This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSP.2014.2365756, IEEE Transactions on Signal Processing

An approach to time-frequency analysis with ridges of the continuous chirplet transform

Mikio Aoi*, Kyle Lepage, Yoonsoeb Lim, Uri Eden, and Timothy J. Gardner

Abstract

We propose a time-frequency representation based on the ridges of the continuous chirplet transform to identify both fast transients and components with well-defined instantaneous frequency in noisy data. At each chirplet modulation rate, every ridge corresponds to a territory in the time-frequency plane such that the territories form a partition of the time-frequency plane. For many signals containing sparse signal components, ridge length and ridge stability are maximized when the analysis kernel is adapted to the signal content. These properties provide opportunities for enhancing signal in noise and for elementary stream segregation by combining information across multiple analysis chirp rates.

I. Introduction

The continuous chirplet transform (CT) [1] is a time-frequency representation (TFR) that generalizes the Gabor transform to represent signals with linear frequency modulation [2], [3]. Since the quality of a TFR, reflected by such properties as energy concentration and the variance of derived parameter estimates, is dependent on how well-adapted the TFR is to the signal, the CT offers more flexibility for parsimonious representation than traditional TFR's with an unmodulated basis.

However, because many signals are formed by multiple components, or display nonlinear frequency modulation, methods that adapt a single CT to the whole signal may fail to represent each component

Manuscript received January 29, 2014. This work was supported in part by NSF grant NSF-DMS-1042134, the NSF Science of Learning Center CELEST (SBE-0354378) and by a Career Award at the Scientific Interface to TJG from the Burroughs Wellcome Fund.

*Correponding author. M. Aoi, K. Lepage, and E. Eden are with the Department of Mathematics & Statistics, Boston University (email: mcaoi@bu.edu, lepage@math.bu.edu, tzvi@bu.edu). Y. Lim is with the Center for Bionics, Biomedical Engineering Research Institute, Korea Institute of Science and Technology, Seoul, Korea (email: yslim@kist.re.kr), and T.J. Gardner is with the Department of Biology, Boston University, Boston, MA, 02215 USA (email: timothyg@bu.edu))

optimally. An improvement can be achieved by taking many transforms, and adapting the representation across these representations to each component separately by leveraging properties of the signal class of interest. One such property is phase continuity.

Many physical signals display continuous phase advancement that persists for several cycles (oscillations) or is coordinated across frequencies (fast transients). Such phase continuity is reflected in the phase of a complex TFR that is continuous in time and frequency. In spite of the ubiquity of phase-continuity in natural signals, the most common discrete-time analyses of frame-based TFRs treat each point in the time-frequency plane independently of every other point, ignoring *a priori* information regarding continuity of phase in time and frequency. In fact, many sparsity-enforcing algorithms impose penalties for the inclusion of correlated coefficients in a representation [4]–[7].

The purpose of this paper is to propose a strategy for leveraging phase continuity of signal components to develop an adaptive TFR in which each component in a signal is represented parsimoniously, relative to the chirplet basis family. The work builds on a contour representation of sound derived from the ridges of the Gabor transform [8], [9]. In the present study, we generalize the contour representation of sound to a derivation based on the ridges of the CT, and consider new signal representation issues that arise in this context.

Using the CT-based approach, bias in instantaneous frequency (IF) estimation of an amplitude-modulated component, with approximately linear frequency modulation, is reduced or eliminated when the analysis chirp rate matches the modulation rate of the signal. For components with nonlinear modulation, the bias reduction comes at the price of requiring the evaluation of the transform at various modulation rates. Thus, ridges that represent the same component, but were extracted using multiple transforms, form complimentary estimates that need to be combined to contribute all available information to a given estimate. This approach of combining estimates across transforms also reduces the need to search for a single analysis parameter (such as a "best" window width [8], [10]).

The organization of this paper is as follows: In Section II we will review the basic notation and concepts used in the rest of the paper, including development of the theory of chirplet ridges for both intrinsic-mode-type (IMT) [11] and click-like signals. In Sec. III we conduct an analytic case study of the ridges extracted from a deterministic, noise-free chirp and compare the properties of the proposed method with the previous method of [8], [9]. In Sec.IV we demonstrate properties of ridges that can be used to identify oscillatory signal components. We show, consistent with the notion of physical signals displaying smooth phase progression over extended regions of the time-frequency plane, that the length of a ridge is a useful metric for signal detection. We further show that the stability of ridge estimates with

respect to analysis parameters can be used to adaptively enhance the representation of signal components on a per-component basis. We illustrate these principles using both simulated data and sample data from zebra finch vocalizations. Finally, in Section V we discuss some properties of the chirplet ridge equations, provide several interpretations and perspectives on them, and highlight the relationship with other ridge extraction methods.

II. Notation and Background

The contour representation of sound developed in [8], [9] is a generalization of methods known more broadly as ridge and skeleton methods. Ridge and skeleton methods [12]–[14] are invertible TFRs which define ridges as curves in the time-frequency plane that represent the IF of a signal component. The restriction of signal resynthesis to these ridges is called the "skeleton". Ridge and skeleton methods have been used for estimation of component IF, phase, and amplitude in fields such as mechanical and civil engineering [15]–[17], analysis of chaotic attractors [18], and neuroscience [19], [20]. These methods are similar in spirit and application to high-resolution TFRs like time-frequency reassignment [21], [22], synchrosqueezing [23] and the Hilbert-Huang transform [24] in that they seek to identify, characterize, and decompose non-stationary, multicomponent signals by the IF of localized components.

Ridges, in the context of [8], [9], are used to explicitly link together associated points in the time-frequency plane resulting in an "object-based" TFR. In this method, objects are defined by the phase (alternatively, modulus [25]) of the Gabor transform. From this perspective, the fundamental objects of representation are not taken to be finer grain "atoms," of the discrete transform, but rather, the fundamental objects are the ridges themselves (although in practice, the continuous transforms are estimated at discretely sampled times and frequencies). While the method implies no constraints on the signal beyond ℓ_2 integrability, parsimonious component representations are possible for signals containing sparse, separable components, such as brief transients, or IMT [11] components with well-defined IF. While "separability" of components traditionally refers to some notion of discriminabilty in the *frequency* dimension, in the present context we generalize separability to refer to discriminability in *some* direction in the time-frequency plane.

In this section, we will define the ridges described in [8], [9] and develop the associated ridge theory for the CT.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSP.2014.2365756, IEEE Transactions on Signal Processing

A. Directed Gabor ridges

The contour representation of sound described in [8], [9] are based upon the ridges defined by directional derivatives of the Gabor transform phase. In this section we provide a formal definition of these ridges and define the territories in the time-frequency plane associated with these ridges.

The Gabor transform of any complex signal $x(t)$ is defined as

$$\begin{aligned}\chi_g(\tau,\omega) &= \langle x, g_{\tau,\sigma,\omega}\rangle,\\ &= |\chi_g(\tau,\omega)|e^{i\phi_g(\tau,\omega)},\end{aligned}$$

where $\langle x, g_{\tau,\sigma,\omega}\rangle = \int_{-\infty}^{\infty} x(t) g^{\dagger}_{\tau,\sigma,\omega}(t)dt$, and $g_{\tau,\sigma,\omega}(s) = (2\pi\sigma^2)^{-1/4}\exp[-(s-\tau)^2/2\sigma^2 + i\omega(s-\tau)]$ is the unit ℓ_2 norm Gabor kernel, and † denotes complex conjugation. The parameter σ determines the width of the analysis window in the time domain.

Commonly, the Gabor ridges are defined by

$$\frac{\partial\phi(\tau,\omega)}{\partial\tau} = \omega, \tag{1}$$

or, equivalently [25], [26] (see Appendix A),

$$\frac{\partial \log|\chi(\tau,\omega)|}{\partial\omega} = 0, \tag{2}$$

where $\chi(\tau,\omega)$ may represent either the Gabor or complex wavelet transforms, and $\phi(\tau,\omega)$ is the transform phase (for examples see [12], [13], [23], [27], [28]). We will show in Section III that ridges defined by (2) give biased estimates of the IF of frequency- and amplitude-modulated components. Furthermore, the ridges of (2) are not useful for the identification of fast transients.

In [8], the ridges (their "contours") of χ_g, with directionality θ are defined by a linear combination of the derivatives of $\phi_g(\tau,\omega)$. These linear combinations can be described as directional derivatives of $\phi_g(\tau,\omega)$. If we define the rotated coordinates $(u_\theta, v_\theta) = (\tau,\omega)\boldsymbol{R}^T$, where $\boldsymbol{R} = (\cos\theta, \sin\theta; -\sin\theta, \cos\theta)$ is the standard rotation matrix, then we can define the directional Gabor ridges in terms of points (τ,ω) that satisfy

$$\frac{\partial\phi_g(\tau,\omega)}{\partial u_\theta} = \omega\cos\theta, \tag{3}$$

or, equivalently[1]

$$Im\left\{\frac{\chi_{\mathcal{T}g}(\tau,\omega)}{\chi_g(\tau,\omega)}\left(\frac{1}{\sigma^2}\cos\theta + i\sin\theta\right)\right\} = 0, \tag{4}$$

[1]While the definition of the ridges in [8] is given by the points satisfying $Im\{\sigma^{-1}e^{i\theta}\chi_g(\tau,\omega)/\chi_{\mathcal{T}g}(\tau,\omega)\} = 0$, we modified the definition for the purposes of this paper to conform more closely with the notion of ridges as solutions to equations with directional derivatives of ϕ_g and for θ to correspond to the standard polar angle coordinate.

where $\chi_{\mathcal{T}g} = \langle x, \mathcal{T}g_{\tau,\sigma,\omega}\rangle$, and $\mathcal{T}g_{\tau,\sigma,\omega}(t') = (t'-\tau)g_{\tau,\sigma,\omega}(t')$. Note that $u_0 = \tau$ and $u_{\pi/2} = \omega$, defining the direction of the derivative as the direction with angle θ from the τ axis. Thus, (3) is a generalization of (1), where (1) is obtained from (3) by setting $\theta = 0$.

The directed Gabor ridges can be used to improve detection and estimation of both types of components [9].

