Sparse time-frequency representations

Timothy J. Gardner[†‡] and Marcelo O. Magnasco[†§]

[†]Center for Studies in Physics and Biology, The Rockefeller University, 1230 York Avenue, New York, NY 10021; and [‡]Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, E19-502B, Cambridge, MA 02139

Communicated by Mitchell J. Feigenbaum, The Rockefeller University, New York, NY, March 1, 2006 (received for review January 20, 2006)

Auditory neurons preserve exquisite temporal information about sound features, but we do not know how the brain uses this information to process the rapidly changing sounds of the natural world. Simple arguments for effective use of temporal information led us to consider the reassignment class of time-frequency representations as a model of auditory processing. Reassigned time-frequency representations can track isolated simple signals with accuracy unlimited by the time-frequency uncertainty principle, but lack of a general theory has hampered their application to complex sounds. We describe the reassigned representations for white noise and show that even spectrally dense signals produce sparse reassignments: the representation collapses onto a thin set of lines arranged in a froth-like pattern. Preserving phase information allows reconstruction of the original signal. We define a notion of ''consensus,'' based on stability of reassignment to time-scale changes, which produces sharp spectral estimates for a wide class of complex mixed signals. As the only currently known class of time-frequency representations that is always ''in focus'' this methodology has general utility in signal analysis. It may also help explain the remarkable acuity of auditory perception. Many details of complex sounds that are virtually undetectable in standard sonograms are readily perceptible and visible in reassignment.

auditory | reassignment | spectral | spectrograms | uncertainty

Time-frequency analysis seeks to decompose a one-dimensional signal along two dimensions, a time axis and a frequency axis; the best known time-frequency representation is the musical score, which notates frequency vertically and time horizontally. These methods are extremely important in fields ranging from quantum mechanics (1–5) to engineering (6, 7), animal vocalizations (8, 9), radar (10), sound analysis and speech recognition (11–13), geophysics (14, 15), shaped laser pulses (16–18), the physiology of hearing, and musicography.[¶] A central question of auditory theory motivates our study: what algorithms do the brain use to parse the rapidly changing sounds of the natural world? Auditory neurons preserve detailed temporal information about sound features, but we do not know how the brain uses it to process sound. Although it is accepted that the auditory system must perform some type of time-frequency analysis, we do not know which type. The many inequivalent classes of time-frequency distributions (2, 3, 6) require very different kinds of computations: linear transforms include the Gabor transform (19), quadratic transforms [known as Cohen's class (2, 6)] include the Wigner–Ville (1) and Choi–Williams (20) distributions, and higher-order in the signal, include multitapered spectral estimates (21–24), the Hilbert–Huang distribution (25, 26), and the reassigned spectrograms (27–32) whose properties are the subject of this article.

Results and Discussion

The auditory nerve preserves information about phases of oscillations much more accurately than information about amplitudes, a feature that inspired temporal theories of pitch perception (33–37). Let us consider what types of computation would be simple to perform given this information. We shall idealize the cochlea as splitting a sound signal $\chi(t)$ into many component signals $\chi(t,\omega)$ indexed by frequency ω

$$\chi(t,\omega) = \int e^{-(t-t')^2/2\sigma^2} e^{i\omega(t-t')} x(t')dt'. \qquad [1]$$

χ is the Gabor transform (19) or short-time Fourier transform (STFT) of the signal $x(t)$. The parameter σ is the temporal resolution or time scale of the transform, and its inverse is the frequency resolution or bandwidth. The STFT χ is a smooth function of both t and ω and is strongly correlated for $\Delta t < \sigma$ or $\Delta\omega < 1/\sigma$. In polar coordinates it decomposes into magnitude and phase, $\chi(t,\omega) = |\chi|(t,\omega)e^{i\phi(t,\omega)}$. A plot of $|\chi|^2$ as a function of (t, ω) is called the spectrogram (3, 38), sonogram (8), or Husimi distribution (2, 4) of the signal $x(t)$. We call $\phi(t,\omega)$ the phase of the STFT; it is well defined for all (t, ω) except where $|\chi| = 0$. We shall base our representation on ϕ.

We can easily derive two quantities from ϕ: the time derivative of the phase, called the instantaneous frequency (31), and the current time minus the frequency derivative of the phase (the local group delay), the instantaneous time:

$$\omega_{ins}(\omega,t) = \frac{\partial\phi}{\partial t}$$
$$t_{ins}(\omega,t) = t - \frac{\partial\phi}{\partial\omega}. \qquad [2]$$

Neural circuitry can compute or estimate these quantities from the information in the auditory nerve: the time derivative, as the time interval between action potentials in one given fiber of the auditory nerve, and the frequency derivative from a time interval between action potentials in nearby fibers, which are tonotopically organized (34).

Any neural representation that requires explicit use of ω or t is unnatural, because it entails "knowing" the numerical values of both the central frequencies of fibers and the current time. Eq. **2** affords a way out: given an estimate of a frequency and one of a time, one may plot the instantaneous estimates against each other, making only implicit use of (t, ω), namely, as the indices in an implicit plot. So for every pair (t, ω), the pair (t_{ins}, ω_{ins}) is computed from Eq. **2**, and the two components are plotted against each other in a plane that we call (abusing notation) the (t_{ins}, ω_{ins}) plane. More abstractly, Eq. **2** defines a transformation T

$$(\omega, t) \xrightarrow{T_{\{x\}}} (\omega_{ins}, t_{ins}). \qquad [3]$$

The transformation is signal-dependent because ϕ has to be computed from Eq. **1**, which depends on the signal x, hence the subscript $\{x\}$ on T.

