FREQUENCY REASSIGNMENT FOR COHERENT MODULATION FILTERING

Steven M. Schimmel, Kelly R. Fitz**, and Les E. Atlas**
*Department of Electrical Engineering, University of Washington,
Seattle WA, 98195-2500, USA
**School of Electrical Engineering and Computer Science, MC 2752,
Washington State University, Pullman WA, 99164-2752, USA

ABSTRACT

Modulation filtering is a technique for filtering slowly-varying envelopes of frequency subbands of a signal, without affecting the signal's phase and fine-structure. Coherent modulation filtering is a promising subtype of such techniques where subband envelopes are determined through demodulation of the subband signal with a coherently detected subband carrier. In this paper we propose a coherent modulation filtering technique that detects the carriers using the frequency reassignment (FR) operator from time-frequency reassignment. We show how this technique avoids the use of finite differences in the computation of instantaneous frequency (IF), and that it estimates IF more accurately than a past technique as a result. We confirm that the FR-enhanced technique retains the desirable modulation filtering properties (superposition and the preservation of zero-crossings) and show that it performs better on the same single-channel music source separation task than the past technique.

1. INTRODUCTION

Modulation filtering is a meaningful type of filtering for many natural signals, for example speech and music, because such signals can often be represented by low frequency modulators which modulate higher frequency carriers. Besides modifying signals, the so-called "modulation frequency" concept is also useful for describing and representing broadband acoustic signals. Modulation frequency representations, commonly referred to as modulation spectrograms, usually consist of a transform of a one-dimensional broadband signal into a two-dimensional joint-frequency representation, where one dimension is standard Fourier frequency and the other dimension is modulation frequency. Modulation analysis and modulation filtering techniques have been used in a wide range of applications, e.g. in music source separation [1], audio encoding and audio compression [2], and psychoacoustic research [3].

One of the most critical components of modulation filtering techniques is the separation of the frequency subband signals into their slowly-varying envelope signals and fine-structure carrier signals. The best technique for carrier detection known to us to date is the instantaneous frequency (IF) estimation technique proposed by Atlas and Janssen [1]. Careful analysis of this technique, however, has revealed a fundamental weakness in its IF estimator. In this paper we present an improved estimator that overcomes this weakness by incorporating some of the theory of *time-frequency reassignment*.

Time-frequency reassignment is a technique for refocusing time-frequency data in a blurred representation like the spectrogram by mapping the time-frequency data to coordinates that are nearer to the true region of support of the analyzed signal [4,5]. Pioneering work on the method of time-frequency reassignment was first published by Kodera, Gendrin, and de Villedary [6]. The method of time-frequency reassignment has been used in a variety of applications for obtaining improved time and/or frequency estimates for time-varying spectral data [7-10].

In the improved modulation transform that we propose, we use only the frequency reassignment (FR) operator from time-frequency reassignment as the carrier detector. As our results demonstrate, this FR-enhanced modulation transform estimates the instantaneous frequency of both synthetic and natural signals more accurately, while still retaining all desirable modulation filtering properties.

The improvement to the IF estimator is relevant to all of the modulation filtering applications mentioned earlier. Audio coding improves because more of a signal's energy will be represented in low modulation frequencies, resulting in a sparser representation. Music source separation and psychoacoustic research benefit because FR-enhanced modulation filtering is more accurate and creates less artifacts.

In section 2 of this paper we present modulation transforms in a general framework, and introduce the necessary definitions and notation to describe modulations transforms and time-frequency reassignment. In section 3 we then analyze and demonstrate the fundamental weakness of the IF estimator proposed in [1]. In section 4 we give an overview of the time-frequency reassignment technique, and show how it overcomes this fundamental weakness. In section 5 we present our results, followed by our conclusions and a discussion in section 6.