1) Gabor ridge territories: Let $\mathcal{C}^*_\theta$ be the set of points which satisfy Eqn. (4), and $\mathcal{C}_\theta \subset \mathcal{C}^*_\theta$ be the subset of $\mathcal{C}^*_\theta$ corresponding to local maxima of $\log|\chi_g(\tau,\omega)|$ with respect to the direction θ. Suppose that, for a given $x(t)$ there are K ridges corresponding to maxima, with respect to direction θ. We can define $\mathcal{C}_{\theta,k}$ as the points on the kth ($k = 1, 2, \ldots, K$) ridge. Since ridges do not intersect [8] each $(u_{\theta,k}, v_{\theta,k}) \in \mathcal{C}_{\theta,k}$ can belong to only one ridge.

For each θ, k we can define a unique territory $\Omega_{\theta,k}$ containing the kth ridge;

Definition II.1. *$\Omega_{\theta,k}$ is the connected set such that $\forall(u_\theta, v_\theta) \in \Omega_{\theta,k}$, and $\forall(u^*_\theta, v^*_\theta) \in \mathcal{C}_{\theta,k}$, the following conditions are met :*

1) $v_\theta \in \mathcal{C}_{\theta,k}$
2) *If* $u_\theta > u^*_{\theta,k}$ *then* $\partial_\theta \log|\chi_g(\tau,\omega)| \leq 0$,
3) *If* $u_\theta < u^*_{\theta,k}$ *then* $\partial_\theta \log|\chi_g(\tau,\omega)| > 0$
4) $\mathcal{C}_{\theta,k} \subset \Omega_{\theta,k}$,

where $\partial_\theta = \partial/\partial u_\theta$.

The connectedness of $\Omega_{\theta,k}$ ensures that each $\Omega_{\theta,k}$ not composed of disjoint regions, and conditions 1)-3) thereby define the boundaries of $\Omega_{\theta,k}$. $\Omega_{\theta,k}$ only composes the entire time-frequency plane in the special case that there is only one ridge ($K = 1$).

Each $\Omega_{\theta,k}$ corresponds to a basin of attraction of the reassignment procedure [26], [29] when reassignment is constrained to displacement only along the u_θ axis. The set of all $\Omega_{\theta,k}$'s forms a partition of the time-frequency plane. Note that an equivalent definition of $\Omega_{\theta,k}$ can be formulated in terms of the phase of the Gabor transform as well (Appendix A). Each ridge in $\mathcal{C}_\theta$ corresponds to a phase-continuous region of $\chi_g(\tau,\omega)$ whose primary influence in the time-frequency domain is constrained to $\Omega_{\theta,k}$. A unique set of ridges exist for each θ and σ. The authors of [8] suggested that the ridges and territories which best represent a frequency-modulated (FM) signal are those for which the direction θ matches the local modulation rate of the signal, but this was not explicitly demonstrated. In Sec. III-C we formalize this assertion and show that this is approximately true for signals with linear frequency modulation, and is strictly true in the limit of a constant-amplitude signal.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSP.2014.2365756, IEEE Transactions on Signal Processing

The above does not imply that $\mathcal{C}_{\theta,k}$ and $\Omega_{\theta,k}$ describe a single component, or that they in fact represent signal components at all (as opposed to noise), only that they are well-defined with respect to a single value of σ and θ for each Gabor transform. In the following sections we develop machinery which can be used to distinguish ridges representing noise from ridges representing deterministic signal components when applied to the CT.

B. Ridges of the CT

The ridge representation defined in the previous section seeks to decompose a time series into time-frequency ridges $\mathcal{C}_{\theta,k}$ and territories $\Omega_{\theta,k}$ that are defined by the data, but the decomposition is influenced by the transform kernel. Here we generalize the approach of [8] to ridges derived from the CT.

The CT of $x(t)$ is defined as

$$\begin{aligned}\chi_c(\tau,\omega) =& \langle x, c_{\tau,\sigma,\omega,\beta}\rangle \\ =& |\chi_c(\tau,\omega)|e^{i\phi_c(\tau,\omega)},\end{aligned} \tag{5}$$

where$\langle x, c_{\tau,\sigma,\omega,\beta}\rangle = \int_{-\infty}^{\infty} x(t)c^{\dagger}_{\tau,\sigma,\omega,\beta}(t)dt$, and $c^{\dagger}_{\tau,\sigma,\omega,\beta}(t) = \left(2\pi\sigma^2\right)^{-1/4}\exp[-\frac{\alpha}{2}(t-\tau)^2 - i\omega(t-\tau)]$, with $\alpha = 1/\sigma^2 + i\beta$. By analogy with (4) and (2), the chirplet ridges will be defined as the points satisfying

$$\begin{aligned}\frac{\partial \log|\chi_c(\tau,\omega)|}{\partial u_\theta} &= \mathcal{R}e\left\{\frac{\chi_{\mathcal{T}c}(\tau,\omega)}{\chi_c(\tau,\omega)}\left(\alpha\cos\theta - i\sin\theta\right)\right\}, \\ &= 0,\end{aligned} \tag{6}$$

where $\chi_{\mathcal{T}c}(\tau,\omega) = \langle x, \mathcal{T}c_{\tau,\sigma,\omega,\beta}\rangle$ and $\mathcal{T}c_{\tau,\sigma,\omega}(t') = (t'-\tau)c_{\tau,\sigma,\omega,\beta}(t')$. We define the local maxima $\mathcal{C}_{\theta,k,\beta}$ and associated territories $\Omega_{\theta,k,\beta}$ of $\chi_c(\tau,\omega)$ similarly to the ridges and territories for χ_g in Sec. II-A. However, this representation now depends on three parameters, θ, σ, and β.

In what follows, we focus on ridges defined by $\theta = 0$ corresponding to the points (τ,ω) satisfying

$$\frac{\partial \log|\chi_c(\tau,\omega)|}{\partial \tau} = 0, \tag{7}$$

and $\theta = \pi/2$, corresponding to

$$\frac{\partial \log|\chi_c(\tau,\omega)|}{\partial \omega} = 0. \tag{8}$$

We will show in Sec. III that the $\theta = \pi/2$ ridges track the IF of brief chirp components when β matches the modulation rate of the component. In contrast, the ridges defined by $\theta = 0$ can resolve fast transients.

In principle, ridges derived from additional values of θ can be used to compute the stability of time-frequency estimates to variations in analysis, and this information used to further enhance a component-wise TFR [9]. This extension, however, is beyond the scope of the present work.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSP.2014.2365756, IEEE Transactions on Signal Processing

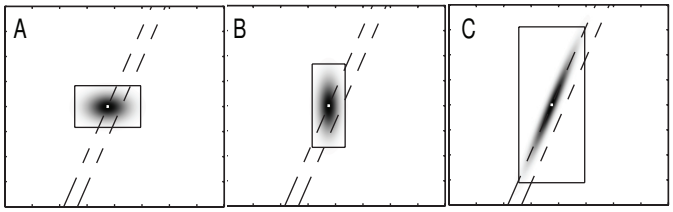


Fig. 1. **Time-frequency resolution enhanced by variation in window orientation.** A comparison between the time-frequency uncertainty ellipses for the Gabor transform for high frequency resolution, but low time resolution (A), low frequency resolution, but high time resolution (B), and with the CT with non-zero modulation rate (C). Time-bandwidth product is indicated by the area of the rectangle bordering the ellipse.

C. Advantage of the CT

The CT permits resolution of closely spaced, FM components which is not possible using monotone basis elements [30]–[32]. To illustrate this point, consider two FM components with IF's that are oriented in parallel in the time-frequency plane. If the components are closely spaced in either time (Fig. 1A) (or frequency (Fig. 1B)) then a large (small) analysis window width parameter σ will make the analyzing Gabor kernel cover both chirps and the two chirps will not be resolved . For components closely spaced in both dimensions there may not exist any scale for which the two chirps can be resolved using the Gabor transform.

In the case of the CT, the orientation of the ellipse is determined by the chirp rate β of the analyzing chirplet. If the ellipse is sufficiently rotated (Fig. 1C), and the scale permits a sufficiently narrow ellipse, then each chirp can be resolved without significant interference. In practice, data that display multiple components, or nonlinear frequency-modulation, will not be well characterized by a single value of β and σ, requiring us to use multiple values of (β, σ).

III. Ridge analysis of chirps

In this section we use the ridges defined by (7) and (8) to conduct an analysis of IMT [11] and asymptotic components, for which the IF is well-defined and the amplitude varies slowly compared to the phase, and click-like non-IMT components, for which the amplitude varies quickly compared to the phase. To this end, we will consider a generalized mono-component signal

$$x(t) = Ae^{-(t-t_0)^2/2s^2 + i\phi(t)}.$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSP.2014.2365756, IEEE Transactions on Signal Processing

To simplify our analysis we approximate this component with a second-order Taylor expansion of the phase about t_0,

$$x(t) \approx A \exp[-(t-t_0)^2/2s^2 + i(b(t-t_0)^2/2 + \omega_0(t-t_0) + \phi_0)], \tag{9}$$

where $\phi_0 = \phi(t_0)$, $\omega_0 = \phi'(t_0)$, and $b = \phi''(t_0)$. The purpose of this exercise is to demonstrate the usefulness of the CT ridges as IF estimators when the signal component has approximately linear FM over a time scale comparable to the analysis window width, σ. Here then, s represents the time scale on which the frequency modulation of the component is approximately linear and the IF is given by

$$\omega(t) = b(t-t_0) + \omega_0. \tag{10}$$

We will use the intuition we gain here in Sec. IV to develop heuristics based on ridge properties which will be useful for methods of component detection. We will then compare the results found here with analogous results for the directed Gabor ridges.

Punctate events such as pulses and clicks can be modeled as special cases of (9) where the signal duration is very short relative to the analysis window width. When $A = A_0(2\pi s^2)^{-1/2}$, where A_0 is any real, positive constant, this component has the property that, in the limit as $s \to 0$, $x(t) \to \delta(t-t_0)$, where $\delta(t-t_0)$ is the Dirac delta function centered at t_0. This property allows us to study the effects of discontinuities and fast transients in the signal by examining the behavior of the ridges in the passage to this limit.