Conflict of interest statement: No conflicts declared.

Freely available online through the PNAS open access option.

Abbreviation: STFT, short-time Fourier transform.

[§]To whom correspondence should be addressed. E-mail: magnasco@rockefeller.edu.

[¶]Hainsworth, S. W., Macleod, M. D. & Wolfe, P. J., IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, Oct. 21–24, 2001, Mohonk, NY, p. 26.

www.pnas.org/cgi/doi/10.1073/pnas.0601707103

Table 1. The values of the estimates for simple test signals

	Tones $x = e^{i\omega_0 t}$	Clicks $x = \delta(t - t_0)$	Sweeps $x = e^{i\alpha t^2/2}$
$\omega_{ins}(\omega, t) = \dfrac{\partial\phi}{\partial t}$	ω_0	ω	αt_{ins}
$t_{ins}(\omega, t) = t - \dfrac{\partial\phi}{\partial\omega}$	t	t_0	$\cdots$

The transformation given by Eqs. **2** and **3** has optimum time-frequency localization properties for simple signals (27, 28). The values of the estimates for simple test signals are given in Table 1.

So for a simple tone of frequency ω_0, the whole (t, ω) plane is transformed into a single line, (t, ω_0); similarly, for a "click," Dirac delta function localized at time t_0, the plane is transformed into the line (t_0, ω); and for a frequency sweep where the frequency increases linearly with time as αt, the plane collapses onto the line $\alpha t_{ins} = \omega_{ins}$. (The full expression in the frequency sweep case is given in *Appendix*.) So for these simple signals the transformation $(t, \omega) \rightarrow (t_{ins}, \omega_{ins})$ has a simple interpretation as a projection to a line that represents the signal. The transformation's time-frequency localization properties are optimal, because these simple signals, independently of their slope, are represented by lines of zero thickness. Under the STFT the simple signals above transform into strokes with a Gaussian profile, with vertical thickness $1/\sigma$ (tones) and horizontal thickness σ (clicks).

These considerations lead to a careful restatement of the uncertainty principle. In optics it is well known that there is a difference between precision and resolution. Resolution refers to the ability to establish that there are two distinct objects at a certain distance, whereas precision refers to the accuracy with which a single object can be tracked. The wavelength of light limits resolution, but not precision. Similarly, the uncertainty principle limits the ability to separate a sum of signals as distinct objects, rather than the ability to track a single signal. The best-known distribution with optimal localization, the Wigner–Ville distribution (1), achieves optimal localization at the expense of infinitely long range in both frequency and time. Because it is bilinear, the Wigner transform of a sum of signals causes the signals to interfere or beat, no matter how far apart they are in frequency or time, seriously damaging the resolution of the transform. This nonlocality makes it unusable in practice and led to the development of Cohen's class. In contrast, it is readily seen from Eq. **1** that the instantaneous time-frequency reassignment cannot cause a sum of signals to interfere when they are further apart than a Fourier uncertainty ellipsoid; therefore, it can resolve signals as long as they are further apart than the Fourier uncertainty ellipsoid, which is the optimal case. Thus, reassignment with instantaneous time-frequency estimates has optimal precision (unlimited) and optimal resolution (strict equality in the uncertainty relation).

We shall now derive the formula needed to implement numerically this method. First, the derivatives of the transformation defined by Eq. **2** should be carried out analytically. The Gaussian window in the STFT has a complex analytic structure; defining $z = t/\sigma - i\sigma\omega$ we can write the STFT as

$$G(z) = \int e^{-(z-t'/\sigma)^2/2} x(t')dt' = \chi e^{(\sigma\omega)^2/2}. \qquad [4]$$

So up to the factor $e^{(\sigma\omega)^2/2}$, the STFT is an analytic function of z (29). Defining

$$\eta(t, \omega) = \frac{1}{\sigma}\int (t' - t)e^{-(t-t')^2/2\sigma^2}\, e^{i\omega(t-t')} x(t')dt', \qquad [5]$$

we obtain in closed form

$$\omega_{ins}(\omega, t) = \partial_t \mathrm{Im}\ln\chi = \omega + \frac{1}{\sigma}\mathrm{Im}\frac{\eta}{\chi}(\omega, t)$$

$$t_{ins}(\omega, t) = t - \partial_\omega \mathrm{Im}\ln\chi = t + \sigma\,\mathrm{Re}\frac{\eta}{\chi}(\omega, t).$$

So the mapping is a quotient of convolutions:

$$z_{ins} = z + (\eta/\chi)^*, \qquad [6]$$

where $*$ is the complex conjugate. Therefore, computing the instantaneous time-frequency transformation requires only twice the numerical effort of an STFT.