2. BACKGROUND

Figure 1 shows a general framework for modulation transforms. A broadband signal $x(t)$ is decomposed into frequency subbands by means of a filter bank that is characterized by a set of band-pass filters $h_\omega(t)$,

$$x_\omega(t) = x(t) * h_\omega(t) \, . \tag{1}$$

The subband signals $x_\omega(t)$ are further separated into an envelope signal $a_\omega(t)$ and a carrier signal $c_\omega(t)$. The envelope signals are subsequently filtered by a linear time-invariant filter $g(t)$,

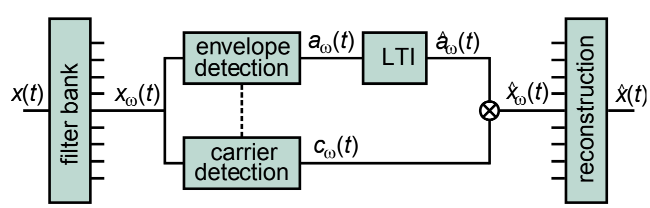

Figure 1 General framework for modulation transforms

$\hat{a}_\omega(t) = a_\omega(t) * g(t)$, and the modified envelopes are recombined with the (unmodified) carriers to form the modified subbands, $\hat{x}_\omega(t) = \hat{a}_\omega(t) c_\omega(t)$. Finally, the frequency subbands are reconstructed into the modulation filtered signal $\hat{x}(t)$.

The filter bank's band-pass filters are often obtained by uniformly modulating a (real-valued) prototype low-pass filter $h(t)$ to the desired frequencies,

$$h_\omega(t) = h(t) e^{j\omega t}, \tag{2}$$

such that the filter bank can be implemented using the short-time Fourier transform as follows:

$$\varepsilon(t,\omega) = \int x(\tau) h(t-\tau) e^{j\omega[t-\tau]} d\tau \tag{3}$$

$$= e^{j\omega t} \int x(\tau) h(t-\tau) e^{-j\omega\tau} d\tau \tag{4}$$

$$= e^{j\omega t} X(t,\omega). \tag{5}$$

Note that the coefficients $\varepsilon(t,\omega) = A(t,\omega)\exp[j\psi(t,\omega)]$ differ from the short-time Fourier transform coefficients $X(t,\omega) = A(t,\omega)\exp[j\phi(t,\omega)]$ only by a phase offset, that is, $\psi(t,\omega) = \phi(t,\omega) + \omega t$. With this type of filter bank, the subband signals $x_\omega(t)$ are defined as $x_\omega(t) = \varepsilon(t,\omega)$. For our discussion it is also convenient to define the discrete versions of both transform coefficients as

$$\varepsilon(n,k) = \varepsilon(nT, k\Omega), \tag{6}$$

$$X(n,k) = X(nT, k\Omega), \tag{7}$$

where T and Ω are the sampling period and the subband frequency spacing of the transforms, respectively.

For modulation analysis and filtering, we distinguish between two types of envelope and carrier detection. In the first type, called incoherent modulation filtering, the envelope and carrier of a subband are detected independently, for example as the magnitude and phase of the subband [2], or as the Hilbert envelope and carrier [3]. In the second type, called coherent modulation filtering, the subband envelope is obtained by demodulating the subband signal with a coherently detected subband carrier [1,11]. The coherent technique is preferred over the incoherent technique for most applications, as it is believed to be a better fit to natural signals [12], and has better filtering properties with less artifacts [11]. The envelope and carrier detection operation of coherent modulation filtering approaches can be expressed as

$$c_\omega(t) = \exp\left[\mathrm{j}\int_0^{\mathrm{t}} \mathrm{IF}\{x_\omega(\tau)\} d\tau\right] \tag{8}$$

$$a_\omega(t) = x_\omega(t) \cdot c_\omega^*(t) \tag{9}$$

where * denotes complex conjugation, and $\mathrm{IF}\{\cdot\}$ is an operation that estimates the instantaneous frequency of its argument. In this paper, we focus on the IF estimator that is used in the coherent modulation filtering technique proposed by Atlas and Janssen [1].