A. Chirplet ridges of oscillatory components

We can identify the $\partial_\omega \log|\chi_c| = 0$ ridges of the CT of (9) by the formula given in (6) with $\theta = \pi/2$. The resulting ridge satisfies

$$\frac{\partial \log|\chi_c(\tau,\omega)|}{\partial \omega} = c_1\left[c_2(\omega_0 - \omega) + c_3(\tau - t_0)\right] = 0, \tag{11}$$

where $c_1 = [(\beta-b)^2 + (1/\sigma^2 + 1/s^2)^2]^{-1}$, $c_2 = 1/s^2 + 1/\sigma^2$, and $c_3 = b/\sigma^2 + \beta/s^2$. For $s, \sigma > 0$, the $\partial_\omega \log|\chi_c(\tau,\omega)| = 0$ points correspond to

$$\omega = \left(\frac{\beta\sigma^2 + bs^2}{\sigma^2 + s^2}\right)(\tau - t_0) + \omega_0, \tag{12}$$

which is a line in the time-frequency plane whose slope is a weighted average of b and β. When the analysis chirp rate and the signal chirp rate match (i.e. $\beta = b$) this line reduces to

$$\omega = b(\tau - t_0) + \omega_0,$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSP.2014.2365756, IEEE Transactions on Signal Processing

which reproduces (10), the IF of the signal, independent of the signal width s or the analysis window width σ. This independence is advantageous as it allows us to identify the IF of the signal in a regime where the estimate is insensitive to s and σ.

Noting that $c_{\tau,\sigma,\omega,\beta}(t)$ reduces to $g_{\tau,\sigma,\omega}(t)$ when $\beta = 0$, we find that the $\partial_\omega \log|\chi_g(\tau,\omega)| = 0$ ridge can be expressed as

$$\omega = \left(\frac{b}{\sigma^2/s^2 + 1}\right)(\tau - t_0) + \omega_0.$$

Therefore, the ridges most commonly used for IF estimation by the ridge technique only recover the IF (10) in the limit as $\sigma/s \to 0$; i.e. as the amplitude of the linear IF regime becomes constant with respect to the analysis window. This may not be possible, even approximately, when there are multiple FM components which are closely spaced in frequency (ex. modulated harmonics with slow fundamental frequency), where making σ too small will result in interference between components. Thus, for most reasonable values of σ, the local maximum of $\log|\chi_g(\tau,\omega)|$ gives a biased estimate of the IF of components with both frequency- and amplitude-modulation.

B. Click-like components

In practical data analysis conditions, due to limits on sampling rate, there will be a lower limit on σ (σ_{min}) such that if $s \ll \sigma_{\text{min}}$ then (12) will be dominated by

$$\omega \approx \beta(\tau - t_0) + \omega_0.$$

Thus, as $s/\sigma \to 0$ and the signal (9) becomes more click-like, the $\partial_\omega \log|\chi_c(\tau,\omega)| = 0$ ridge becomes a function of only the analysis chirp rate β, and not the signal chirp rate.

Under these conditions, the amplitude modulation of the signal is much faster than the phase progression, resulting in a signal component that more closely resembles a fast transient than a well-defined oscillation. We would like to be able to characterize these components with a set of ridges that are not as sensitive to the analysis parameters, analogous to the analysis demonstrated for asymptotic components in Sec. III-A.

To this end, we consider the $\partial_\tau \log|\chi_c| = 0$ ridge of $x(t)$, where, using Eq. (6) with $\theta = 0$

$$\frac{\partial \log|\chi_c(\tau,\omega)|}{\partial\tau} = c_1\left[c_3(\omega - \omega_0) + c_4(t_0 - \tau)\right] = 0 \tag{13}$$

where c_1 and c_3 are defined as above and $c_4 = \beta^2/s^2 + b^2/\sigma^2 + 1/s^4\sigma^2 + 1/s^2\sigma^4$. The points that satisfy $\partial_\tau \log|\chi_c(\tau,\omega)| = 0$, obey the relation

$$\tau = \frac{(bs^2 + \beta\sigma^2)(\omega - \omega_0)}{\beta^2\sigma^2 + b^2s^2 + 1/s^2 + 1/\sigma^2} + t_0. \tag{14}$$

In the limit as $s \to 0$, Eqn. (14) becomes

$$\tau = t_0,$$

giving a $\partial_\tau \log|\chi_c| = 0$ ridge at all ω, independent of the analysis parameters σ and β. Thus, signals that are sparse in time are well localized by the $\partial_\tau \log|\chi_c| = 0$ ridge and the representation is less sensitive to the analysis parameters than the $\partial_\omega \log|\chi_c| = 0$ ridges.

C. Chirp analysis by directed Gabor ridges

We can assess the value of the chirplet ridge approach with respect to the prior approach of [8] by examining the signal (9) with the ridges defined in (4). Equation (4) can be written as

$$\cos\theta(\partial_\tau \phi_c(\tau,\omega) - \omega) + \sin\theta \partial_\omega \phi_c(\tau,\omega) = 0.$$

Using the phase - magnitude relations provided in Appendix A, and expressions (11) and (13), we find that the phase derivatives are given by

$$\partial_\tau \phi_c(\tau,\omega) - \omega = c_1 \left(c_2'(\omega_0 - \omega) + c_3'(\tau - t_0) \right),$$

and

$$\partial_\omega \phi_c(\tau,\omega) = c_1 \left(c_4'(\omega_0 - \omega) + c_5(\tau - t_0) \right),$$

where c_1 is the same as in Sec. III-A and $c_2' = \beta(\beta - b) + 1/s^2\sigma^2 + 1/\sigma^4$, $c_3' = b/\sigma^4$, $c_4' = b$, and $c_5 = b(b-\beta) + 1/s^2\sigma^2 + 1/s^4$. We can identify the ridges of χ_g of the deterministic chirp, in direction angle θ by setting $\beta = 0$. The resulting ridge representation of the signal in (9) is given by points (τ,ω) satisfying

$$\omega - \omega_0 = \hat{b}_\theta(\tau - t_0),$$

where

$$\hat{b}_\theta = \left(\frac{(b/\sigma^4)\cos\theta + (b^2 + 1/s^2\sigma^2 + 1/s^4)\sin\theta}{(1/s^2\sigma^2 + 1/\sigma^4)\cos\theta + b\sin\theta} \right). \tag{15}$$

In the limit as $s \to \infty$ (locally constant amplitude), we have $\hat{b}_\theta \to b$, giving a ridge identical to (10), the IF of the component, for all σ and θ. Similarly, as $s \to 0$ (click-like), $1/\hat{b}_\theta \to 0$ for all b, σ, θ (except for $\theta = \pi/2$, in which case, $\hat{b}_\theta \to 0$), giving a ridge defined by $\tau = t_0$, which identifies the location in time of the click. For $0 < s < \infty$, the ridge will match the IF of the chirp ($\hat{b}_\theta = b$) when

$$\tan\theta = \frac{b}{1 + \sigma^2/s^2}. \tag{16}$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSP.2014.2365756, IEEE Transactions on Signal Processing

Thus, in the limit as $\sigma/s \to 0$, the directed Gabor ridge estimate of the IF of the linear FM signal is unbiased when $\tan\theta = b$, as suggested in [8].

These results show that, as long as the component is sufficiently isolated (with respect to the uncertainty window), the directed Gabor ridges can identify the IF of the component or the location of transients. However, even for this simple component, in order to correctly recover the IF of the component, both θ and σ must be chosen to satisfy (16), except in cases where the signal either has approximately constant amplitude (s/σ large), or where this signal is click-like (s/σ small). In contrast, the linear IF can be recovered by the chirplet ridges while only having to satisfy $\beta = b$.

IV. Signal enhancement by ridge properties

In this section we demonstrate the application of the chirplet ridges to data and how their properties can be used to enhance signal-to-noise ratio (SNR) for click-like and IMT components by restricting the analysis to relatively long ridges. First, some heuristics for discriminating between signal ridges and noise ridges will be described. We then show that phase-continuous components can be discriminated using the methods described even when their IFs cross in the time-frequency plane.

A. Signal amplification by ridge length

Sparse signals such as clicks or IMT components can introduce long-range dependencies in the time-frequency plane in the phase of $\chi_g(\tau,\omega)$. These dependencies correspond to standard notions of the phase evolution of physical signals, and are therefore displayed as persistent phase continuity in the time-frequency plane. Since white noise has random phase, this persistent phase continuity of the signal may be exhibited as ridges that are longer than the typical ridge length for white noise-only data. Thus, we hypothesized that ridge length may be an effective metric for distinguishing between ridges formed by noise, and those formed by the presence of signal components. For simplicity, we defined ridge length simply as the number of pixels belonging to a ridge and the ridge is defined as in (8)[2].