Any transformation F: $(\omega, t) \rightarrow (\omega_{ins}, t_{ins})$ can be used to transform a distribution in the (ω, t) plane to its corresponding distribution in the (ω_{ins}, t_{ins}) plane. If the transformation is invertible and smooth, the usual case for a coordinate transformation, this change of coordinates is done by multiplying by the Jacobian of F the distribution evaluated at the "old coordinates" $F^{-1}(\omega_{ins}, t_{ins})$. Similarly, the transformation T given by Eqs. **2** and **3** transforms a distribution f in the (ω, t) plane to the (ω_{ins}, t_{ins}) plane, called a "reassigned f" (28–30, 38)[§]. However, because T is neither invertible nor smooth, the reassignment requires an integral approach, best visualized as the numerical algorithm shown in Fig. 1: generate a fine grid in the (t, ω) plane, map every element of this grid to its estimate (t_{ins}, ω_{ins}), and then create a two-dimensional histogram of the latter. If we weight the histogrammed points by a positive-definite distribution $f(t, \omega)$, the

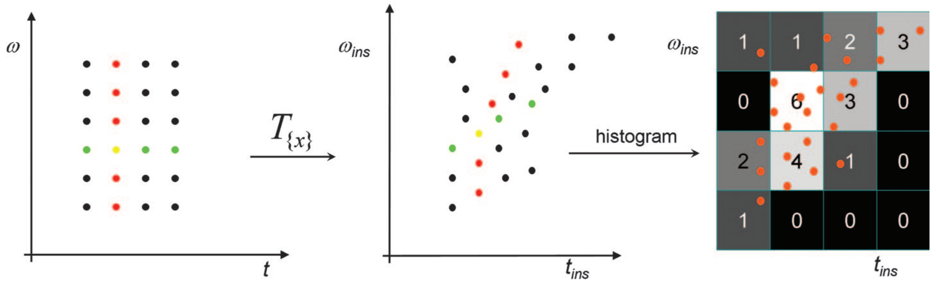

Fig. 1. Reassignment. $T_{\{x\}}$ transforms a fine grid of points in (t, ω) space into a set of points in (t_{ins}, ω_{ins}) space; we histogram these points by counting how many fall within each element of a grid in (t_{ins}, ω_{ins}) space. The contribution of each point to the count in a bin may be unweighted, as shown above, or the counting may be weighted by a function $g(t, \omega)$, in which case we say we are computing the reassigned g. The weighting function is typically the sonogram from Eq. **1**. An unweighted count can be viewed as reassigning 1, or more formally, as the reassigned Lebesgue measure. For a given grid size in (t_{ins}, ω_{ins}) space, as the grid of points in the original (t, ω) space becomes finer, the values in the histogram converge to limiting values.

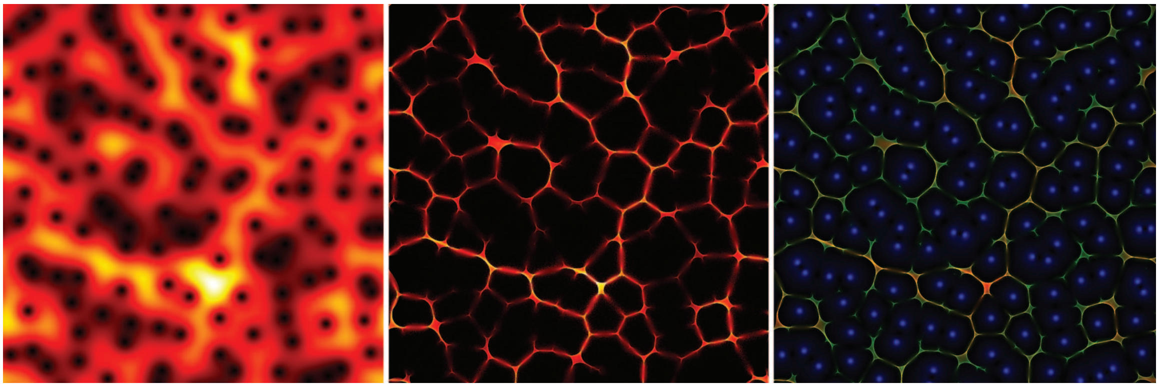

Fig. 2. Analysis of a discrete white-noise signal, consisting of *N* independent identically distributed Gaussian random variables. (*Left*) $|\chi|$, represented by false colors; red and yellow show high values, and black shows zero. The horizontal axis is time and the vertical axis is frequency, as in a musical score. Although the spectrogram of white noise has a constant expectation value, its value on a specific realization fluctuates as shown here. Note the black dots pockmarking the figure; the zeros of χ determine the local structure of the reassignment transformation. (*Center*) The reassigned spectrogram concentrates in a thin, froth-like structure and is zero (black) elsewhere. (*Right*) A composite picture showing reassigned distributions and their relationship to the zeros of the STFT; the green channel of the picture shows the reassigned Lebesgue measure, the red channel displays the reassigned sonogram, and the blue channel shows the zeros of the original STFT. Note that both distributions have similar footprints (resulting in yellow lines), with the reassigned histogram tracking the high-intensity regions of the sonogram and form a froth- or Voronoi-like pattern surrounding the zeros of the STFT.

histogram $g(t_{ins},\ \omega_{ins})$ is the reassigned or remapped *f*; if the points are unweighted (i.e., $f = 1$), we have reassigned the uniform (Lebesgue) measure. We call the class of distributions so generated the reassignment class. Reassignment has two essential ingredients: a signal-dependent transformation of the time-frequency plane, in our case the instantaneous time-frequency mapping defined by Eq. **2**, and the distribution being reassigned. We shall for the moment consider two distributions: that obtained by reassigning the spectrogram $|\chi|^2$, and that obtained by reassigning 1, the Lebesgue measure. Later, we shall extend the notion of reassignment and reassign χ itself to obtain a complex reassigned transform rather than a distribution.

