A SUM-OF-PRODUCTS MODEL FOR EFFECTIVE COHERENT MODULATION FILTERING

Pascal Clark and Les Atlas

Department of Electrical Engineering, University of Washington,
Seattle WA, 98195-2500, USA
{clarkcp, atlas}@u.washington.edu

ABSTRACT

Modulation filtering is a technique for filtering slowly-varying envelopes of frequency subbands of a nonstationary signal, ideally 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 demonstrate how modulation filtering, when done coherently, is far more effective than standard incoherent methods. We show that empirical results can be made to be almost ideal, and significantly better than previous coherent attempts, as long as fine-structure information is retained as side information and the filterbank reduces subband interference.

Index Terms— modulation, time-varying filters, acoustic signal processing, speech processing

1. INTRODUCTION

Modulation filtering is potentially a useful type of filtering for many natural and machine-made signals because such signals can often be represented by low frequency modulators which modulate higher frequency carriers. 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 extensively in perceptual speech research (e.g. [3, 4]).

Two critical components of modulation filtering techniques are 1) the separation of the broadband signal into frequency subbands and 2) the subsequent separation of each subband signal into its slowly-varying envelope signal and fine-structure carrier signal. Thus the joint-frequency representation arises from an implicit sum-of-products signal model. For modulation filtering, considerable attention has focused on distortion effects caused by the subband product model, starting with the work of Ghitza [5] and later addressed with a partial solution by Schimmel and Atlas [6] using a form of coherent demodulation. However, [6] did not address the full sum-of-products model, including both the "sum" and "product" parts, which [5] touched upon within the context of overlapping cochlear filters.

This research was partially supported by AFOSR Grant FA95500610191.

In this paper, we show that modulation filtering, when done coherently, can perform considerably better than previously demonstrated in [6]. We achieve nearly ideal results with the use of bandwidth and recovery constraints derived in [7], as well as a careful filterbank design that isolates adjacent subbands. These two developments lead to new qualifications on the overall sum-of-products signal model for effective coherent modulation filtering.

2. MODULATION FILTERING BACKGROUND

For a real-valued signal $x[n]$ we assume the following sum-of-products signal model

$$x[n] = \sum_{k=0}^{K-1} s_k[n] = \sum_{k=0}^{K-1} m_k[n] \cdot c_k[n] \tag{1}$$

where $s_k[n]$ is the kth analytic subband signal from a K-channel filterbank. Each subband is then demodulated into a high-frequency carrier $c_k[n]$, containing temporal fine structure, and a low-frequency modulator $m_k[n]$, containing temporal envelope information.

Strictly speaking, $s_k[n]$ need not be subband signals. We adopt the filterbank convention in order to remain consistent with the methodologies of [3] and [5], wherein bandpass filters approximated the cochlear filterbank in the human ear.

After decomposing $x[n]$ according to (1), modulation filtering is the act of convolving each $m_k[n]$ with a kernel $g[n]$ to yield $\hat{m}_k[n]$, which is then recombined with the original carrier $c_k[n]$. The final modulation-filtered signal is synthesized as

$$\begin{aligned} \hat{x}[n] &= \sum_{k=0}^{K-1} \hat{m}_k[n] \cdot c_k[n] \\ &= \sum_{k=0}^{K-1} (g[n] * m_k[n]) \cdot c_k[n]. \end{aligned} \tag{2}$$

To measure the performance of a modulation filtering system, Ghitza [5] devised the test that appears in block-diagram form in Fig. 1. After forming $\hat{x}[n]$, we reapply the same demodulation algorithm and compare each recovered modulator $\tilde{m}_k[n]$ to the corresponding $\hat{m}_k[n]$. In other words, we should be able to perfectly recover the modified components that went into constructing $\hat{x}[n]$. Any resulting discrepancy is regarded as artifact introduced by a flaw in the modulation decomposition.