We examined the distribution of ridge lengths in the presence and absence of signal by simulating data with a single linear, Gaussian modulated chirp as described by (9) in Gaussian white noise and extracted the ridges at varying analysis modulation rates β. Repeating the analysis for 1000 trials, we compared the estimated distribution of resulting ridge lengths between data with signal+noise and data

[2]Because the phase of $\chi_c(\tau,\omega)$ is smooth with respect to both τ and ω everywhere except at its zeros [33], [34] we reasoned that ridges never branch, since branching would be associated with a discontinuity in the phase of χ_c. Thus, the "length" of a ridge is unambiguous.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSP.2014.2365756, IEEE Transactions on Signal Processing

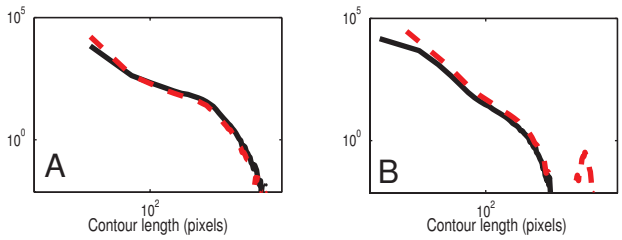


Fig. 2. **Histograms of ridge length at two analysis chirp rates** Simulated chirps in noise (- -) compared to noise only (-). The data were 50 sec long with sampling rate $F_s = 100$ Hz. The signal was a linear chirp as in (9) with $b = 1$ Hz/sec, $s = 8$ sec, $f_0 = 25$ Hz, $t_0 = 25$ sec, and $A = 2$. Noise was Gaussian white noise with zero mean and standard deviation = 1.

containing only white noise across varying analysis parameters. Fig. 2 shows the histograms of ridge length for $\beta = -1$ Hz/sec (Fig. 2A) and $\beta = b = 1$ Hz/sec (Fig. 2B) for a single chirp in noise (- -), and for noise with no signal (–). The parameters of the simulations are provided in the caption of Fig. 2. At $\beta = -1$ Hz/sec (Fig. 2A) the distribution of lengths from data containing signal does not appear to differ from the distribution of lengths from data with noise only. However, at $\beta = 1$ Hz/sec (Fig. 2B) the distribution shows an additional mode at the right tail for the signal-plus-noise condition. Noting that $\beta = b$ in this case, Fig. 2B suggests the presence of consistently longer ridges when the analysis and signal chirp rates match[3]. This result suggests that by rejecting all ridges but those in an upper quantile of length we can restrict our analysis to only those ridges most likely to correspond to signal *and* that represent high-quality IF estimates. We will demonstrate the usefulness of ridge length for improving SNR in Sec. IV-B, IV-C, and IV-E by using the top 1% of ridges by length.[4]

B. Signal enhancement by ridge density

In Sec. III we showed that, as σ/s decreases, the ridge estimate of the IF will have decreasing bias and be less sensitive to the value of β. Thus, even for nonlinear modulation, if we can set $\sigma/s \ll 1$ (provided this is possible with respect to the sampling constraints discussed in Sec. III-B) then the ridge estimate of the IF will be close to the true IF in a neighborhood of points where $\beta = \phi''(t)$. If we obtain multiple

[3]Simulations were conducted for various values of σ, however, those results were qualitatively similar to the results presented here.

[4]While this threshold was somewhat arbitrary, it is supported by the analysis in Fig. 2 and the quality of the resulting representations across the examples that we present here. However, the optimal threshold is likely to be dependent on the data SNR and the amplitude modulation of the signal. Future studies should include a more systematic study of how the ridge length distribution is influenced by the data properties.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSP.2014.2365756, IEEE Transactions on Signal Processing

ridge estimates (with varying analysis parameters) and sum them, then insensitivity to analysis parameters would result in a high density of ridges in regions of the time-frequency plan containing signal. In this section, we develop this argument and identify conditions in which such a procedure would be successful in discriminating between signal ridges and noise ridges.

Consider the variation in the IF with respect to β. Assuming $\phi'''(t) \approx 0$ in the neighborhood of the analysis so that the modulation is approximately linear and $\omega(\tau)$ is determined from Eq. (12), the variation in $\omega(\tau)$ due to a small variation in β, ($\delta\beta$) is

$$\delta\omega(\tau) = \frac{\delta\beta(\tau - t_0)}{1 + s^2/\sigma^2}.$$

If we let Δt be the resolution of the time discretization and $\Delta\beta$ be a finite change in β then, k ($k = 1, 2, \dots$) time steps from t_0 we have the frequency distance equal to

$$\delta\omega(k\Delta t) = \frac{k\Delta\beta\Delta t}{1 + s^2/\sigma^2}.$$

Thus, for the discretized CT, a small change in β ($\Delta\beta$) will not change the coefficients that a ridge occupies, k time steps from t_0 so long as $\Delta\beta$ satisfies

$$\Delta\beta < \Delta\beta_{\text{max}} \equiv (1 + s^2/\sigma^2)\frac{\Delta\omega}{k\Delta t}, \tag{17}$$

where $\Delta\omega$ refers to the discretization in frequency. Thus, when this condition is met the representation of the ridges will overlap for $2k - 1$ time steps.

Heuristically, if we note that white noise has random phase, then the time scale of phase continuity for white noise is essentially zero ($s^2/\sigma^2 \approx 0$). It stands to reason therefore, that white noise would also not have a directionality preference in the time-frequency plane. Thus, we can approximate the white noise-only agreement condition as[5],

$$\Delta\beta < \Delta\beta^0_{\text{max}} \equiv \frac{\Delta\omega}{k\Delta t}. \tag{18}$$

Since we must have $s/\sigma > 0$, we always have $\Delta\beta_{\text{max}} > \Delta\beta^0_{\text{max}}$. Thus, ridges formed by signal components are less sensitive to variations in β than ridges due to noise.

Similarly, we can determine the sensitivity of ridges to variations in σ^2. The variation in $\omega(\tau)$ due to a small change in σ^2 ($\Delta\sigma$) is given by

$$\delta\omega(k) = \frac{(\beta - b)\Delta\sigma}{s^4(\sigma^2/s^2 + 1)^2}k\Delta t.$$

[5]Condition (18) corresponds to our experience with simulated signals in white noise and is corroborated by the noise ridges shown in Fig. 3.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSP.2014.2365756, IEEE Transactions on Signal Processing

TABLE I

VALUES OF COMPONENT PARAMETERS

Component	s	t_0	t_1	γ	b	f_0
y_1	3	3.5	4	1	4	12
y_2	3	3	4.5	1	4	10

Therefore, for $\beta \approx b$, there should be little to no difference in ridge position at any k due to a small change in σ.

Thus, insensitivity of the ridge to β and σ indicates the presence of signal. This is similar in principle to the "stability" of contours described in [9]. The above results suggest that signal components can be identified by building a histogram in the time-frequency plane of the ridges derived from different values of β and σ. If we estimate ridges formed over a range of β's such that $\Delta\beta^0_{\text{max}} < \Delta\beta < \Delta\beta_{\text{max}}$, we are assured that ridges due to signal components will tend to overlap in the time-frequency plane more than ridges due to noise alone. A high ridge density associated with variations in β therefore indicates regions which contain signal components, and including variations in σ can further amplify this result when $\beta \approx b$. In practice, with a course sampling of the frequency axis and large s/σ, condition (17) turns out to be easy to satisfy. Points on the ridges of (7) for click-like components can be similarly grouped, but the expression for this grouping is more complicated and less intuitive. We use ridge density in Sec. IV-C and IV-E to highlight salient features of data with nonlinear frequency modulation.

C. Amplification of nonlinear signal components

In this section we demonstrate signal amplification using the methods described in Sections IV-A, IV-B of two, closely-spaced signal components with nonlinear frequency modulation in noise. We simulated noisy data with two signal components

$$d(t) = y_1(t) + y_2(t) + n(t),$$

where $y_1(t)$ and $y_2(t)$ are cubic-phase components of the form

$$y(t) = \sqrt{\frac{2}{s^2}} e^{\pi(t-t_0)^2/s^2 + 2\pi i(\gamma/3(t-t_1)^3 + b/2(t-t_1)^2 + f_0(t-t_1))},$$

and $n(t)$ is Gaussian white noise with $\text{Var}[n(t)] = 0.03$, giving a peak SNR = 15 dB. Parameter values for the two components are given in Table I.

Chirplet transform magnitudes and ridges are shown in Fig. 3 for two chirp rates. The quadratic evolution of the IF of the components makes the linear modulation of the analysis window a poor fit for the signal at any one value of β. The Gabor transform magnitude ($\beta = 0$ Hz/sec, $\sigma = 3/5$ sec), shown in Fig. 3A reveals that the close spacing of the signal components and the nonlinear modulation combine in both the time and frequency axes such that we observe significant cross terms in all directions, indicating that there is no single pair of parameters (σ, β) that will prevent cross terms from interfering with the signal representation at some point in the TF plane. For example, at a low analysis chirp rate (Fig. 3A,B, $\beta = 0$ Hz/sec) ridges break apart in response to cross terms near the regions of the TF plane where the signals have fast modulation. Conversely, for the relatively high analysis chirp rate (Fig. 3C,D, $\beta = 4$ Hz/sec) cross terms effect ridges at the regions with slow modulation. Similar effects would be present even for a mono-component signal due to the bias in the ridge estimates described in Sec. III. Thus, there is no single best σ, or β that will allow us to characterize both signals simultaneously.

We can compare the ridges for the CT in Fig. 3 with another multi-scale method, the ridges of the Morlet wavelet transform for the same data, displayed in Fig. 4D. Note that the ridges of the Morlet wavelets transform lack the continuity and smoothness of the chirplet transform. However, the central frequency of the signal, the rate of change in modulation, and the scaling factor of the wavelet scales all influence the quality of this representation and better or worse representations are achievable.

The territories belonging to each ridge are outlined for each component in Fig. 3B,D and show a similar failure to perfectly segregate any one component. We defined these territories as in Definition II.1 where here, $\theta = \pi/2$.

As predicted in Sec. IV-A the ridges corresponding to signal components tend to be longer than the ridges due to noise alone. Furthermore, as predicted in Sec. IV-B the component ridges are strongly biased towards the IF of the corresponding components (shown in blue and green), while the ridges corresponding to variations due to noise alone are well-aligned with the analysis chirp rate (Fig. 3B,D).