Neurons could implement a calculation homologous to the method shown in Fig. 1, e.g., by using varying delays (39) and the "many-are-equal" logical primitive (40), which computes histograms.

Despite its highly desirable properties, the unwieldy analytical nature of the reassignment class has prevented its wide use. Useful signal estimation requires us to know what the transformation does to both signals and noise. We shall now demonstrate the usefulness of reassignment by proving some important results for white noise. Fig. 2 shows the sonogram and reallocated sonogram of a discrete realization of white noise. In the discrete case, the signal is assumed to repeat periodically, and a sum replaces the integral in Eq. **1**. If the signal has *N* discrete values we can compute *N* frequencies by Fourier transformation, so the time-frequency plane has N^2 "pixels," which, having been derived from only *N* numbers, are correlated (19). Given a discrete realization of white noise, i.e., a vector with *N* independent Gaussian random numbers, the STFT has exactly *N* zeros on the fundamental tile of the (t,ω) plane, so, on average, the area per zero is 1. These zeros are distributed with uniform density, although they are not independently distributed.

Because the zeros of the STFT are discrete, the spectrogram is almost everywhere nonzero. In Fig. 2 the reassigned distributions are mostly zero or near zero: nonzero values concentrate in a froth-like pattern covering the ridges that separate neighboring zeros of the STFT. The Weierstrass representation theorem permits us to write the STFT of white noise as a product over the zeros × the exponential of an entire function of quadratic type:

$$G(z) = e^{Q(z)} \prod_i (1 - z/z_i),$$

where $Q(z)$ is a quadratic polynomial and z_i is the zeros of the STFT. The phase $\phi = \operatorname{Im} \ln G$ and hence the instantaneous estimates in Eq. **2** become sums of magnetic-like interactions

$$\frac{\partial\phi}{\partial t} = \frac{\partial}{\partial t} \operatorname{Im} \ln G = \operatorname{Im}\left(\frac{\partial}{\partial z} \ln G\ \frac{\partial z}{\partial t}\right),$$

where

$$\partial_x \ln G = \partial_z Q/Q - \sum_i \frac{1}{z_i - z},$$

and similarly for the instantaneous time; so the slow manifolds of the transformation *T*, where the reassigned representation has its support, are given by equations representing equilibria of magnetic-like terms.

The reassigned distributions lie on thin strips between the zeros, which occupy only a small fraction of the time-frequency plane; see *Appendix* for an explicit calculation of the width of the stripes in a specific case. The fraction of the time-frequency plane occupied by the support of the distribution decreases as the sequence becomes longer, as in Fig. 3; therefore, reassigned distributions are sparse in the time-frequency plane. Sparse representations are of great interest in neuroscience (41–43), particularly in auditory areas, because most neurons in the primary auditory cortex A1 are silent most of the time (44–46).

Signals superposed on noise move the zeros away from the representation of the pure signal, creating crevices. This process is shown in Fig. 4. When the signal is strong and readily detectable, its reassigned representation detaches from the underlying "froth" of noise; when the signal is weak, the reassigned representation merges into the froth, and if the signal is too weak its representation fragments into disconnected pieces.

Distributions are not explicitly invertible; i.e., they retain information on features of the original signal, but lose some information (for instance about phases) irretrievably. It would be desirable to reassign the full STFT χ rather than just its spectrogram $|\chi|^2$. Also the auditory system preserves accurate timing information all of the way to primary auditory cortex (47). We shall now extend the reassignment class to complex-valued functions; to do this we need to reassign phase information, which requires more care than reassigning positive values,

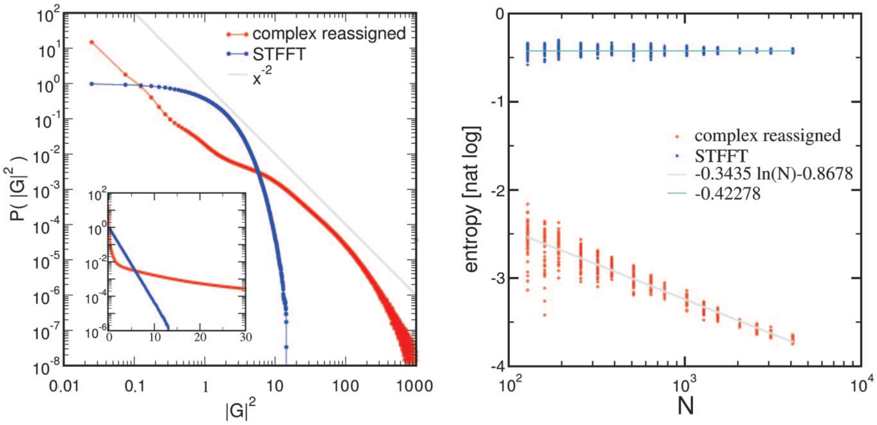

Fig. 3. The complex reassigned representation is sparse. We generated white-noise signals with N samples and computed both their STFFT and its complex reassigned transform on the $N \times N$ time-frequency square. The magnitude squared of either transform is its ''energy distribution.'' (*Left*) The probability distribution of the energy for both transforms computed from 1,000 realizations for $N = 2{,}048$. The energy distribution of the STFT (blue) agrees exactly with the expected e^{-x} (see the log-linear plot inset). The energy distribution of the complex reassigned transform (red) is substantially broader, having many more events that are either very small or very large; we show in gray the power-law x^{-2} for comparison. For the complex reassigned transform most elements of the 2,048 × 2,048 time-frequency plane are close to zero, whereas a few elements have extremely large values. (*Right*) Entropy of the energy distribution of both transforms; this entropy may be interpreted as the natural logarithm of the fraction of the time-frequency plane that the footprint of the distribution covers. For each N, we analyzed 51 realizations of the signal and displayed them as dots on the graph. The entropy of the STFT remains constant, close to its theoretical value of 0.42278 as N increases, whereas the entropy of the complex reassigned transform decreases linearly with the logarithm of N. The representation covers a smaller and smaller fraction of the time-frequency plane as N increases.

