

AM-FM Separation Using Auditory-Motivated Filters

Thomas F. Quatieri, *Senior Member, IEEE*, Thomas E. Hanna, and Gerald C. O'Leary

Abstract—An approach to the joint estimation of sine-wave amplitude modulation (AM) and frequency modulation (FM) is described based on the transduction of frequency modulation into amplitude modulation by linear filters, being motivated by the hypothesis that the auditory system uses a similar transduction mechanism in measuring sine-wave FM. An AM-FM estimation algorithm is described that uses the amplitude envelope of the output of two transduction filters of piecewise-linear spectral shape. The piecewise-linear constraint is then relaxed, allowing a wider class of transduction-filter pairs for AM-FM separation under a monotonicity constraint on the filters' quotient. The particular case of Gaussian filters is shown to yield a closed-form solution to AM-FM estimation while gammatone filters, used as a simplified model of auditory filters, and measured auditory filters, although not leading to a solution in closed form, provide for iterative AM-FM estimation. Solution stability analysis and error evaluation are performed and the FM transduction method is compared with the energy separation algorithm, based on the Teager energy operator, and the Hilbert transform method for AM-FM estimation. Finally, a generalization to two-dimensional (2-D) filters is described.

Index Terms—Amplitude and frequency modulation, auditory-filter transduction, energy separation algorithm, FM-to-AM transduction, gammatone filter, Teager energy operator.

I. INTRODUCTION

IN MODELS of the early stage of auditory processing, a near constant-Q filterbank approximates frequency-tuned cochlear filters; the amplitude envelope of each cochlear filter output is determined and passed on to higher processing levels. Although this simple model tracks amplitude fluctuations in an input sine wave, it does not necessarily track frequency modulations because the amplitude envelope of a frequency-modulated sine wave is constant. A hypothesis given by Saberi and Hafter [1] for the measurement of frequency modulation by the auditory system is that the cochlear filters, and perhaps higher level neurophysiological tuning curves, use transduction of frequency modulation (FM) to amplitude modulation (AM); the instantaneous frequency of the FM sweeps through the nonflat passband of the filter, thus inducing a change in the amplitude envelope of the filter output. Psychoacoustic experiments by Saberi and Hafter indicate that FM and AM

may be transformed into a common neural code in the brain stem. Goldstein [2], using an approximate Bessel function representation of FM, earlier demonstrated the importance of auditory filter shape in locating the place within each filter where FM modulation may be optimally detected; Bessel components of the modulation are weighted according to auditory filter slope, resulting in distinct amplitude envelopes with different location of the components. Others such as McEachern [3] have also argued for the importance of auditory filter shape in detecting frequency modulation. When both AM and FM are present simultaneously, these modulations are combined nonlinearly within the filter-output envelope. This paper addresses the problem of separating AM and FM from the amplitude envelope of the output of transduction filters whose spectral shape is motivated by the tuning curves of typical auditory filters [4], [5].

The approach to AM-FM separation is based on the amplitude envelope of the output of two linear transduction filters. Separation algorithms are described for numerous classes of discrete-time transduction-filters¹ using the difference or quotient of their output envelopes. In one case, the filters take on a piecewise-linear spectral shape; under certain conditions, the resulting solution to AM-FM separation is shown to reduce to an early method of FM demodulation for radio broadcasting, referred to as *balanced frequency discrimination* [6]. The AM-FM estimation method is then generalized to allow transduction by way of nonpiecewise-linear spectral shape. Although the amplitude envelope of such filter outputs may be a nonlinear function of the desired AM and FM, the relative amplitude of the two filter outputs provides a means of AM-FM separation under a monotonicity constraint on their quotient. Gaussian filters² are a particular class of these filters that can result in a closed-form solution. A second such class are gammatone filters, used to represent auditory filter dynamics [5]; although a closed-form solution is not found for this case, as well as for measured auditory filters, a unique solution to AM-FM separation exists over certain frequency ranges, obtainable by table look-up and iterative methods.