A composite image of the longest 1% of ridges determined from nine chirp rates (uniformly sampled from 0–4 Hz/sec) and weighted by their modulus, enhances ridges corresponding to high-energy signal components with good agreement between ridges formed at different modulation rates (Fig. 4A). This stage of signal enhancement is important as it restricts the detection procedure to the points in the time-frequency plane most likely to correspond to the IF of a signal component. If we were to simply take the average modulus of the data over all transforms and take only the largest 1% of coefficients, we obtain the coefficients shown in Fig. 4B. In contrast, when we take largest 1% of coefficients from the representation in Fig. 4A, the representation is reduced to the coefficients shown in Fig. 4C, highlighting

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSP.2014.2365756, IEEE Transactions on Signal Processing

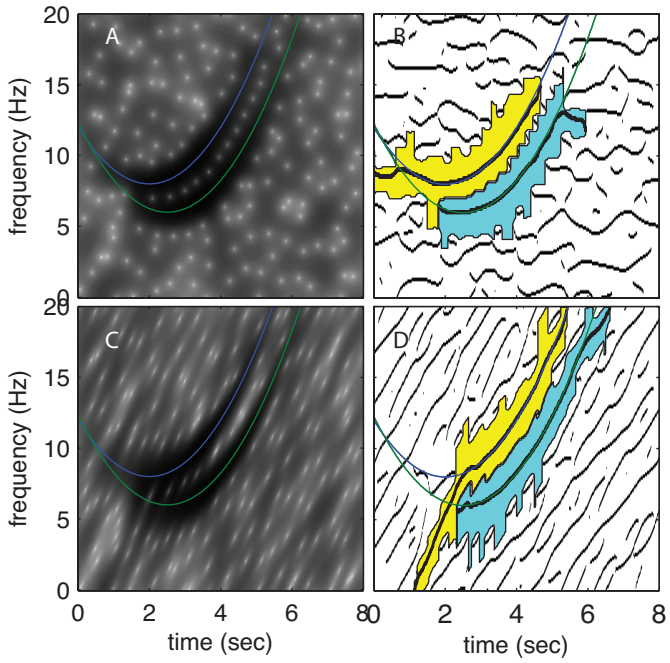


Fig. 3. **Sample ridges and territories for simulated data at two chirp rates.** $\sigma = 3/5$ sec. The exact IFs of the signal components are shown in blue and green. A,B: β = 0 Hz/sec (Gabor). C,D: $\beta = 4$ Hz/sec. A,C: $\log|\chi_c|$. B,D: $\partial_\omega \log|\chi_c| = 0$ ridges. Shaded areas are ridge territories.

the IF of the components with greater precision than thresholding the modulus alone.

D. Segregation of crossed components

The process of choosing the longest and most stable ridges across a family of transforms can lead to a simple form of stream segregation for components with IF's that cross in the time-frequency plane in which representations displaying phase continuity are favored.

To demonstrate this capability, we simulated three pairs of components, one pair of sinusoids, one pair of clicks, and one pair of linear frequency sweeps, with a background of white noise. Simulation details are given in the caption of the Fig. 5. Based on our analyses in Sections IV-B and IV-A, we used ridge length to discriminate between noise ridges and component ridges.

Fig.'s 5 and 6 show histograms of the longest 1% of ridges from the $\partial_\omega \log|\chi_c(\tau,\omega)| = 0$ and $\partial_\tau \log|\chi_c(\tau,\omega)| = 0$ ridges, respectively. The ridges were generated using a range of analysis parameter values and the colors reveal which analysis parameters contributed to each pixel in the image. In Fig. 5A, each color corresponds to a specific value of σ, and the intensity of each color in a pixel indicates

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSP.2014.2365756, IEEE Transactions on Signal Processing

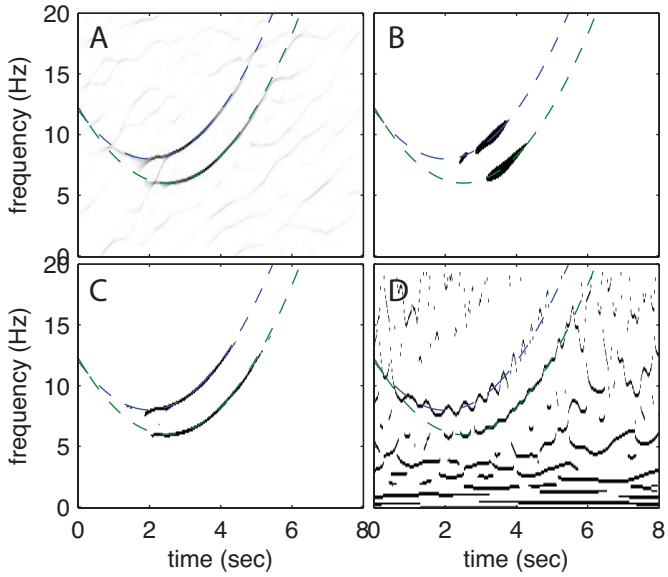


Fig. 4. **Signal is amplified when the representation in restricted to the long, stable ridges.** A) Ridge density for the longest 1% of ridges from all chirp rates, weighted by transform modulus. Darker regions indicate higher cumulative weighted-ridge density. B) Coefficients with cumulative modulus (over all transforms) in the highest 1%. C) Coefficients from A with highest 1% cumulative modulus. D) Ridges of the Morlet wavelet transform.

the density of ridges over all values of β. Fig. 5B shows the same histogram but with color indicating the value of β and intensity indicating the pixel density with respect to σ. The variations in color reflect that this composite image is built from multiple transforms that contribute differentially to the representation of each component. If a component is represented in all angles or all time-scales equally, it will appear white. If a component appears in a specific color range, then a narrow range of analysis parameters contributed to the representation of that component.

Fig. 5 indicates that variations over β of the $\partial_\omega \log|\chi_c| = 0$ ridges help to discriminate between components which have different modulation rates. In particular, chirplet modulation rates close to the modulation of a component, will uniquely identify that component. We can see this in Fig. 5B, where the chirps are dominated by ridges with analysis chirp rate $\beta = 6.25$kHz/sec, while the tones are dominated by ridges formed by relatively low modulation rates (β between 1.87-.62 kHz/sec). Fig. 5C,D show that there exist specific parameter combinations that bias the ridge estimates in favor of each set of components without substantial interference from the other signal components.

In agreement with our analysis in Sec. III, Figs. 5A,B show that the $\partial_\omega \log|\chi_c| = 0$ ridges are insensitive to the presence of the click components with the exception of interference terms crossing the

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSP.2014.2365756, IEEE Transactions on Signal Processing

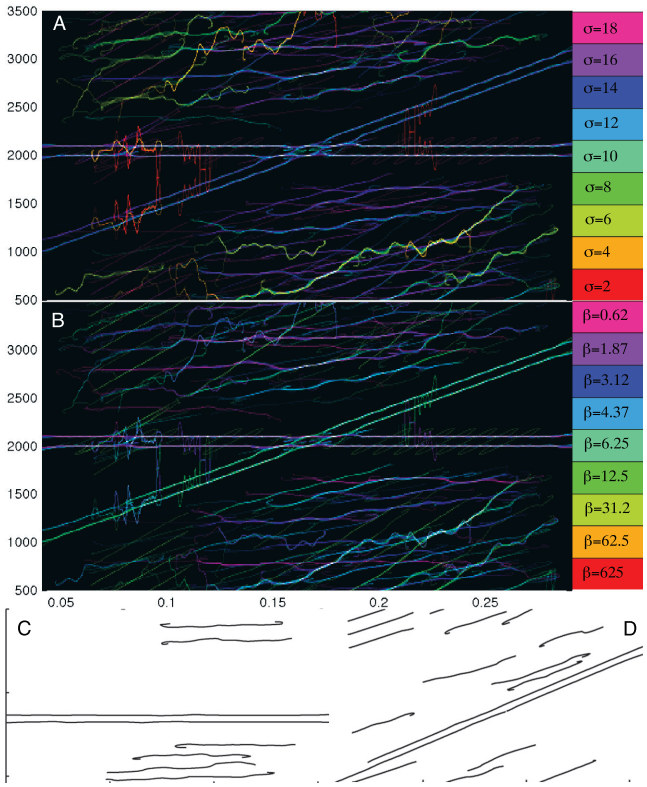


Fig. 5. **Stream segregation of crossed signal components.** Histogram of the 99% longest $\partial_\omega \log|\chi_c| = 0$ ridges for a multi-component signal. Colors indicate ridges estimated from different values of the analysis parameters. A) Colors indicate different values of the analysis window width σ, with color intensity indicating overlap over different values of the modulation rate β. B) The same data as A, but the color indicates the chirp rates β that contributed to each pixel. C) Ridges formed from one set of analysis parameters which identifies the tones. D) Ridges formed from one set of analysis parameters which identify the chirps. y-axes are in Hertz, x-axes are in seconds.The data were simulated with a sampling rate of 25kHz. Chirps and tones had an amplitude of 1 in arbitrary units. Chirp and click pairs were spaced by 0.1kHz. Chirps had modulation rate of 8kHz/sec . Clicks were one sample in duration with amplitude of 40 and are spaced 5 ms apart. The noise was distributed $n(t) \sim U(-5 \times 10^{-4}, 5 \times 10^{-4})$.

chirps and tones for short window widths and moderate modulation rates (at $\approx$ .08 sec). Fig. 6 shows the $\partial_\tau \log|\chi_c| = 0$ ridges of the same signal, with colors and intensities corresponding to those in Fig. 5. Fig. 6A,B shows clear discrimination of the clicks by the $\partial_\tau \log|\chi_c| = 0$ ridges by varying over both σ and β. The best click detection occurs at small σ, and low modulation rates (Fig. 6C). At large β and σ, the $\partial_\tau \log|\chi_c| = 0$ ridges do not identify any signal components (Fig. 6D).

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSP.2014.2365756, IEEE Transactions on Signal Processing

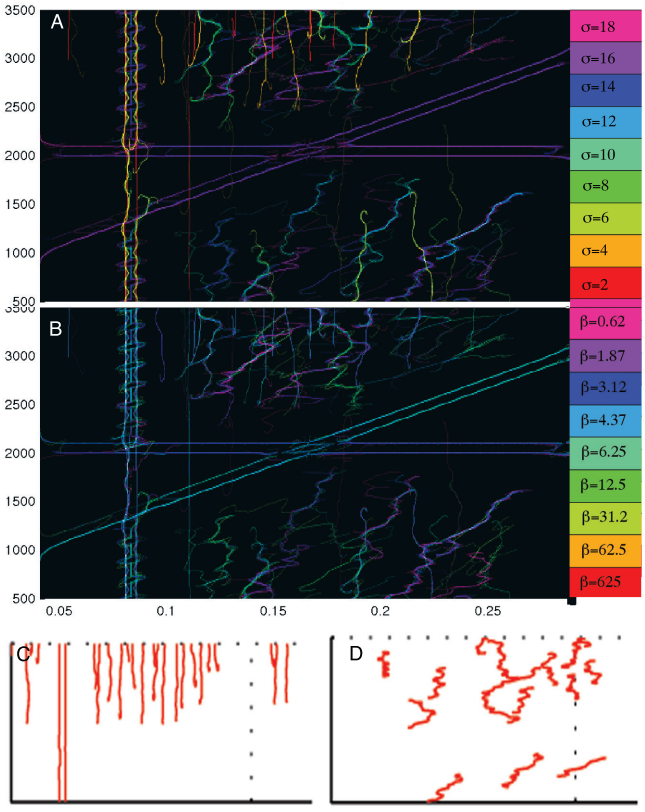


Fig. 6. **Histogram of the 99% longest $\partial_\tau \log|\chi| = 0$ ridges for a multi-component signal.** Units and the meaning of the colors in A and B are the same as those in Fig. 5A,B. C) Ridges formed from one set of analysis parameters which identified the clicks. D) Ridges formed from one set of analysis parameters which did not identify any of the signal components.