because complex values with rapidly rotating phases can cancel through destructive interference. We build a complex-valued histogram where each occurrence of (t_{ins}, ω_{ins}) is weighted by $\chi(t, \omega)$. We must transform the phases so preimages of (t_{ins}, ω_{ins}) add coherently. The expected phase change from (t, ω) to (t_{ins}, ω_{ins}) is $(\omega + \omega_{ins})(t_{ins} - t)/2$, i.e., the average frequency times the time difference. This correction is exact for linear frequency sweeps. Therefore, we reassign χ by histogramming (t_{ins}, ω_{ins}) weighted by $\chi(t, \omega)e^{i(\omega+\omega_{ins})(t_{ins}-t)/2}$. Unlike standard reassignment, the weight for complex reassignment depends on both the point of origin and the destination.

The complex reassigned STFT now shares an important attribute of χ that neither $|\chi|^2$ nor any other positive-definite distribution possesses: explicit invertibility. This inversion is not exact and may diverge significantly for spectrally dense signals. However, we can reconstruct simple signals directly by integrating on vertical slices, as in Fig. 5, which analyzes a chirp (a Gaussian-enveloped frequency sweep). Complex reassignment also allows us to define and compute synchrony between frequency bands: only by using the absolute phases can we

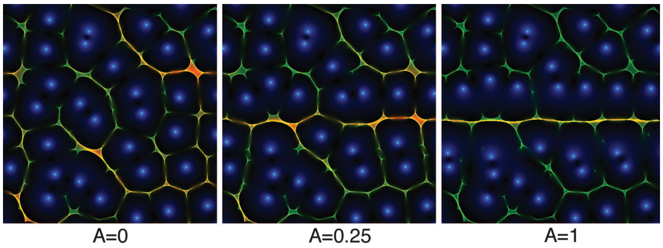
A=0 A=0.25 A=1

Fig. 4. Detection of a signal in a background of noise. Shown are the reassigned distributions and zeros of the Gabor transform as in Fig. 2 *Right*. The signal analyzed here is $x = \zeta(t) + A\sin\omega_0 t$, where $\zeta(t)$ is Gaussian white noise and has been kept the same across the panels. As the signal strength A is increased, a horizontal line appears at frequency ω_0. We can readily observe that the zeros that are far from ω_0 are unaffected; as A is increased, the zeros near ω_0 are repelled and form a crevice whose width increases with A. For intermediate values of A a zigzagging curve appears in the vicinity of ω_0. Note that because the instantaneous time-frequency reassignment is rotationally invariant in the time-frequency plane, detection of a click or a frequency sweep operates through the same principles, even though the energy of a frequency sweep is now spread over a large portion of the spectrum.

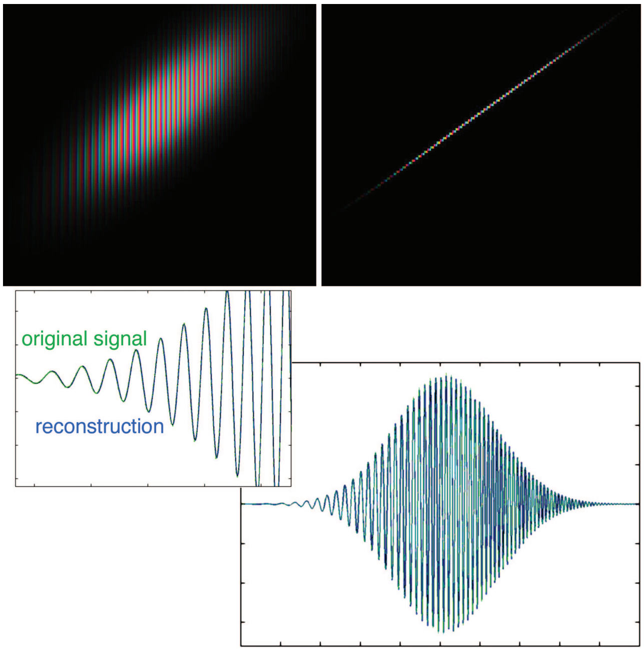

Fig. 5. Reconstruction of a chirp from the complex reassigned STFT. (*Upper Left*) STFT of a chirp; intensity represents magnitude, and hue represents complex phase. The spacing between lines of equal phase narrows toward the upper right, corresponding to the linearly increasing frequency. (*Upper Right*) Complex reassigned STFT of the same signal. The width of this representation is one pixel; the oscillation follows the same pattern. (*Lower*) A vertical integral of the STFT (blue) reconstructs the original signal exactly; the vertical integral of the complex reassigned transform (green) agrees with the original signal almost exactly. (Note the integral must include the mirror-symmetric, complex conjugate negative frequencies to reconstruct real signals.) (*Lower Right*) Full range of the chirp. (*Lower Left*) A detail of the rising edge of the waveform, showing the green and blue curves superposing point by point.