3. INSTANTANEOUS FREQUENCY ESTIMATION

Atlas and Janssen [1] define their IF estimator as follows. Given the discrete subband coefficients from (6) in polar form,

$$\varepsilon(n,k) = A(n,k) e^{j\psi(n,k)}, \tag{10}$$

the intermediate instantaneous phase signal $z(n,k)$ of a signal $x(t)$ is determined by

$$z(n,k) = z_i(n,k) + j z_q(n,k), \tag{11}$$

where the in-phase and quadrature parts are defined as

$$z_i(n,k) = i(n-1,k) i(n+1,k) + q(n-1,k) q(n+1,k), \tag{12}$$

$$z_q(n,k) = i(n-1,k) q(n+1,k) - q(n-1,k) i(n+1,k), \tag{13}$$

with

$$i(n,k) = \mathrm{Re}\{\varepsilon(n,k)\}, \tag{14}$$

$$q(n,k) = \mathrm{Im}\{\varepsilon(n,k)\}. \tag{15}$$

The instantaneous frequency estimate is obtained from the instantaneous phase signal through

$$\alpha(n,k) = z(n,k) / |z(n,k)|. \tag{16}$$

Using the definitions in (10)–(15) and some trigonometric identities, this expression can be simplified to

$$\alpha(n,k) = \exp\{j[\psi(n+1,k) - \psi(n-1,k)]\}. \tag{17}$$

In other words, the IF estimator used in [1] is equivalent to the finite central phase difference at each point in the discrete short-time Fourier transform.

Normally, a signal's instantaneous frequency is computed from signal phase, rather than from subband phase as in (17), and is defined for an analytic signal $x(t) = A(t)\exp[j\theta(t)]$ as

$$\omega(t) = \frac{d\theta(t)}{dt}. \tag{18}$$

Kodera *et al.* [6] showed that the instantaneous frequency of a subband signal is equal to the partial derivative with respect to time of the subband phase, i.e.

$$\alpha(t,\omega) = \frac{\partial\psi(t,\omega)}{\partial t}. \tag{19}$$

Comparing (17) with this expression, it is clear that the finite difference used in the IF estimator by Atlas and Janssen is only a linear approximation to the true partial derivative, and that its accuracy depends on the sampling period T. By adopting the frequency reassignment operator from time-frequency reassignment, however, we are able to find the partial derivative of subband phase without resorting to finite approximations.

4. TIME-FREQUENCY REASSIGNMENT

A time-domain signal $x(t)$ can be reconstructed from the coefficients $\varepsilon(t,\omega)$ in the filter-bank decomposition given in (3) by

$$x(t) = \iint \varepsilon(\tau,\omega) h(\tau - t) e^{-j\omega[\tau - t]} d\omega d\tau \tag{20}$$

$$= \iint A(\tau,\omega) h(\tau - t) e^{j[\psi(\tau,\omega) - \omega\tau + \omega t]} d\omega d\tau, \tag{21}$$

where $h(t)$ is a (real-valued) low-pass kernel function, or analysis window. For signals having magnitude spectra $A(t,\omega)$, whose time variation is slow relative to the phase variation, the maximum contribution to the reconstruction integral comes from the vicinity of the point t,ω satisfying the phase stationarity condition

$$\frac{\partial}{\partial\omega}[\psi(\tau,\omega) - \omega\tau + \omega t] = 0, \tag{22}$$

$$\frac{\partial}{\partial\tau}[\psi(\tau,\omega) - \omega\tau + \omega t] = 0, \tag{23}$$

or equivalently, around the point $\hat{t},\hat{\omega}$ defined by

$$\hat{t}(\tau,\omega) = \tau - \frac{\partial\psi(\tau,\omega)}{\partial\omega}, \tag{24}$$

$$\hat{\omega}(\tau,\omega) = \frac{\partial\psi(\tau,\omega)}{\partial\tau}. \tag{25}$$