It is important to emphasize that, although a motivation for the paper's approach is the *possibility* that the human auditory system exploits FM-to-AM transduction using the spectral shapes of auditory filters, the actual mechanism for frequency

Manuscript received April 24, 1996; revised March 13, 1997. The work of T. F. Quatieri and G. C. O'Leary was supported by the Naval Submarine Medical Research Laboratory. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Dennis R. Morgan.

T. F. Quatieri and G. C. O'Leary are with Lincoln Laboratory, Massachusetts Institute of Technology, Lexington, MA 02173 USA (e-mail: tfq@sst.ll.mit.edu).

T. E. Hanna is with the Naval Submarine Medical Research Laboratory, Groton, CT 06349 USA.

Publisher Item Identifier S 1063-6676(97)06386-4.

¹The approaches of this paper are also valid in continuous time; however, because the algorithms are implemented by digital computer, sampled data representation is invoked.

²Possible use by the auditory system of the ratio of two Gaussian filter outputs as a means for FM estimation was proposed independently by McEachern [3].

demodulation in the ear is not known, and there are several candidates. For example, since the cochlea consists of a large number of overlapping filters, the auditory system may track the frequencies at which local maxima in the neural firing rate occur. Shifts in the peaks of the excitation pattern could be used to track frequency modulation of the input. On the other hand, because an FM signal can be expressed as a sum of sines through a Bessel function expansion, sine components of the FM signal may be analyzed across as well as within auditory filters. This Bessel function viewpoint was taken, for example, in the early work of Goldstein [2] in recognizing that differences between temporal amplitude envelopes from auditory filters are potential cues for discriminating FM from AM. These observations as well as more recent work, as for example by Edwards and Viemeister [7], serve to illustrate the complexity of the modeling problem. The algorithms of this paper, therefore, do not purport to describe AM-FM separation by the auditory system.

The paper is organized as follows. In Section II, the FM-to-AM transduction mechanism is described for a class of linear filters with an AM-FM sine input. In Section III, transduction filters with piecewise-linear spectra are investigated. The amplitude envelopes of the output of two distinct transduction filters are determined and their difference or quotient is used in the AM-FM separation. In Section IV, the piecewise-linear constraint is relaxed and the algorithms of Section III are generalized for arbitrary linear-filter pairs. A solution “sensitivity measure” is defined, which quantifies a change in the solution with a perturbation in the filter or input signal. A closed-form solution is then derived for the specific case of Gaussian filters. In Section V, an error analysis is performed for piecewise-linear and Gaussian filters, including the effect of change in bandwidth, change in input carrier and modulation frequencies, and addition of noise. A performance comparison is made with the Hilbert transform method [8]–[10] and the energy separation algorithm [11], based on the Teager energy operator [12], for AM-FM separation. In Section VI, AM-FM separation is performed with gammatone filters, used as a simplified model of auditory filters, as well as with measured auditory filters. In Section VII, a generalization of the method to two dimensions is described, presenting an AM-FM separation algorithm for a particular class of two-dimensional (2-D) separable Gaussian filters. The paper ends in Section VIII with a summary, current work, and a speculative discussion.

II. TRANSDUCTION OF FM TO AM

Consider a discrete-time AM-FM sine wave of the form

$$x(n) = A(n) \cos[\omega_c n + I \sin(\omega_m n)] \quad (1)$$

where ω_c and ω_m are the carrier and modulation frequencies, respectively, $A(n)$ is the time-varying amplitude, and I is the index of modulation. The instantaneous frequency $\omega(n)$ is defined as the phase derivative,³ i.e.,

$$\omega(n) = \omega_c + I\omega_m \cos(\omega_m n) \quad (2)$$

³More generally, the instantaneous frequency can be expressed as $\omega(n) = \omega_c + Iq(n)$ where $q(n)$ is a bandlimited frequency modulation signal. For the examples of this paper, however, $q(n)$ is restricted to a sine.