E. Analysis of bird song data

Time-frequency analysis is common practice among scientists and engineers studying acoustic data. Zebra finch vocalizations, for example, often include sequences of modulated harmonics and repeated clicks. An example of $\log|\chi_g|$ of a zebra finch vocalization is shown in Fig. 7A ($\sigma = 4.7$ ms).

Both $\partial_\tau \log|\chi_c| = 0$ ridges and $\partial_\omega \log|\chi_c| = 0$ ridges were generated for ten chirp rates (uniformly sampled between -0.45 kHz/ms and 0.29 kHz/ms) and four window widths (uniformly sampled between 2-4.7 ms). Fig. 7B shows the superposition of all ridges generated by the analysis. Salient signal features are over-represented across analysis parameters but, as predicted in Sec IV-A, the most distinct signal features are easily identified among the longest 1% of ridges shown in Fig. 7C,D. Fig.'s 7C and 7D show the longest 1% of the $\partial_\omega \log|\chi_c| = 0$ and $\partial_\tau \log|\chi_c| = 0$ ridges, respectively. The sum of Fig. 7C and

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSP.2014.2365756, IEEE Transactions on Signal Processing

7D is shown in Fig. 7E. The pixel intensities are weighted only by the ridge density and not by the CT modulus. Thus, points with high ridge density are brighter while regions with no ridges are dark.

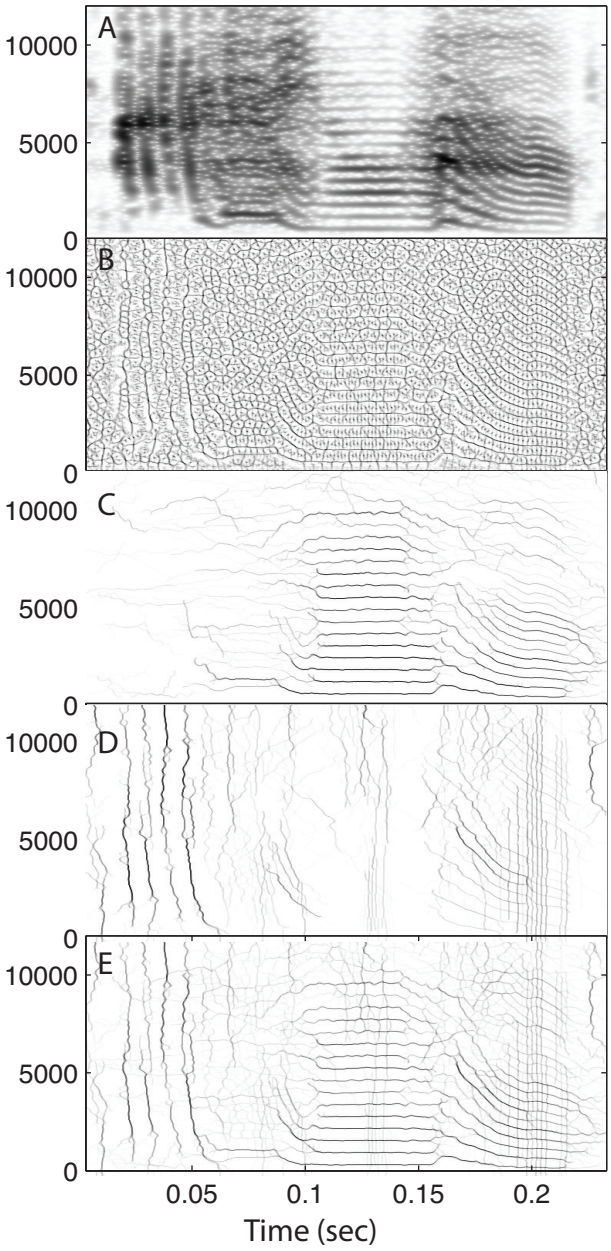


Fig. 7. Chirplet ridge analysis of a zerbrafinch vocalization. A: Log modulus of Gabor spectrogram of song at one time scale ($\sigma = 4.7$ ms). Dark colors indicate higher intensities. (The contrast of this image has be modified in order to improve visibility of details at higher frequencies). B: Ridges averaged over ten chirp rates and four window widths. C: Average of longest 1% of $\partial_f \log|\chi_c| = 0$ ridges, D: Average of longest 1% $\partial_\tau \log|\chi_c| = 0$ ridges. E: Sum of C and D. Y-axis scales are in kHz. X-axis is in ms.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSP.2014.2365756, IEEE Transactions on Signal Processing

Signal features with less distinct time-frequency localization are more poorly represented by the ridges. For example, the ridge representation shows distinct epochs with transitions from a series of clicks (0–75 ms), to a broad-band, noise-like hiss (80–100ms), to a distinct tone with highly organized harmonic structure (125–175ms). The broad-band hiss epoch clearly has less organized phase structure, resulting in a less distinct, and less stable (with respect to analysis parameters) ridge representation. The transition from the hiss epoch to the harmonic epoch shows the formation of narrow-band components which are modulated in parallel bands, but in different directions around 5kHz.

V. Discussion

In this paper we developed a method based on a contour TFR of sound [8], [9] that identifies signal components by persistent phase progression across time or frequency. The distinguishing features of the present approach is the use of the CT at various chirp rates and time scales to extract multiple ridges per component, and then select from this overcomplete representation the longest, and most stable ridges; the ridges that occur with low probability relative to noise. We demonstrated the viability of this approach for signal enhancement in noise, and elementary stream segregation of crossed components. Quantifying the quality of signal denoising by resynthesis of the chirplet skeleton, and characterization of ridge sampling properties are the subjects of on-going work.

A. From directional derivatives to ridges

Guillemain and Kronland-Martinet [35] and Lim et al. [8] suggested the utilization of separate partial derivatives of the $\phi_g(\tau,\omega)$ in order to identify signal components that are either horizontally- or vertically-oriented in the time-frequency plane. Lim et al. [8] however, generalized this notion to suggest that ridges could be defined as a linear combination of partial phase derivatives. As we showed in Sec. III-C the best linear combination to represent a component depends upon the orientation of the component. This approach helps to correct for the bias in IF estimates made by the $\partial_\omega \log|\chi(\tau,\omega)| = 0$ ridge alone as suggested by others [12], [13].

In the present study, we used ridges defined by both (8) and (7) in order to identify both IMT and click-like events, respectively. Use of condition (7) allows the detection of the components which have amplitude variations that are too fast to satisfy the conditions of an IMT function [11]; a property that is not shared with most ridge detection methods, although notable exceptions are [35] and [8].

Although [35] and [8] were developed in the context of derivatives of ϕ_g, there are relationships between the time and frequency derivatives of ϕ_g and $\log|\chi_g|$, allowing us to formulate equivalent definitions of

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSP.2014.2365756, IEEE Transactions on Signal Processing

ridges in terms of $\log|\chi_g|$. This suggests that the two types of ridges introduced in this paper, defined by $\log|\chi_c|$, also have equivalent formulations based on ϕ_c. Indeed, it can be shown (Appendix A) that the partial derivatives of $\log|\chi|$ with respect to ω and τ are equivalent to linear combinations of the ω and τ derivatives of the CT phase.

According to Eqns. (15) and (16), the approach proposed by Lim et al. [8] suggests that the modulation rate of sufficiently isolated components can be identified by a single transform, rather than the multiple transforms required for the chirplet ridges described in the present study. This is a distinct computational advantage of the directed Gabor ridges as compared to the chirplet ridges. A further advantage of the directional Gabor ridges is that both chirp-like and click-like ridges can be identified in a single, unified framework. However, based on (16), these ridges are likely to be best suited to use in regimes where amplitude modulation of the signal is slow compared to the width of the analysis kernel (i.e. $\sigma/s \ll 1$). Chirplet ridges have the advantage of identifying linear frequency modulation, even for fast amplitude modulated components and they have the potential to discriminate between closely-spaced modulated components without significant interference.

B. Contour bias and prior knowledge

Classical ridge extraction methods using Gabor or Morlet wavelet transforms produce biased estimates of the IF of signals with finite duration and frequency modulation [12], [13]. From the simple example of Section III when $\beta = 0$ we find that this bias becomes worse as the signal modulation rate gets larger and the ratio σ/s becomes large. We showed that use of the CT can help to mitigate this bias if the modulation rate of the analysis kernel approximates the modulation rate of the component.

As $\sigma/s \to 0$, the slope of the IF estimate is weighted in favor of b (Eq. (12)), the modulation rate indicated by the data, and the bias in the estimation is minimized. However, by using only a short analysis window relative to the signal length (small σ/s), we are not making use of all available data and we risk increasing the variance of the estimate in the presence of noise. On the other hand, as $\sigma/s \to \infty$, the slope estimate is weighted in favor of β, the modulation rate of the analysis kernel. Thus, we can view the ratio σ/s as a parameter which reflects our *a priori* certainty that $\beta = b$ or, more generally that, at time τ, $\beta \approx \phi_k(\tau)''$ on a time scale commensurate with σ, where $\phi_k(\tau)$ is the phase of the kth component.

While we expect the bias in the IF estimate to decrease as $\sigma/s \to 0$, this will also imply an increasing lower bound for the variance of IF estimates. This assertion is supported by the bias and variance of ridge estimates of the IF using the short-time Fourier transform [36] and it can be shown that this result

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSP.2014.2365756, IEEE Transactions on Signal Processing

holds in general for short-time Fourier transforms of demodulated signals [37]. The size of σ will also influence the separability of closely-space components (Fig.1). Thus, for the more general setting of nonlinear modulation and multiple components, such as the signal components in Sec. IV-C, the optimal size of σ for each component will be a function of the distance in the time-frequency plane between components as well as the goodness of the linear modulation approximation.