This phenomenon has long been known in such fields as optics as the *principle of stationary phase* (see for example [13]). The method of time-frequency reassignment changes, or reassigns, the point of attribution of $X(t,\omega)$ from t,ω to this point of maximum contribution $\hat{t}(t,\omega),\hat{\omega}(t,\omega)$, sometimes called the "center of gravity" of the distribution, by way of analogy to a mass distribution. The reassigned coordinates in (24) and (25) can be computed by

$$\hat{t}(t,\omega) = t - \mathrm{Re}\left\{\frac{X_{Th}(t,\omega)\cdot X^*(t,\omega)}{|X(t,\omega)|^2}\right\}, \tag{26}$$

$$\hat{\omega}(t,\omega) = \omega + \mathrm{Im}\left\{\frac{X_{Dh}(t,\omega)\cdot X^*(t,\omega)}{|X(t,\omega)|^2}\right\}, \tag{27}$$

where $X(t,\omega)$, $X_{Th}(t,\omega)$ and $X_{Dh}(t,\omega)$ are short-time Fourier transforms of the signal $x(t)$ that are computed using an analysis window $h(t)$, a time-weighted analysis window $h_T(t) = t\cdot h(t)$, and a time-derivative analysis window $h_D(t) = \frac{d}{dt}h(t)$ respectively [2].

Using the auxiliary window functions $h_T(t)$ and $h_D(t)$, the reassignment operations can be computed at any time-frequency coordinate t,ω from an algebraic combination of the values of three Fourier transforms evaluated at t,ω, without directly evaluating or approximating the partial derivatives of phase. Since this algorithm operates only on spectral data evaluated at a single time and frequency, and does not explicitly compute any derivatives, it can easily be implemented in digital systems using discrete times and frequencies.

The time-weighted window function, $h_T(t)$, is trivially computed by point-wise multiplication of the original window function $h(t)$ by a time ramp. In discrete time,

$$h_T(n) = n\cdot h(n)\,. \tag{28}$$

If the derivative of the window function is unknown, then $h_D(t)$ can also be computed numerically. The derivative theorem for Fourier transforms states that if $X(\omega) = \mathrm{FT}\{x(t)\}$ then $j\omega X(\omega) = \mathrm{FT}\left\{\frac{d}{dt}x(t)\right\}$. We can therefore construct the time-derivative window used in the evaluation of the frequency reassignment operator by computing the Fourier transform of $h(t)$, multiplying by $j\omega$, and inverting the Fourier transform. That is,

$$\frac{d}{dt}h(t) = \mathrm{FT}^{-1}\left\{j\omega H(\omega)\right\} = -\mathrm{Im}\left\{\mathrm{FT}^{-1}\left\{\omega H(\omega)\right\}\right\}, \tag{29}$$

and so, in discrete time,

$$h_D(n) = -\mathrm{Im}\left\{\mathrm{FT}^{-1}\left\{\frac{2\pi k}{N}H(k)\right\}\right\}. \tag{30}$$

An example of the auxiliary short-time analysis windows employed in our computation of the reassignment operations are shown in Figure 2. To adapt the frequency reassignment operator for use in our modulation transform, it is converted to a carrier signal,

$$c_\omega(t) = \exp\left[j\int_0^t \hat{\omega}(\tau,\omega)d\tau\right]. \tag{31}$$

5. RESULTS

To illustrate the improved IF estimates that are obtained with the IF estimator based on the frequency reassignment operator, Figure 3 shows the IF estimates of both methods for a linear chirp. The improved estimator clearly performs better, since the subband

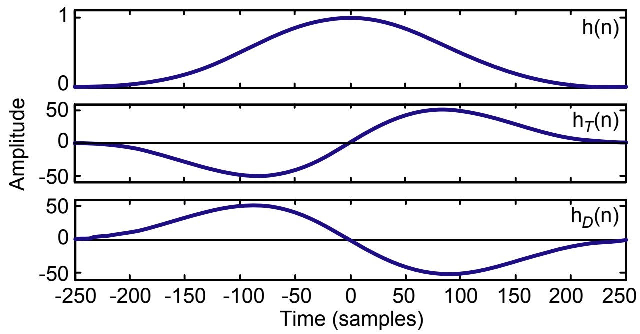

Figure 2 Representative analysis windows employed in the three short-time transforms used to compute reassigned times and frequencies. (*top*) The original window function $h(n)$ (a Kaiser window with shaping parameter $\beta = 9$); (*middle*) The time-weighted window function $h_T(n)$; (*bottom*) The time-derivative window function $h_D(n)$.