which represents a discretized form of the continuous-time derivative $(d/dt)[\omega_c t + I \sin(\omega_m t)]$ with time sampling interval normalized to unity. Consider the class of discrete-time filters $h(n)$ with frequency response that is zero for $\omega < 0$, i.e., $H(\omega) = 0$ for $\omega < 0$. Under the condition that $[A(n)/2][e^{-j[\omega_c n + I \sin(\omega_m n)]}] * h(n) = 0$, i.e., that the Fourier transform of the “negative frequency component” of $x(n)$ does not leak into positive frequencies,⁴ then the output of the discrete-time filter $h(n)$ to the input sequence $x(n)$ is given by

$$\begin{aligned} s(n) &= \frac{A(n)}{2} [e^{j[\omega_c n + I \sin(\omega_m n)]} + e^{-j[\omega_c n + I \sin(\omega_m n)]}] \\ &\quad * h(n) \\ &= \frac{A(n)}{2} [e^{j[\omega_c n + I \sin(\omega_m n)]}] * h(n). \end{aligned} \quad (3)$$

Observe that the filter output of the form (3) can be used for “direct” AM-FM estimation with certain filters. For example, suppose that $H(\omega)$ is unity in the region of the FM. Then the output is the analytic signal representation of the input, its imaginary part being the Hilbert transform of $x(n)$. Therefore, the amplitude $A(n)$ is simply $2|s(n)|$, and the frequency $\omega(n)$ can be computed by the derivative of the phase of $s(n)$ [8]–[10]. In discrete-time, this phase derivative can be obtained approximately by first-differencing the unwrapped phase or approximately through first-differences of the real and imaginary parts of $s(n)$. We return to the Hilbert transform approach to AM-FM estimation in Section V.

The methods of this paper rely on only the amplitude envelope, i.e., the magnitude of the output of $h(n)$, by exploiting the property of filter *transduction*, i.e., the linear-filter output can be obtained approximately by sweeping $\omega(n)$ through the filter’s transfer function. The approximation is given by [13]

$$\begin{aligned} y(n) &= \frac{A(n)}{2} [e^{j[\omega_c n + I \sin(\omega_m n)]}] H(\omega(n)) \\ &\approx s(n) \end{aligned} \quad (4)$$

where $\omega(n)$ is given in (2). The magnitude of the error in this approximation is written as $e(n) = |y(n) - s(n)|$, (4) being valid when the relative error $e(n)/|s(n)| \ll 1$. Under this condition, the amplitude envelope of the instantaneous output is given by the approximation

$$|y(n)| = A(n) |H[\omega(n)]| \quad (5)$$

where, for convenience, the factor $\frac{1}{2}$ in (4) has been discarded. We see that in using the envelope of the output, it follows that only the magnitude of the filter $H(\omega)$ is used. The approximation (5) is the basis for the AM-FM separation methods of this paper.

An upper bound⁵ on the error $e(n)$ can be expressed as a function of the duration or temporal “localization” of the filter $h(n)$ and the temporal “smoothness” of $A(n)$ and $\omega(n)$

⁴The approximation does not hold for very low carrier with large frequency modulation. For this case, the Fourier transform of $[A(n)/2][e^{-j[\omega_c n + I \sin(\omega_m n)]}] * h(n) \neq 0$ for $\omega > 0$, and thus (3) is approximate. This distortion will influence the accuracy of any AM-FM separation method that relies on (3) such as the Hilbert transform method, as well as the methods described in this paper.

⁵An upper bound on $e(n)$ also serves as an upper bound on $\|y(n) - |s(n)|\|$ because $\|y(n) - |s(n)|\| \leq |y(n) - s(n)|$.