Schimmel [6] later formalized Ghitza's test by computing the empirical modulation frequency response (EMFR) as

$$G_e(\omega) = \frac{1}{K} \sum_{k=0}^{K-1} \left| \frac{\tilde{M}_k(\omega)}{M_k(\omega)} \right| \tag{3}$$

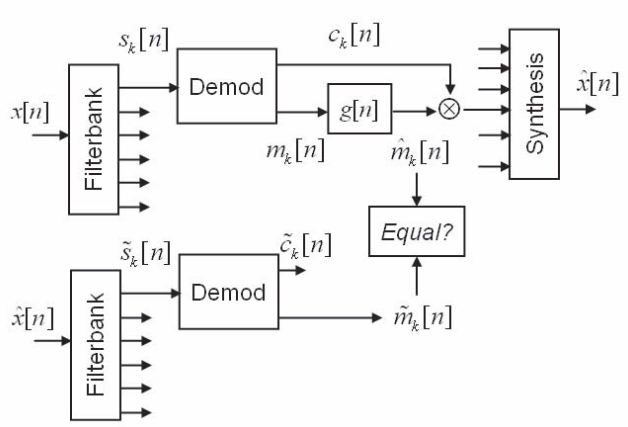

Fig. 1. System diagram for testing modulation filtering efficacy. where the desired outcome is for the recovered modulator to equal the filtered modulator used in synthesis.

which directly computes the change in spectral power between the original and the recovered modulators. If $G(\omega)$ is the frequency response of $g[n]$, then ideally $G_e(\omega) = G(\omega)$. As both Ghitza and Schimmel discovered, however, $G_e(\omega)$ depends heavily upon the methods of the chosen modulation decomposition [5, 6].

The rest of this paper will use the empirical modulation transfer function as a measure of performance. As detailed over the next several sections, we show that a well-chosen modulation decomposition, plus a carefully designed filterbank, ensures that $G_e(\omega)$ is essentially equivalent to $G(\omega)$.

3. SUBBAND DEMODULATION

In Fig. 1, the "Demod(ulate)" block consists of two steps. First, we define a carrier detection operator $\mathcal{D}$ such that

$$\mathcal{D}\{s_k[n]\} = c_k[n] = \exp(j\phi_k[n]). \tag{4}$$

Consequently, the envelope is given by

$$m_k[n] = s_k[n] \cdot c_k[n]^* \tag{5}$$

where superscript * denotes complex conjugation. We should be clear in stating that (5) allows the envelope to be complex-valued, depending on the choice for $c_k[n]$.

Another important result from demodulation is the instantaneous frequency (IF) of a subband, defined in continuous time as

$$\omega_k(t) = \frac{d}{dt}\phi_k(t). \tag{6}$$

In practice, we compute a discrete-time version of (6) by applying a derivative-approximating, finite-impulse response filter to the phase signal $\phi_k[n]$.

At this point, the specifics of demodulation remain undetermined. In fact, there are an infinite number of ways to factor a given $s_k[n]$ into a modulator and carrier. Broadly speaking, demodulation methods can be categorized as either *coherent* or *incoherent*. We explain this distinction in the following subsections along with an example of each.

3.1. Incoherent Demodulation (Hilbert Envelope)

A common method of incoherent demodulation is to separate magnitude and phase, which defines $\phi_k[n] = \angle s_k[n]$ so that

$$m_k[n] = |s_k[n]|, \quad c_k[n] = \exp(j\angle s_k[n]). \tag{7}$$

In this case, $m_k[n]$ is called the Hilbert envelope of $s_k[n]$ and is restricted to be real-valued and non-negative. Note that (7) is consistent with (4) and (5) yet does not require an explicit carrier estimate in order to find the modulator. For this reason the Hilbert envelope is a form of *incoherent demodulation*.

The drawbacks of the Hilbert envelope are well-documented. Most prominently, neither $m_k[n]$ nor $c_k[n]$ are necessarily bandlimited [6], and the IF often extends beyond the bandwidth of $s_k[n]$ and can even contain infinite spikes [8, 9]. Nevertheless, the Hilbert envelope is standard in speech studies [3, 4] and will represent the baseline for this paper.

3.2. Coherent Demodulation (Complex Envelope)

In contrast to incoherent demodulation, coherent demodulation defines the modulator solely in terms of a detected carrier and the relationship in (5). As a result, the coherent modulator is likely to be complex-valued as $m_k[n]$ adopts any phase that is not captured by the carrier estimate.

For this paper we use a form of coherent demodulation based on spectral center-of-gravity (COG) [10, 7], which estimates the carrier IF in terms of energy concentration in a time-frequency representation of $s_k[n]$. As prescribed by [10], the coherent IF is

$$\omega_k[n] = \frac{\int \omega\, S_k(\omega, n)\, d\omega}{\int S_k(\omega, n)\, d\omega} \tag{8}$$

where $S_k(\omega, n)$ is a short-time power-spectral density estimate of $s_k[n]$ using an analysis window centered on time sample n. In other words, $\omega_k[n]$ is the centroid of the local spectrum in the vicinity of $s_k[n - L/2], ..., s_k[n + L/2]$, where L is the length of the analysis window. In this way, the IF follows the average spectral energy of the subband signal over time. Upon determining the IF, we proceed to form the carrier by integrating $\omega_k[n]$ and using the righthand side of (4), by which the modulator $m_k[n]$ follows from (5).