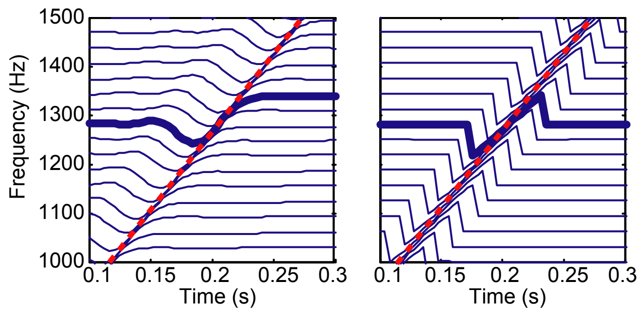

Figure 3 Instantaneous frequency estimates for subbands of a linear chirp signal. Each line (blue) indicates the estimated IF for one subband. One subband is highlighted. The dotted line (red) indicates the signal's true IF. (*left*) IF estimator proposed by Atlas and Janssen; (*right*) Frequency reassignment based IF estimator.

IF estimates adapt faster to the true instantaneous frequency, and follow it more accurately.

The improved modulation filtering system was also applied to the same test signals as in [1] to confirm that it preserves the two desirable properties for IF estimation and modulation filtering: superposition and preservation of zero-crossings. The left panel in Figure 4 shows the effect of low-pass modulation filtering on a signal's overall envelope. A low-pass modulation filter was applied to each subband independently, and since the new system still satisfies the superposition property, the overall envelope has also been low-pass filtered as a result.

To demonstrate that the new modulation filtering system also preserves zero-crossings, the right panel in Figure 4 shows a very short segment of the signals on the left. On this time-scale, the fine-structure of the signals is visible. It can be seen that the zero-crossings of the overall signal are not affected by modulation filtering and signal reconstruction.

To compare the performance of the FR-enhanced modulation filtering system to the system defined by Atlas and Janssen, we have applied our system to the same single channel music source separation task as in [1], involving a mix of a castanets and a flute signal. Figure 5 shows the spectrograms of the original and sepa-

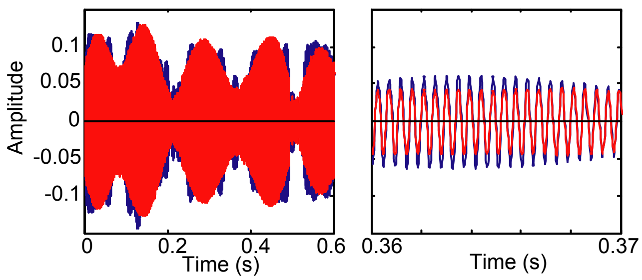

Figure 4 Flute signal before (blue, or black in b/w) and after (red, or grey in b/w) low-pass modulation filtering. (*left*) The overall envelope has also been low-passed; (*right*) Zero-crossings are preserved.

rated signals. We have found that our system achieves 10–20 dB signal separation, which is up to 5 dB more than reported in [1].

6. CONCLUSIONS AND DISCUSSION

We have identified that the IF estimator proposed by Atlas and Janssen uses a linear approximation of the partial derivative of subband phase. We have demonstrated that this linear approximation causes inaccuracies in the estimated IF. Furthermore, we have shown how the frequency reassignment operator from time-frequency reassignment can be adapted for use in our modulation transform, and how it improves the IF estimates. Our experiments confirm that the FR-enhanced modulation transform still preserves zero-crossings and satisfies superposition, as expected. It also performs better than Atlas' and Janssen's modulation transform on the same single channel music source separation task.

The Kaiser window that we used to compute the frequency reassignment operator worked well in the experiments we have presented here, but its effect on modulation filtering effectiveness in general needs further study.

Another issue that needs further study is the phase-shift that off-center frequencies get from the window. As components of a signal move through a subband, they observe a linear phase-shift from the analysis window that depends on their distance to the subband's center frequency. These shifts cause small distortions in the instantaneous frequency estimation, but they can be corrected for because the phase response of the window is known.

7. ACKNOWLEDGEMENTS

The authors wish to acknowledge discussions with Alain de Cheveigné of Ircam-CNRS in France and Bishnu Atal of the University of Washington.