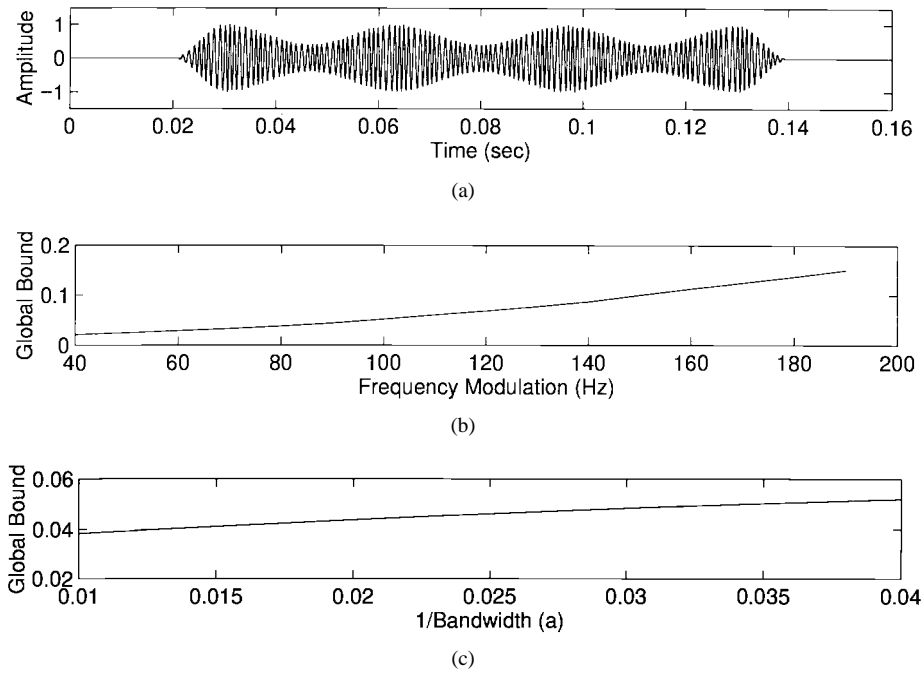


Fig. 1. Global error bounds for an AM-FM sine. (a) Waveform. (b) Bound with increasing FM. (c) Bound with decreasing bandwidth.

[13]. These conditions reflect the requirement that the varying amplitude and frequency be almost constant over the duration of the filter's impulse response at each time instant, the input sine appearing as an eigenfunction of the linear filter. An upper bound on $e(n)$ can be quantified (see Appendix A) as

$$e(n) \leq \frac{\pi}{\sqrt{3}} [\Delta_1(h)D(A) + A_{\max}\Delta_2(h)D(\omega)] \quad (6)$$

where $\Delta_1(h)$ and $\Delta_2(h)$ are the second and fourth moments of $h(n)$, i.e.,

$$\Delta_k(h) = \left[\sum_n n^{2k} |h(n)|^2 \right]^{1/2}$$

giving measures of time localization of $h(n)$. $D(A)$ and $D(\omega)$ are measures of the average rate of change of $A(n)$ and $\omega(n)$, respectively; specifically, for an arbitrary signal $z(n)$

$$D(z) = \int_{-\infty}^{\infty} |\dot{z}(\tau)| d\tau.$$

For an input of the form in (1), the error bound grows as the change in instantaneous amplitude and modulating frequency increases, or as the filter bandwidth decreases.

Fig. 1(a) shows an input sequence $x(n)$ of (1) of amplitude envelope with a 30 Hz modulation and tapered on each end with a half cycle of a von Hann window. The instantaneous frequency of this particular sequence has carrier $\omega_c = 1000$ Hz, index of modulation $I = 1$, and modulation frequency $\omega_m = 70$ Hz. This sequence serves throughout the paper as the basis for a variety of experiments, but where the carrier and modulation frequencies are varied. Fig. 1(b), in particular, shows how the error bound (6) increases with increasing the frequency modulation ω_m from 40 Hz to 190 Hz. A Gaussian filter of the form $H(\omega) = e^{-a(\omega-\omega_o)^2}$ is used as the transduction filter. Fig. 1(c), using the specific sequence of

Fig. 1(a) with frequency modulation ω_m held fixed at 70 Hz, illustrates the increasing trajectory of the error bound for the Gaussian transduction filter when its bandwidth is decreased, i.e., the value of parameter a is increased from 0.01 to 0.04.