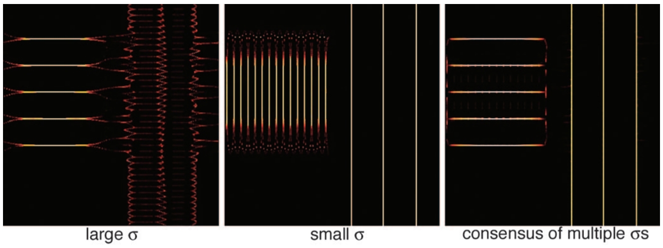

Fig. 6. Consensus finds the best local bandwidth. Analysis of a signal $x(t)$ composed of a series of harmonic stacks followed by a series of clicks; the separation between the stacks times the separation between the clicks is near the uncertainty limit 1/2, so no single σ can simultaneously analyze both. If the analyzing bandwidth is small (*Center*), the stacks are well resolved from one another, but the clicks are not. If the bandwidth is large (i.e., the temporal localization is high, *Left*), the clicks are resolved but the stacks merge. Using several bandwidths (*Right*) resolves both simultaneously.

check whether different components appearing to be harmonics of a single sound are actually synchronous.

We defined the transformation for a single value of the bandwidth σ. Nothing prevents us from varying this bandwidth or using many bandwidths simultaneously, and indeed the auditory system appears to do so, because bandwidth varies across auditory nerve fibers and is furthermore volume-dependent. Performing reassignment as a function of time, frequency, and bandwidth we obtain a reassigned wavelet representation, which we shall not cover in this article. We shall describe a simpler method: using several bandwidths simultaneously and highlighting features that remain the same as the bandwidth is changed. When we intersect the footprints of representations for multiple bandwidths we obtain a consensus only for those features that are stable with respect to the analyzing bandwidth (31), as in Fig. 6. For spectrally more complex signals, distinct analyzing bandwidths resolve different portions of the signal. Yet the lack of predictability in the auditory stream precludes choosing the right bandwidths in advance. In Fig. 6, the analysis proceeds through many bandwidths, but only those bands that are locally optimal for the signal stand out as salient through consensus. Application of consensus to real sounds is illustrated in Fig. 7. This principle may also support robust pattern recognition in the presence of primary auditory sensors whose bandwidths depend on the intensity of the sound.

A final remark about the use of timing information in the auditory system is in order. Because $G(z)$ (Eq. **4**) is an analytic function of z, its logarithm is also analytic away from its zeros, and so

$$\ln G(z) = \ln|\chi| + (\sigma\omega)^2/2 + i\phi$$

satisfies the Cauchy–Riemann relations, from where the derivatives of the spectrogram can be computed in terms of the derivatives of the phase and vice versa, as shown (29):

$$-\sigma^2\frac{\partial}{\partial t}\phi = \frac{1}{|\chi|}\frac{\partial|\chi|}{\partial\omega} + \sigma^2\omega, \quad \frac{\partial}{\partial\omega}\phi = \sigma^2\frac{1}{|\chi|}\frac{\partial|\chi|}{\partial t}.$$

So, mathematically, time-frequency analysis can equivalently be done from phase or intensity information. In the auditory system, though, these two approaches are far from equivalent: estimates of $|\chi|$ and its derivatives must rely on estimating firing rates in the auditory nerve and require many cycles of the signal to have any accuracy. As argued before, estimating the derivatives of ϕ only requires computation of intervals between few spikes.

Our argument that time-frequency computations in hearing use reassignment or a homologous method depends on a few simple assumptions: (*i*) we must use simple operations from information readily available in the auditory nerve, mostly the phases of oscillations; (*ii*) we must make only implicit use of t and ω; (*iii*) phases themselves are reassigned; and (*iv*) perception uses a multiplicity of bandwidths. Assumptions *i* and *ii* led us to the definition of reassignment, *iii* led us to generalize reassignment by preserving phase information, and *iv* led us to define consensus. The result is interesting mathematically because reassignment has many desirable properties. Two features of the resulting representations are pertinent to auditory physiology. First, our representations make explicit information that is readily perceived yet is hidden in standard sonograms. We can perceive detail below the resolution limit imposed by the uncertainty principle that is stable across bandwidth changes, as in Fig. 5. Second, the resulting representations are sparse, which is a prominent feature of auditory responses in primary auditory cortex.