8. REFERENCES

[1] L. Atlas and C. Janssen, "Coherent modulation spectral filtering for single-channel music source separation," Proceedings of ICASSP, 2005, pp. 461–464.

[2] M. S. Vinton and L. E. Atlas, "Scalable and progressive audio codec," Proceedings of ICASSP, 2001, pp. 3277–3280.

[3] R. Drullman, J. M. Festen, and R. Plomp, "Effect of temporal envelope smearing on speech reception," *Journal of the Acoustical Society of America*, vol. 95, pp. 1053–1064, 1994.

[4] P. Flandrin, F. Auger, and E. Chassande-Mottin, "Time-Frequency Reassignment: From Principles to Algorithms," in *Applications in Time-Frequency Signal Processing*, P.-S. Antonia, Ed., 2003, pp. 179-203.

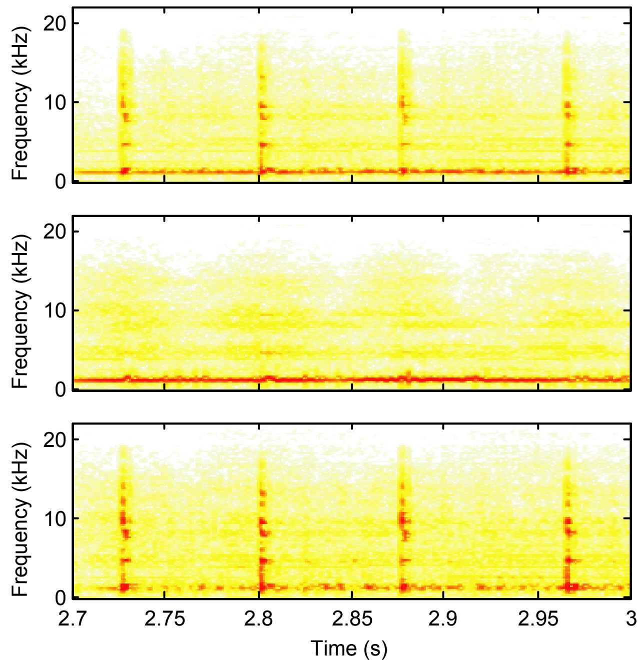

Figure 5 Single channel music source separation. (*top*) Mix of castanets and flute; (*middle*) Flute only signal; (*bottom*) Castanets only signal.

[5] F. Auger and P. Flandrin, "Improving the readability of time-frequency and time-scale representations by the reassignment method," *IEEE Transactions on Signal Processing*, vol. 43, pp. 1068–1089, 1995.

[6] K. Kodera, R. Gendrin, and C. de Villedary, "Analysis of Time-Varying Signals with Small BT Values," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. ASSP-26, pp. 64–76, 1978.

[7] F. Plante, G. Meyer, and W. A. Ainsworth, "Improvement of Speech Spectrogram Accuracy by the Method of Spectral Reassignment," *IEEE Transactions on Speech and Audio Processing*, vol. 6, pp. 282–287, 1998.

[8] S. W. Hainsworth, M. D. Macleod, and P. J. Wolfe, "Analysis of reassigned spectrograms for musical transcription," in *IEEE Workshop on the Applications of Signal Processing to Audio and Acoustics*, 2001, pp. 23–26.

[9] K. Fitz and L. Haken, "On the Use of Time-Frequency Reassignment in Additive Sound Modeling," *Journal of the Audio Engineering Society*, vol. 50, pp. 879–893, 2002.

[10] T. J. Gardner and M. O. Magnasco, "Instantaneous frequency decomposition: an application to spectrally sparse sounds with fast frequency modulations," *Journal of the Acoustical Society of America*, vol. 117, pp. 2896–2903, 2005.

[11] S. M. Schimmel and L. E. Atlas, "Coherent envelope detection for modulation filtering of speech," Proceedings of ICASSP, 2005, pp. 221–224.

[12] L. Atlas, Q. Li, and J. Thompson, "Homomorphic modulation spectra," Proceedings of ICASSP, 2004, pp. 761–764.

[13] A. Papoulis, *Systems and Transforms with Applications to Optics*, 1968.