Another possible form of coherent demodulation appeared originally in [6], and operates by lowpass filtering the unwrapped $\angle s_k[n]$ and placing the residual phase into the modulator. Although achieving a bandlimited decomposition, it is difficult to interpret the effects of smoothing over discontinuities in the Hilbert phase. By comparison, the spectral COG is conceptually more elegant because of its time-frequency interpretation.

4. MODULATION FILTERING A SINGLE SUBBAND

In our treatment of the sum-of-products model, we first focus on the "product" part. As shown in [7], there are two necessary and sufficient conditions on $\mathcal{D}$ for effective, arbitrary modulation filtering of a single subband. To paraphrase, the conditions are:

1. the IF must be bounded and relatively smooth, and
2. the original carrier must be recoverable from the remodulated subband.

The first condition stems from an underlying need to constrain the bandwidth of $\hat{s}_k[n]$ within the bandwidth of $s_k[n]$, as observed

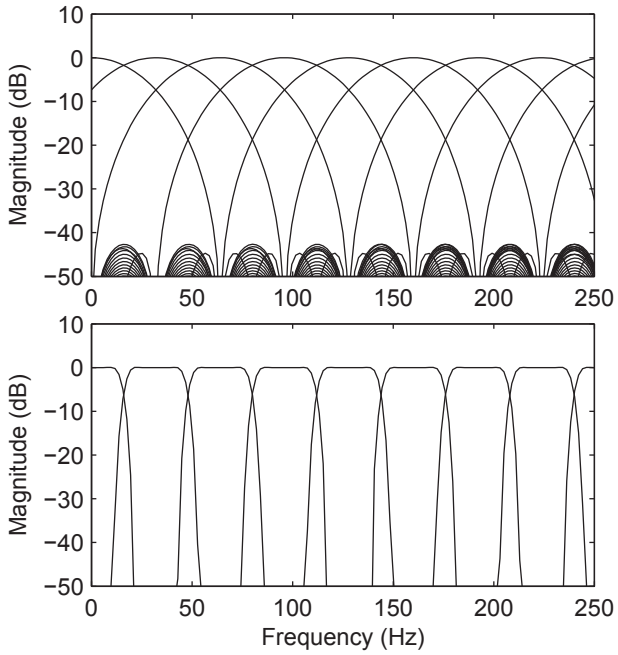

Fig. 2. Filterbank frequency responses for the Hamming-window design with subband overlap (top) and the higher-order non-overlapping design (bottom). Both exhibit subbands spaced 32 Hz apart.

by [6]. Due to the erratic, often discontinuous nature of analytic subband phase, the Hilbert decomposition does not comply with the first condition and therefore fails Ghitza's test. The coherent method, on the other hand, does satisfy the first condition by expressly controlling the bounds of $\omega_k[n]$.

The second condition was not considered by [6] but is just as important, requiring the general recovery property of

$$\mathcal{D}\{\hat{m}_k[n] \cdot c_k[n]\} = c_k[n]. \tag{9}$$

Unfortunately, the spectral COG fails to satisfy the second condition [7]. Roughly speaking, coherent demodulation defines the carrier in a signal-dependent way, so that modifying the signal (via modulation filtering) can result in a different carrier estimate. In practice the difference is small, on the order of Hertz, but can amount to a large discrepancy when observing low-frequency modulator spectra.

Although it may appear to be a concession, the use of side information reminds us that a modulation-filtered signal is modified *with respect to a set of arbitrary carrier estimates*. The perceptual relevance of detected carriers and modulators is certainly a crucial matter, but is beyond the scope of this paper. Instead, we are concerned with establishing mathematical groundwork for modulation filtering, which we argue is essential for understanding the perceptual roles of carrier and modulation frequency. In other words, our hope is that a proper mathematical understanding will lead to a basis for evaluating the relevance of the sum-of-products model.

5. FILTERBANK DESIGN

Relating to the "sum" component of the sum-of-products model, an experimental factor affecting $G_e(\omega)$ is the effect of interference between subbands in the filterbank design. Subbands that overlap in frequency are related by some common information, but modulation-filtering each subband separately can destroy the consensus and lead to interference when the subbands are summed together. This is related to issues in the inversion of modified short-time Fourier transforms [11], and a similar point was noted in [5] for cochlea-emulating filterbanks.