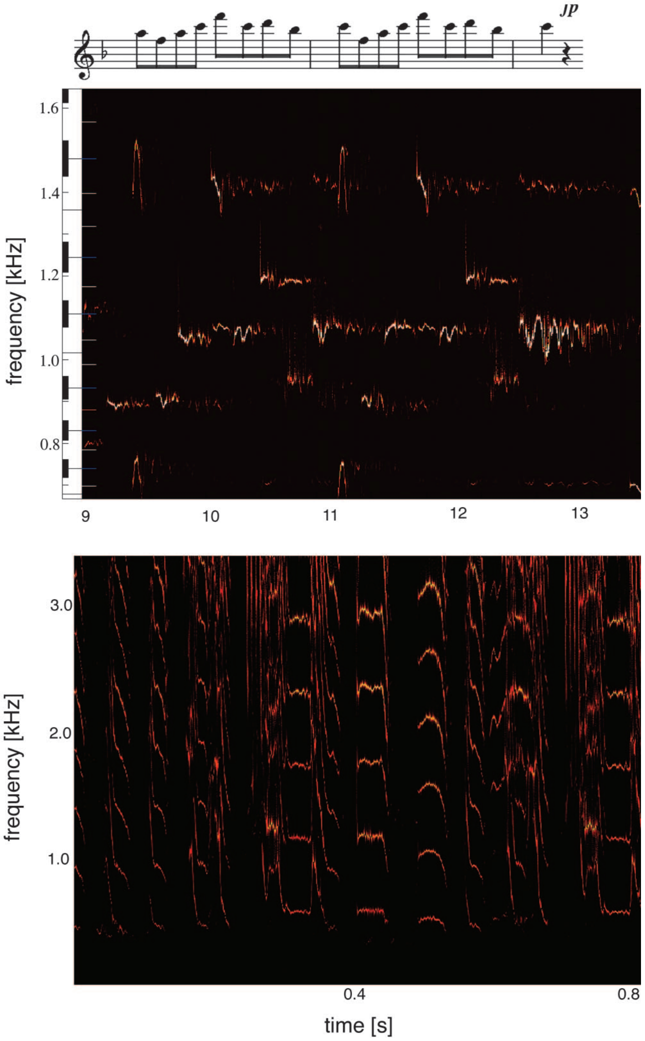

Fig. 7. Application of this method to real sounds. (*Upper*) A fragment of Mozart's ''Queen of the Night'' aria (*Der Hölle Rache*) sung by Cheryl Studer. (*Lower*) A detail of zebra-finch song.

Appendix: Instantaneous Time Frequency for Some Specific Signals

Frequency Sweep. $x(t) = e^{i\alpha t^2/2}$:

$$\chi = \sqrt{\frac{2\pi}{\sigma^{-2} - i\alpha}}\exp\left(-\frac{i\sigma^2\omega^2 + \alpha t(t - 2i\sigma^2\omega)}{2i + 2\alpha\sigma^2}\right)$$

$$\phi = -\mathrm{Im}\,\frac{i\sigma^2\omega^2 + \alpha t(t - 2i\sigma^2\omega)}{2i + 2\alpha\sigma^2} + \mathrm{Im}\ln\sqrt{\frac{2\pi}{\sigma^{-2} - i\alpha}} = \frac{\alpha t^2 + 2\alpha^2\sigma^4\omega t - \alpha\sigma^4\omega^2}{2(1 + \alpha^2\sigma^4)},$$

from where $t_{ins} = \dfrac{t + \alpha\sigma^4\omega}{1 + \alpha^2\sigma^4}$, $\omega_{ins} = \alpha t_{ins}$.

Gaussian-Enveloped Tone. $x(t) = \exp(-(t - t_0)^2/2\lambda^2 + i\omega_0(t - t_0))$, then the STFT has support on an ellipsoid centered at (t_0, ω_0) with temporal width $\sqrt{\sigma^2 + \lambda^2}$ and frequency width $\sqrt{\sigma^{-2} + \lambda^{-2}}$; the total area of the support is $(\sigma^2 + \lambda^2)/\sigma\lambda$, which is bounded by below by 2 and becomes infinite for either clicks or tones. The instantaneous estimates are

$$t_{ins} = t_0 + \frac{\lambda^2}{\sigma^2 + \lambda^2}(t - t_0), \quad \omega_{ins} = \omega_0 + \frac{\sigma^2}{\sigma^2 + \lambda^2}(\omega - \omega_0),$$

from where the two limits, $\lambda \to \infty$ and $\lambda \to 0$ give the first two columns of Table 1, respectively. The support of the reassigned sonogram has temporal width $\lambda^2/\sqrt{\sigma^2 + \lambda^2}$ and frequency width $(\sigma/\lambda)/\sqrt{\sigma^2 + \lambda^2}$, so the reassigned representation is tone-like when $\lambda > \sigma$ (i.e., the representation contracts the frequency axis more than the time axis) and click-like when $\lambda > \sigma$ (the time direction is contracted more than the frequency). The area of the support has become the reciprocal of the STFT's, $\sigma\lambda/(\sigma^2 + \lambda^2)$, whose maximum is $1/2$ when $\sigma = \lambda$ (i.e., when the signal matches the analyzing wavelet).

We thank Dan Margoliash, Sara Solla, David Gross, and William Bialek for discussions shaping these ideas; A. James Hudspeth, Carlos Brody, Tony Zador, Partha Mitra, and the members of our research group for spirited discussions and critical comments; and Fernando Nottebohm and Michale Fee for support of T.J.G. This work was supported in part by National Institutes of Health Grant R01-DC007294 (to M.O.M.) and a Career Award at the Scientific Interface from the Burroughs-Wellcome Fund (to T.J.G.).