The bound on the instantaneous error $e(n)$ in (6) is a single value, given in terms of a global measure of rate of change in amplitude and frequency. Appendix A gives a tighter, time-varying, error bound in terms of a local rate of change in these functions; when the instantaneous amplitude or frequency changes more rapidly, then the local error bound becomes larger [13]. Both the global and local error bounds give guidelines for predicting behavior of AM-FM separation algorithms throughout the paper.

III. AM-FM SEPARATION WITH PIECEWISE-LINEAR FILTERS

Consider a frequency response $H(\omega)$ with piecewise-linear magnitude and arbitrary phase, one specific subset being real and positive bandpass filters⁶ with a positive-sloped and a negative-sloped region as illustrated in Fig. 2. The filter characteristic in the positive-sloped region is expressed as

$$H(\omega) = a\omega + b \quad \omega \in H^+ \quad (7)$$

where H^+ denotes the frequency interval over which $H(\omega)$ has positive slope. It is assumed that the maximum and minimum frequencies about ω_c , i.e., $\omega_c \pm I\omega_m$, fall within⁷ H^+ . Under these conditions, using (5) and (7), the amplitude of the filter output can be written approximately as

$$|y(n)| = A(n)[a\omega(n) + b] \quad (8)$$

⁶In this paper, we study positive, zero-phase piecewise-linear, Gaussian, and gammatone filters because they are localized about the time origin. Although the filter phase is not used in the approximation (5), a nonzero-phase function can increase the approximation error. Additional discussion on the use of zero-phase filters is given in Section VI and Appendix A.

⁷The input spectrum, however, can fall outside this range.

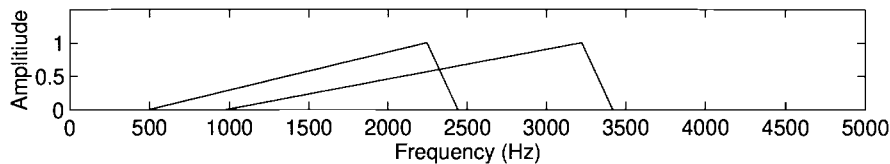


Fig. 2. Filters $H_1(\omega)$ (left) and $H_2(\omega)$ (right) with piecewise-linear spectra.

where the FM is linearly transduced to an AM.

One approach to separating the AM and FM from the nonlinear function of (8) is to utilize the amplitude envelope of the output of two distinct filters of the form of $H(\omega)$ in (7). Consider two such discrete-time filters in their positive-sloped regions, i.e., $H_i(\omega) = a_i\omega + b_i$ for $\omega \in H_i^+$ where $i = 1, 2$. The output amplitude envelopes for $\omega(n)$ in the intersection of the regions H_i^+ are then expressed as $|y_i(n)| = A(n)[a_i\omega(n) + b_i]$ for $i = 1, 2$, which expanded gives

$$\begin{aligned} |y_1(n)| &= a_1 A(n)\omega(n) + b_1 A(n) \\ |y_2(n)| &= a_2 A(n)\omega(n) + b_2 A(n). \end{aligned} \quad (9)$$

Multiplying the top component of (9) by (a_2/a_1) , the following equation pair results:

$$\begin{aligned} a_2 A(n)\omega(n) + \left(\frac{a_2 b_1}{a_1}\right) A(n) &= \left(\frac{a_2}{a_1}\right) |y_1(n)| \\ a_2 A(n)\omega(n) + b_2 A(n) &= |y_2(n)|. \end{aligned}$$

