) magnitude spectra. The lower plot shows the mapping of nominal (Fourier transform bin) frequency to reassigned frequency for the same fragment of speech. Near the frequencies of strong harmonics, the mapping is flat, as all nearby transform data are reassigned to the frequency of the dominant harmonic component. This *consensus*, or clustering among reassigned frequency estimates in the vicinity of spectral peaks can be used as an indicator of the reliability of the time-frequency data.[4] If the reassigned frequencies for neighboring short-time Fourier transform channels are all very similar, then there is said to be a high degree of consensus and the quality of the frequency estimates is assumed to be good. In fact, consensus is exactly what is measured, in a very local way, by the mixed partial phase derivative. Moreover, an analogous argument can be made in the time dimension, wherein consensus would be a measure of impulsiveness, or more specifically a measure of the appropriateness of the analysis window when the signal is sparse in time.

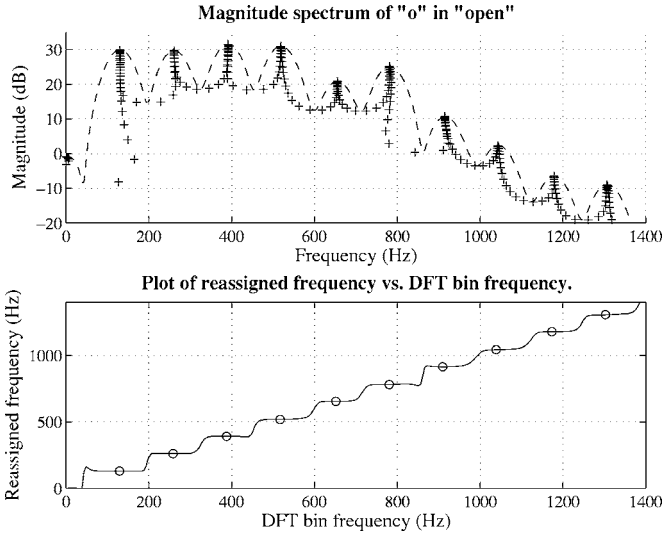


FIG. 9. Demonstration of frequency reassignment in a single spectrum for a fragment of speech (the "o" in "open"), computed using a 33.6 ms Kaiser analysis window with a shaping parameter of 12. The upper plot shows the conventional (dashed lines) and reassigned (crosses) magnitude spectra. The lower plot shows the mapping of nominal (Fourier transform bin) frequency to reassigned frequency for the same fragment of speech. The flat portions of the lower curve represent regions in which the energy in many transform bins is reassigned to the same frequency. The circled points show the samples having locally minimal frequency reassignments.

ACKNOWLEDGMENTS

Thanks to Doug Nelson for his encouragement, and for work worth following.

[1]D. J. Nelson, "Cross-spectral methods for processing speech," J. Acoust. Soc. Am. **110**, 2575–2592 (2001).

[2]S. A. Fulop and K. Fitz, "Algorithms for computing the time-corrected instantaneous frequency (reassigned) spectrogram, with applications," J. Acoust. Soc. Am. **119**, 360–371(2006).

[3]D. J. Nelson, "Instantaneous higher order phase derivatives," Digit. Signal Process. **12**, 416–428 (2002).

[4]T. J. Gardner and M. O. Magnasco, "Instantaneous frequency decomposition: An application to spectrally sparse sounds with fast frequency modulations," J. Acoust. Soc. Am. **117**, 2896–2903 (2005).

[5]K. Kodera, C. de Villedary, and R. Gendrin, "A new method for the numerical analysis of non-stationary signals," Phys. Earth Planet. Inter. **12**, 142–150 (1976).

[6]F. Auger and P. Flandrin, "Improving the readability of time-frequency and time-scale representations by the reassignment method," IEEE Trans. Signal Process. **43**, 1068–1089 (1995).

[7]H. A. Murthy, K. V. Madhu Murthy, and B. Yegnanarayana, "Formant extraction from phase using weighted group delay function," Electron. Lett. **25**, 1609–1611 (1989).

[8]A. V. Oppenheim and R. W. Shafer, *Digital Signal Processing* (Prentice-Hall, Englewood Cliffs, NJ, 1975).