C. A per-component adaptive TFR

While there are a number of adaptive TFR methods, prior methods tend to either adapt the representation to the entire recording [38], or adapt the window parameters on a per-time and/or per-frequency basis (for examples, see [2], [30], [39]–[45]). Other methods adapt the representation at an atomic level, choosing "best" individual members of a highly redundant dictionary while actively suppressing correlation between coefficients [4], [6], [46].

The prospect of a per-component adaptive TFR was, perhaps, first suggested by Huang at al. with the development of empirical mode decomposition and the Hilbert-Huang transform [24]. More recent methods have been proposed which try to directly address the issue of per-component analysis using synchrosqueezing [37], [47]–[49], but these methods are developed for the "well-separated" IMT components [11] for which the IFs of the components do not cross. Indeed this is feature that is directly exploited for the method developed in [49]. Another problem with other per-component approaches is that adaptation is in terms of a single analysis kernel per component (with the exception of [49] which seems to be able to mitigate bias associated with kernel parameters), rather than aggregating information across multiple values of analysis kernel parameters, as in the present study.

Our approach displays the ability to resolve components that have IF laws that intersect, a direct violation of the "well-separated" condition. While techniques such as the synchrosqueezed wavelet transform can successfully reconstruct crossed components in the resynthesis process (see Fig 9 of [11]), synchrosqueezing does not represent the components as separate on its own. The analyst must decide when two components are different signals prior to reconstructing them independently. In contrast, our approach displays the potential for the representation itself to help distinguish among crossed components, prior to the resynthesis step.

The assumption of phase continuity of the signal is an important factor in the performance of our method. Phase continuity has been recognized by other adaptive TFR methods as well. For example, [50] expressed this importance through the total variation norm, that is, the method sought to decompose a time series into IMF-like functions that were smooth enough at each iteration to minimize a third-order

total variation norm.

Phase continuity influences our method in two ways. First, this property is essential to the length and ridge density metrics for signal amplification discussed in Sec. IV. Fig.s 5 and 6 support the analysis of Sec. IV-B and IV-A by showing that the longest ridges correspond to good matches with signal components and that ridge stability can be used to identify separate signal components, even when they cross in the time-frequency plane.

Second, phase averaging helps to explain how the chirplet ridges can achieve ridge detection in spite of crossed components. We conjecture that crossed-component discrimination works in the following way: The filter kernel has contributions from the signal which are determined by the Heisenberg-Gabor uncertainty ellipse (Fig. 1). When we orient the kernel to match a given component, the coefficient estimate is a weighted average of contributions from the time-frequency support of the analysis kernel. When the kernel matches the modulation rate of a component in Fig.5, the phase of the coefficients near the crossing point will be dominated by contributions due to the components with matched modulation rate, thereby biasing the phase evolution of the transform coefficients to more closely resemble the matched components. Likewise, when the analysis kernel is mismatched, the phase of the coefficients, and their relationships among each other, are dominated either by contributions by noise, or by the mixing of nearby components. This is demonstrated by the matched kernel ridges in Fig. 5C,D and 6C. Kernels mis-matched to all components will have contributions from multiple components, or will have a large share of the phase estimate be influenced by noise, making the phase estimates less likely to display continuous variations on length scales any larger than that determined by the kernel.

VI. Conclusion

The ridges of the chirplet transform magnitude may be used as per-component, adaptive time-frequency representations for both IMT signal components and click-like components. Ridges formed using multiple chirp rates can be combined to form individual component estimates in order to reduce the bias inherent in Gabor/Morlet wavelet ridge extraction methods.

Appendix A

Phase-magnitude relations of the continuous chirplet transform

Distinct relationships are evident between the phase and magnitude of the Gabor transform [25], [26] which allows one to define the phase derivatives in terms of the derivatives of the log-modulus of the

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSP.2014.2365756, IEEE Transactions on Signal Processing

transform, and vice versa. For the Gabor transform (given by (5) with $\beta = 0$) the phase-magnitude relations are given by

$$\sigma^{-2}\frac{\partial \log|\chi_g(\tau,\omega)|}{\partial\omega} = \frac{\partial\phi_g(\tau,\omega)}{\partial\tau} - \omega, \tag{19}$$

$$\sigma^{2}\frac{\partial \log|\chi_g(\tau,\omega)|}{\partial\tau} = -\frac{\partial\phi_g(\tau,\omega)}{\partial\omega}. \tag{20}$$

Analogous phase-magnitude relationships can be derived for the CT as well. However, in contrast to (19) and (20), the coupling takes the form of linear combinations of time and frequency derivatives where the gradient of $\log|\chi_c|$ can be expressed as

$$\begin{pmatrix} \partial_\omega \log|\chi_c(\tau,\omega)| \\ \partial_\tau \log|\chi_c(\tau,\omega)| \end{pmatrix} = \boldsymbol{M}\begin{pmatrix} \partial_\tau\phi_c(\tau,\omega) - \omega \\ -\partial_\omega\phi_c(\tau,\omega), \end{pmatrix} \tag{21}$$

where

$$\boldsymbol{M} = \begin{pmatrix} \sigma^2 & -\beta\sigma^2 \\ -\beta\sigma^2 & 1/\sigma^2 + \beta^2\sigma^2 \end{pmatrix}.$$

ACKNOWLEDGMENT

The authors would like to thank T. Oberlin and S. Meignan for helpful comments on an earlier draft of this paper. They would also like to thank the anonymous reviewers for their effort in evaluating this manuscript.

REFERENCES

[1] S. Mann and S. Haykin, "The chirplet transform: A generalization of Gabor's logon transform," *Vision Interface '91*, pp. 205–212, June 3-7, iSSN 0843-803X.

[2] ——, "Adaptive chirplet transform: an adaptive generalization of the wavelet transform," *Optical Engineering*, vol. 31, no. 6, pp. 1243–1256, 1992.

[3] ——, "The chirplet transform: Physical considerations," *IEEE Trans. Signal Processing*, vol. 43, no. 11, p. 2745, 1995.

[4] S. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Transactions on Signal Processing*, vol. 41, no. 12, pp. 3397–3412, 1993.

[5] M. Zibulevsky and B. Pearlmutter, "Blind source separation by sparse decomposition in a signal dictionary," *Neural Computation*, vol. 13, pp. 863–882, 2001.

[6] S. Chen, D. Donoho, and M. Saunders, "Atomic decomposition by basis pursuit," *SIAM Journal of Scientific Computing*, vol. 20, no. 1, pp. 33–61, 1998.

[7] D. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

[8] Y. Lim, B. Shinn-Cunningham, and T. Gardner, "Sparse contour representations of sound," *IEEE Signal Processing Letters*, vol. 19, no. 10, pp. 684–687, 2012.

[9] ——, "Stable time-frequency contours for sparse signal representation," in *Signal Processing conference (EUSIPCO), Proceedings of the 21st European*. Marrakech, Morocco: IEEE, September 2013, pp. 1–5.

[10] H. Xie, J. Lin, Y. Lei, and Y. Liao, "Fast-varying am-fm components extraction based on an adaptive stft," *Digital Signal Processing*, vol. 22, pp. 664–670, 2012.

[11] I. Daubechies, L. J, and W. H-T, "Synchrosqueezed wavelet transforms: An empirical mode decomposition-like tool," *Applied and Computational Harmonic Analysis*, vol. 30, pp. 243–261, 2011.

[12] N. Delprat, B. Escudié, P. Guillemain, R. Kronland-Martinet, P. Tchamitchian, and Torrésani, "Asymptotic wavelet and gabor analysis: Extraction of instantaneous frequency," *IEEE Transactions on Information Theory*, vol. 38, no. 2, pp. 644–664, 1992.

[13] P. Guillemain and R. Kronland-Martinet, "Characterization of acoustic signals through continuous linear time-frequency representations," *Proceedings of the IEEE*, vol. 84, no. 4, pp. 561–585, 1996.

[14] R. Carmona, W. Hwang, and B. Torrésani, "Characterization of signals by the ridges of their wavelet tranforms," *IEEE Transactions on Signal Processing*, vol. 45, no. 10, pp. 2586–2590, 1997.

[15] T. Kijewski and A. Kareem, "Wavelet transforms for system identification in civil engineering," *Computer-Aided Civil and Infrastructure Engineering*, vol. 18, pp. 339–355, 2003.

[16] T.-P. Le and P. Argoul, "Continuous wavelet transform for modal identification using free decay response," *Journal of Sound and Vibration*, vol. 277, no. 1-2, pp. 73–100, 2004.

[17] C. Wang, W.-X. Ren, Z.-C. Wang, and H.-P. Zhu, "Instantaneous frequency identification of time-varying structures by continuous wavelet transform," *Engineering Structures*, vol. 52, pp. 17–25, 2013.

[18] N. Özkurt and F. Savaci, "Reconstruction of non stationary signals alone the wavelet ridges," *International Journal of Bifurcation and Chaos*, vol. 16, no. 1, pp. 191–198, 2006.

[19] S. Roux, T. Cenier, S. Garcia, P. Litaudon, and N. Buoniso, "A wavelet-based method for local phase extraction from a multi-frequency oscillatory signal," *Journal of Neroscience Methods*, vol. 160, pp. 135–143, 2007.

[20] P. Litaudon, S. Garcia, and N. Buonviso, "Strong coupling between pyramidal cell activity and network oscillations in the olfactory cortex," *Neuroscience*, vol. 156, pp. 781–787, 2008.

[21] K. Kodera, C. De Villedary, and R. Gendrin, "A new method for the numerical analysis of non-stationary signals," *Physics of the Earth and Planetary Interiors*, vol. 12, pp. 142–150, 1976.

[22] F. Auger and P. Flandrin, "Improving the readability of time-frequency and time-sale representations by the reassignment method," *IEEE Transactions on Signal Processing*, vol. 43, no. 5, pp. 1068–1089, 1995.

[23] I. Daubechies and S. Maes, "A nonlinear squeezing of the continuous wavelet transform based on auditory nerve models," in *Wavelets in Medicine and Biology*, A. Aldroubi and M. Uner, Eds. CRC Press, 1996, pp. 527–546.