To reduce the amount of subband interference, we adjust frequency-domain overlap between subbands in our filterbank design. For our purpose the short-time Fourier transform (STFT) acts as a multirate filterbank for which a single parameter controls subband overlap.

Consider the STFT definition, given as

$$s_k[n] = \sum_{m=0}^{N-1} x[m]\, h[Rn - m]\, \exp\left(\frac{-j2\pi km}{K}\right) \tag{10}$$

where $h[n]$ is a lowpass window and $k = 0, 1, ..., K-1$. In the above, each $s_k[n]$ is equivalent to a bandpass portion of $x[n]$ that has been frequency-shifted to baseband and then downsampled by a factor of R.

For modulation filtering, we compare two filterbank designs: standard and non-overlapping. There are $K = 250$ subbands in each design, which results in uniformly-spaced subbands 32 Hz apart. In the standard design, $h[n]$ is a 250-point Hamming window with about 2/3 frequency overlap between subbands. This is essentially the same filterbank used in [6]. By contrast, the non-overlapping design uses a 2250-point windowed Dirichlet kernel with -6 dB bandwidth equal to $2\pi/K$. In both cases, the time-decimation rate R is chosen for computational efficiency but is small enough to avoid aliasing the subband mainlobes. Refer to Fig. 2 for frequency-domain plots of the subband filters for each filterbank design.

We should also mention that an STFT filterbank requires a synthesis stage, shown in Fig. 1, to reassemble the subbands and form $\hat{x}_k[n]$. We use the synthesis equations and architecture presented in [12] to design invertible STFT filterbanks.

6. EXPERIMENTAL RESULTS

To assess the performances of coherent and incoherent modulation filtering, we conducted a series of experimental simulations on recorded speech. We compared the Hilbert envelope and spectral COG demodulation under identical conditions: with or without carrier side information, and using the standard or the non-overlapping filterbank. The test signal in each experiment was a 10-second clip of male speech sampled at 8 kHz. The lowpass modulation filter was designed for a 2 Hz cutoff frequency and 40 dB stopband attenuation. For both of the coherent demodulation methods, we set the respective parameters such that the IF signals $\omega_k[n]$ had a bandwidth of approximately 10 Hz.

Fig. 3a shows the EMFRs for both demodulation methods when using the standard filterbank and no carrier side information. This is approximately the same result as in [6], which already displays a marked improvement in the coherent technique over incoherent. Also similar to [6], the empirical responses peak at the location of the next adjacent subband at 32 Hz. We plot only to 16 Hz, though, which is analogous to the Nyquist rate in modulation frequency when the subbands are 32 Hz wide.

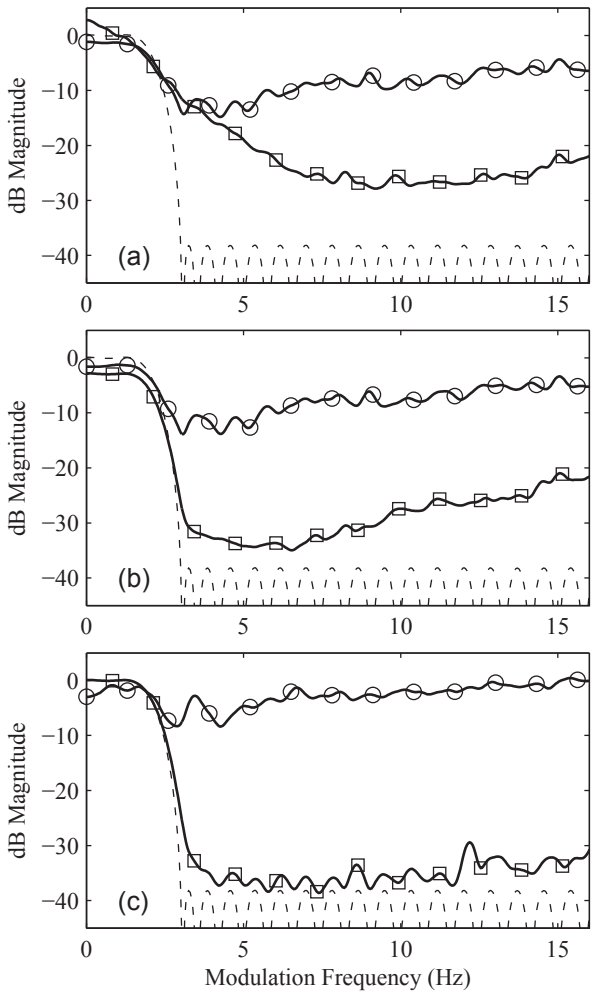

Fig. 3. Empirical modulation frequency responses for Hilbert (circles) and spectral COG (squares) compared to the ideal (dashed) in three experimental setups: standard filterbank and no side info (a); standard filterbank with side-info (b), and non-overlapping filterbank with side info (c).