1. Wigner, E. P. (1932) *Phys. Rev*. **40,** 749–759.
2. Lee, H. W. (1995) *Phys. Rep.* **259,** 147–211.
3. Cohen, L. (1995) *Time-Frequency Analysis* (Prentice–Hall, Englewood Cliffs, NJ).
4. Korsch, H. J., Muller, C. & Wiescher, H. (1997) *J. Phys. A* **30,** L677–L684.
5. Wiescher, H. & Korsch, H. J. (1997) *J. Phys. A* **30,** 1763–1773.
6. Cohen, L. (1989) *Proc. IEEE* **77,** 941–981.
7. Hogan, J. A. & Lakey, J. D. (2005) *Time-Frequency and Time-Scale Methods: Adaptive Decompositions, Uncertainty Principles, and Sampling* (Birkhauser, Boston).
8. Greenewalt, C. H. (1968) *Bird Song: Acoustics and Physiology* (Smithsonian Institution, Washington, DC).
9. Margoliash, D. (1983) *J. Neurosci.* **3,** 1039–1057.
10. Chen, V. C. & Ling, H. (2002) *Time-Frequency Transforms for Radar Imaging and Signal Analysis* (Artech House, Boston, MA).
11. Riley, M. D. (1989) *Speech Time-Frequency Representations* (Kluwer, Boston).
12. Fulop, S. A., Ladefoged, P., Liu, F. & Vossen, R. (2003) *Phonetica* **60,** 231–260.
13. Smutny, J. & Pazdera, L. (2004) *Insight* **46,** 612–615.
14. Steeghs, P., Baraniuk, R. & Odegard, J. (2002) in *Applications in Time-Frequency Signal Processing*, ed. Papandreou-Suppappola, A. (CRC, Boca Raton, FL), pp. 307–338.
15. Vasudevan, K. & Cook, F. A. (2001) *Can. J. Earth Sci.* **38,** 1027–1035.
16. Trebino, R., DeLong, K. W., Fittinghoff, D. N., Sweetser, J. N., Krumbugel, M. A., Richman, B. A. & Kane, D. J. (1997) *Rev. Sci. Instrum.* **68,** 3277–3295.
17. Hase, M., Kitajima, M., Constantinescu, A. M. & Petek, H. (2003) *Nature* **426,** 51–54.
18. Marian, A., Stowe, M. C., Lawall, J. R., Felinto, D. & Ye, J. (2004) *Science* **306,** 2063–2068.
19. Gabor, D. (1946) *J. IEE (London)* **93,** 429–457.
20. Choi, H. I. & Williams, W. J. (1989) *IEEE Trans. Acoustics Speech Signal Processing* **37,** 862–871.
21. Thomson, D. J. (1982) *Proc. IEEE* **70,** 1055–1096.
22. Slepian, D. & Pollak, H. O. (1961) *Bell System Tech. J.* **40,** 43-63.
23. Tchernichovski, O., Nottebohm, F., Ho, C. E., Pesaran, B. & Mitra, P. P. (2000) *Anim. Behav.* **59,** 1167–1176.
24. Mitra, P. P. & Pesaran, B. (1999) *Biophys. J.* **76,** 691–708.
25. Huang, N. E., Shen, Z., Long, S. R., Wu, M. L. C., Shih, H. H., Zheng, Q. N., Yen, N. C., Tung, C. C. & Liu, H. H. (1998) *Proc. R. Soc. London Ser. A* **454,** 903–995.
26. Yang, Z. H., Huang, D. R. & Yang, L. H. (2004) *Adv. Biometric Person Authentication Proc.* **3338,** 586–593.
27. Kodera, K., Gendrin, R. & Villedary, C. D. (1978) *IEEE Trans. Acoustics Speech Signal Processing* **26,** 64–76.
28. Auger, F. & Flandrin, P. (1995) *IEEE Trans. Signal Processing* **43,** 1068–1089.
29. ChassandeMottin, E., Daubechies, I., Auger, F. & Flandrin, P. (1997) *IEEE Signal Processing Lett.* **4,** 293–294.
30. Chassande-Mottin, E., Flandrin, P. & Auger, F. (1998) *Multidimensional Systems Signal Processing* **9,** 355–362.
31. Gardner, T. J. & Magnasco, M. O. (2005) *J. Acoust. Soc. Am.* **117,** 2896–2903.
32. Nelson, D. J. (2001) *J. Acoust. Soc. Am.* **110,** 2575–2592.
33. Licklider, J. C. R. (1951) *Experientia* **7,** 128-134.
34. Patterson, R. D. (1987) *J. Acoust. Soc. Am.* **82,** 1560–1586.
35. Cariani, P. A. & Delgutte, B. (1996) *J. Neurophysiol.* **76,** 1698–1716.
36. Cariani, P. A. & Delgutte, B. (1996) *J. Neurophysiol.* **76,** 1717–1734.
37. Julicher, F., Andor, D. & Duke, T. (2001) *Proc. Natl. Acad. Sci. USA* **98,** 9080–9085.
38. Flandrin, P. (1999) *Time-Frequency/Time-Scale Analysis* (Academic, San Diego).
39. Hopfield, J. J. (1995) *Nature* **376,** 33–36.
40. Hopfield, J. J. & Brody, C. D. (2001) *Proc. Natl. Acad. Sci. USA* **98,** 1282–1287.
41. Hahnloser, R. H. R., Kozhevnikov, A. A. & Fee, M. S. (2002) *Nature* **419,** 65–70.
42. Olshausen, B. A. & Field, D. J. (2004) *Curr. Opin. Neurobiol.* **14,** 481–487.
43. Olshausen, B. A. & Field, D. J. (1996) *Nature* **381,** 607–609.
44. Coleman, M. J. & Mooney, R. (2004) *J. Neurosci.* **24,** 7251–7265.
45. DeWeese, M. R., Wehr, M. & Zador, A. M. (2003) *J. Neurosci.* **23,** 7940–7949.
46. Zador, A. (1999) *Neuron* **23,** 198–200.
47. Elhilali, M., Fritz, J. B., Klein, D. J., Simon, J. Z. & Shamma, S. A. (2004) *J. Neurosci.* **24,** 1159–1172.