[24] N. Huang, Z. Shen, S. Long, M. Wu, H. Shih, Q. Zheng, N. Yen, C. Tung, and L. H.H., "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis," *Proceedings: Mathematical, Physical and Engineering Sciences*, vol. 454, pp. 903–995, 1998.

[25] F. Auger, E. Chassande-Mottin, and P. Flandrin, "On phase-magnitde relationships in the short-time Fourier transform," *IEEE Signal Processing Letters*, vol. 19, no. 5, pp. 267–270, 2012.

[26] E. Chassande-Mottin, I. Daubechies, F. Auger, and P. Flandrin, "Differential reassignment," *IEEE Signal Processing Letters*, vol. 4, no. 10, pp. 293–294, 1997.

[27] R. Carmona, W. Hwang, and B. Torrésani, *Practical Time-Frequency Analysis: Gabor and Wavelet Transforms, with an Implementation in S*. San Diego, CA: Academic Press, 1998.

[28] S. Borum and K. Jensen, "Additive analysis/synthesis using analytically derived windows," in *Proceedings of the 2nd COST G-6 Workshop on Digital Audio Effects, NTNU, Trondheim, December 9-11*, 1999.

[29] K. Kodera, R. Gendrin, and C. De Villedary, "Analysis of time-varying signals with small BT values," *IEEE Transactions of Acoustics, and Signal Processing*, vol. ASSP-26, no. 1, pp. 64–76, 1978.

[30] D. Jones and T. Parks, "A high resolution data-adaptive time-frequency representation," *Acoustics, Speech and Signal Processing, IEEE Transactions on*, vol. 38, no. 12, pp. 2127–2135, 1990.

[31] B. Weisburn, T. Parks, and R. Shenoy, "Separation of transient signals," in *Sixth IEEE Signal Processing Workshop, Yosemite National Park, CA, Oct 02-05*, 1994.

[32] B. Weisburn and R. Shenoy, "Time-frequency strip filters," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP-96, NTNU, Atlanta, GA, May 07-10*, vol. 3, 1996.

[33] T. Gardner and M. Magnasco, "Sparse time-frequency representations," *Proceedings of the National Academy of Sciences*, vol. 103, no. 16, pp. 6094–6099, 2006.

[34] P. Balazs, D. Bayer, F. Jaillet, and P. Sondergaard, "The phase derivative around zeros of the short-time fourier transform," *eprint arXiv:1103.0409*, 2011.

[35] P. Guillemain and R. Kronland-Martinet, "Horizontal and vertical ridges associated to the continuous wavelet transform," in *Proceedings of the IEEE-SP international Symposium, Time-Frequency and Time-Scale Analysis*, Victoria, BC, October 1992, pp. 63–66.

[36] L. Stanović, M. Dakonović, and V. Ivanović, "Performance of spectrogram and if estimator," *IEEE Electronics Letters*, vol. 37, no. 12, pp. 797–799, 2001.

[37] S. Wang, X. Chen, G. Cai, B. Chen, X. Li, and Z. He, "Matching demodulation transform and synchrosqueezing in time-frequency analysis," *IEEE Transactions on Signal Processing*, In Press.

[38] R. Baraniuk and D. Jones, "Signal-dependent time-frequency analysis using a radially gaussian kernel," *Signal Processing*, vol. 32, pp. 263–284, 1993.

[39] D. Jones and R. Baraniuk, "A simple scheme for adapting time-frequency representations," *Signal Processing, IEEE Transactions on*, vol. 42, no. 12, pp. 3530–3535, 1994.

[40] ——, "An adaptive optimal-kernel time-frequency representation," *IEEE Transactions on Signal Processing*, vol. 43, no. 10, pp. 2361–2371, 1995.

[41] R. Czerwinski and D. Jones, "Adaptive short-time Fourier analysis," *Signal Processing Letters, IEEE*, vol. 4, no. 2, pp. 42–45, 1997.

[42] H. Kwok and D. Jones, "Improved instantaneous frequency estimation using an adaptive short-time Fourier transform," *Signal Processing, IEEE Transactions on*, vol. 48, no. 10, pp. 2964–2972, 2000.

[43] L. Stanković and V. Katkovnik, "Wigner distribution of noisy signals with adaptive time-frequency varying window," *IEEE Transaction on Signal Processing*, vol. 47, no. 4, pp. 1099–1108, 1999.

[44] J. Gilles, "Empirical wavelet transform," *IEEE Trans on Signal Processing*, vol. 61, no. 16, pp. 3999–4010, August 2013.

[45] K. Dragomiretskiy and D. Zosso, "Variational mode decomposition," *IEEE Transactions on Signal Processing*, vol. 62, no. 3, pp. 531–544, 2014.

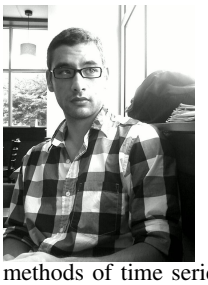
[46] R. Coifman and M. Wickerhauser, "Entropy-based algorithms for best basis selection," *Information Theory, IEEE Transactions on*, vol. 38, no. 2, pp. 713–718, 1992.

[47] S. Meignen, T. Oberlin, and S. McLaughlin, "A new algorithm for mulitcomponent signals analysis based on synchrosqueezing: With an application to signal sampling and denoising," *IEEE Transactions on Signal Processing*, vol. 60, no. 11, pp. 5787–5798, 2012.

[48] G. Thakur, E. Brevdo, N. Fuckar, and H.-T. Wu, "The synchrosqueezing algorithm for time-varying spectral analysis: Robustness properties and new paleoclimate applications," *Signal Processing*, vol. 93, pp. 1079–1094, 2013.

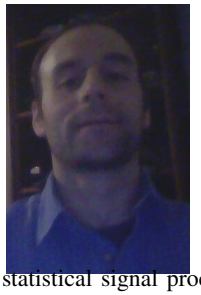
[49] Z.-C. Wang, W.-X. Ren, and J.-L. Liu, "A synchrosqueezed wavelet transform enhanced by extended analytical mode decomposition method for dynamic signal reconstruction," *Journal of Sound and Vibration*, vol. 332, no. 22, pp. 6016–6028, 2013.

[50] T. Hou and S. Zuoqiang, "Adaptive data analysis via sparse time-frequency representation," *Advances in Adaptive Data Analysis*, vol. 3, no. 1& 2, pp. 1–28, 2011.



Mikio Aoi received a B.S. degree in kinesiology and exercise science from California State University – Long Beach, Long Beach, CA, in 2006 and a Ph.D. degree in Mathematical Biology at North Carolina State University in 2011 where his dissertation was on the assessment of cerebral blood flow regulation using nonlinear biophysical models and non-invasively recorded data.

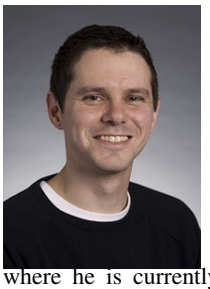
He is currently a postdoctoral associate in the Department of Mathematics & Statistics at Boston University, Boston, MA, in association with the Cognitive Rhythms Collaborative, where he is developing methods of time series analysis for use with neural data.



Kyle Lepage was born in Edmonton, Canada, in 1976. He received the BASc degree in engineering physics/electrical engineering and the MSc degree in physics from the University of British Columbia, Canada, in 2000 and 2004, respectively. In 2009 he received a PhD in mathematics and engineering from Queen's University, Canada. He joined the Department of Mathematics and Statistics at Boston University in 2009 as a post-doctoral scholar, becoming in 2012 a research fellow within the Cognitive Rhythms Collaborative. His current interests are applied statistics in neuroscience, time-series analysis, statistical signal processing, tracking and estimation of evoked connectivity, experimental design and system identification. He has authored fifteen papers in peer reviewed journals spanning the fields of astrophysics, optics, geophysics, neuroscience and signal processing. Of these, he is the first author of eleven. In 2000, Dr. Lepage was a developer at Point Grey Research, and, in 2001, a senior software engineer at DSI Datotech Systems.

Yoonsoeb Lim received the B.S. degree in mechanical engineering and M.S. degree in electrical engineering from the Seoul National University, Seoul, Korea, in 2002 and 2004 respectively. In 2014, he received a Ph.D. in cognitive and neural system, Boston University, Boston, MA, USA. From 2004 to 2006, he was a research engineer at Hyundai Motor Company, Hwaseong, Gyeonggi-do, Korea. And he worked at Center for Cognitive Robotics, Korea Institute of Science and Technology, Seoul, Korea, between 2006 and 2008.

He is currently working at Center for Bionics, Biomedical Engineering Research Institute, Korea Institute of Science and Technology, Seoul, Korea. His current research interests include time-frequency analysis, speech recognition and auditory stream analysis.



Uri T. Eden received the B.S. degree in mathematics and in engineering and applied sciences from the California Institute of Technology, Pasadena, in 1999, and the S.M. degree in engineering sciences and the Ph.D. degree in engineering sciences from Harvard University, Cambridge, MA, in 2002 and 2005, respectively. He was a Postdoctoral Fellow in the Brain and Cognitive Sciences Department at the Massachusetts Institute of Technology, Cambridge, from 2005-2006.

In 2006, he joined the Department of Mathematics and Statistics an Boston University, Boston, MA, where he is currently an Associate Professor and the Assistant Director of the Program in Statistics. He received an NSF CAREER award in 2007. His research focuses on developing mathematical and statistical methods to analyze neural spiking activity, integrating methodologies related to model identification, statistical inference, signal processing, and stochastic estimation and control.

Timothy J. Gardner received a B.A. degree from the department of Physics, Princeton University, Princeton, NJ in 1996, and a Ph.D. degree in Biology and Physics from Rockefeller University, New York, NY in 2003. From 2003-2008 he pursued postdoctoral training in neuroscience from Rockefeller University, New York, NY and Massachusetts Institute of Technology, Boston MA.

In 2009, he joined the Biology department at Boston University where he is currently Assistant Professor and member of the Center for Computational Neuroscience and Neural Technology. He received a Career Award at the Scientific Interface from the Burroughs Wellcome Fund in 2005, and Smith Family award for Excellence in Biomedical Research in 2009. His research interests include brain machine interfaces, sensory-motor learning, and signal processing.