Fig. 3b repeats the experiment using the same filterbank but with the inclusion of original carrier side information. A dramatic improvement can be seen in the coherent method, with close adherence to the ideal passband and deeper stopband attenuation for low modulation frequencies.

Finally, Fig. 3c shows further improvement after incorporating the non-overlapping filterbank in addition to carrier side information. Using this configuration, the coherent EMFR closely matches the ideal filter response, with roughly 20-25 dB more stopband attenuation than even the best incoherent EMFR in Fig. 2a.

7. CONCLUSION

In this paper we showed that coherent modulation filtering can be more effective than previously believed. As measured by the empirical modulation frequency response, we demonstrated near-ideal performance and 25 dB improvement in the ability to suppress undesired modulation frequencies compared to the standard incoherent technique. This result was made possible by a full assessment of the sum-of-products signal model, where the "sum" component refers to subband summation and the "product" component refers to individual subband modulation. Addressing each component in turn, we employed a filterbank with reduced subband interference while retaining carrier side information to enhance the already beneficial bandwidth-preserving properties of coherent modulation. With these new developments, coherent modulation filtering offers a mathematically consistent framework for measuring the perceptual relevance of modulation frequency in a sum-of-product decomposition, and is a potentially useful new form of signal modification. For more resources and Matlab code, refer to the Modulation Toolbox located at http://isdl.ee.washington.edu/projects/modulationtoolbox/.

The authors wish to thank S. Schimmel of the University of Zurich and Prof. B. Atal of the University of Washington for many helpful insights.

8. REFERENCES

[1] S.M. Schimmel, K.R. Fitz, and L.E. Atlas, "Frequency reassignment for coherent modulation filtering," *Proc. IEEE ICASSP*, vol. 5, pp. 261–264, May 2006.

[2] M.S. Vinton and L.E. Atlas, "Scalable and progressive audio codec," *Proc. IEEE ICASSP 2001*, vol. 5, pp. 3277–3280, 2001.

[3] R. Drullman, J.M. Festen, and R. Plomp, "Effect of temporal envelope smearing on speech reception," *J. Acoust. Soc. Am.*, vol. 95, no. 2, pp. 1053–1064, 1994.

[4] Z.M. Smith, B. Delgutte, and A.J. Oxenham, "Chimaeric sounds reveal dichotomies in auditory perception," *Nature*, vol. 416, pp. 87–90, March 2002.

[5] O. Ghitza, "On the upper cutoff frequency of the auditory critical-band envelope detectors in the context of speech perception," *J. Acoust. Soc. Am.*, vol. 110, no. 3, pp. 1628–1640, 2001.

[6] S. Schimmel and L. Atlas, "Coherent envelope detection for modulation filtering of speech," *Proc. IEEE ICASSP 2005*, vol. 1, pp. 221–224, March 18-23, 2005.

[7] P. Clark and L. Atlas, "Time-frequency coherent modulation filtering of nonstationary signals," *To appear*.

[8] L. Mandel, "Interpretation of instantaneous frequencies," *Am. J. Physics*, vol. 42, no. 10, pp. 840–846, 1974.

[9] L. Cohen, P. Loughlin, and D. Vakman, "On an ambiguity in the definition of the amplitude and phase of a signal," *Elsevier Sig. Process.*, vol. 79, pp. 301–307, June 1999.

[10] P.J. Loughlin and B. Tacer, "On the amplitude- and frequency-modulation decomposition of signals," *J. Acoust. Soc. Am.*, vol. 100, no. 3, pp. 1594–1601, 1996.

[11] D.W. Griffin and J.S. Lim, "Signal estimation from modified short-time fourier transform," *IEEE Trans. Acoust., Speech, Sig. Process.*, vol. ASSP-32, no. 2, pp. 236–243, 1984.

[12] M. Portnoff, "Time-frequency representation of digital signals and systems based on short-time Fourier analysis," *IEEE Trans. Acoust., Speech, Sig. Process.*, vol. 28, no. 1, pp. 55–69, Feb 1980.