Differencing the equation pair, combining terms, and solving for $A(n)$ yields

$$A(n) = \frac{a_2 |y_1(n)| - a_1 |y_2(n)|}{a_2 b_1 - a_1 b_2} \quad (10)$$

from which $\omega(n)$ is obtained using either component of (9). Denoting the above denominator by the variable U , i.e., $U \triangleq a_2 b_1 - a_1 b_2$, then (10) is valid under the condition $U \neq 0$. The condition $U \neq 0$ can be shown to be equivalent to constraining the two positively-sloped regions of the filters H_i to be “distinct” in the sense that they are not related by a scale factor, i.e., $a_2\omega + b_2 \neq \alpha(a_1\omega + b_1)$. If in (9), we set $a_1 = -a_2$ and $b_1 = b_2$ then the AM-FM separation algorithm reduces to a discrete form of a *balanced frequency discriminator*, which is a classical early method of FM demodulation for radio broadcasting [6], [14] (see Appendix B).

Alternatively, the instantaneous frequency $\omega(n)$ can be estimated first, followed by the amplitude modulation $A(n)$. This can be performed by dividing the pair of (9), i.e.,

$$\frac{a_2\omega(n) + b_2}{a_1\omega(n) + b_1} = \frac{|y_2(n)|}{|y_1(n)|} \quad (11)$$

to obtain

$$\omega(n) = \frac{b_2 |y_1(n)| - b_1 |y_2(n)|}{a_1 |y_2(n)| - a_2 |y_1(n)|} \quad (12)$$

from which $A(n)$ can be estimated using either component of (9). This alternative solution serves as a dual to (10) and provides the motivation for the generalization given in the following section.

IV. EXTENSION TO NONPIECEWISE-LINEAR FILTERS

In this section, the approach introduced through (11) is extended to nonpiecewise-linear filters. Closed-form solutions are derived for the special case of Gaussian-filter pairs.

A. Generalized Filter Structures

Consider arbitrary frequency responses, $H_1(\omega)$ and $H_2(\omega)$, equal to zero for $\omega \leq 0$. Using (5), the following equation pair can be written:

$$\begin{aligned} |y_1(n)| &= A(n)|H_1[\omega(n)]| \\ |y_2(n)| &= A(n)|H_2[\omega(n)]|. \end{aligned} \quad (13)$$

Motivated by (11), we write

$$\frac{|H_2[\omega(n)]|}{|H_1[\omega(n)]|} = \frac{|y_2(n)|}{|y_1(n)|} = v(n).$$

Defining

$$g[\omega(n)] \triangleq \frac{|H_2[\omega(n)]|}{|H_1[\omega(n)]|} \quad (14)$$

then whenever $g(\omega)$ is invertible

$$\omega(n) = g^{-1}[v(n)]. \quad (15)$$

An estimate of the amplitude envelope then follows from either component of (13).

There will be a unique solution (15) when $g(\omega)$ is strictly monotonic in a frequency interval in which $\omega(n)$ lies. The derivative of $g(\omega)$ can be expressed as

$$\begin{aligned} \dot{g}(\omega) &= \frac{d}{d\omega} \frac{|H_2[\omega(n)]|}{|H_1[\omega(n)]|} \\ &= \frac{\left[|H_1(\omega)| \frac{d}{d\omega} |H_2(\omega)| - |H_2(\omega)| \frac{d}{d\omega} |H_1(\omega)| \right]}{|H_1(\omega)|^2} \end{aligned}$$

so that a sufficient condition⁸ for a unique solution in the region of interest becomes

$$\begin{aligned} U(\omega) \triangleq & |H_1(\omega)| \frac{d}{d\omega} |H_2(\omega)| \\ & - |H_2(\omega)| \frac{d}{d\omega} |H_1(\omega)| \neq 0. \end{aligned} \quad (16)$$

The condition (16), as we shall show, is similar to our previous condition $U \neq 0$ that was derived for a piecewise-linear frequency response; hence the functional notation $U(\omega)$. We

⁸The derivative may be zero at the borders of the region of interest and at certain inflection points.

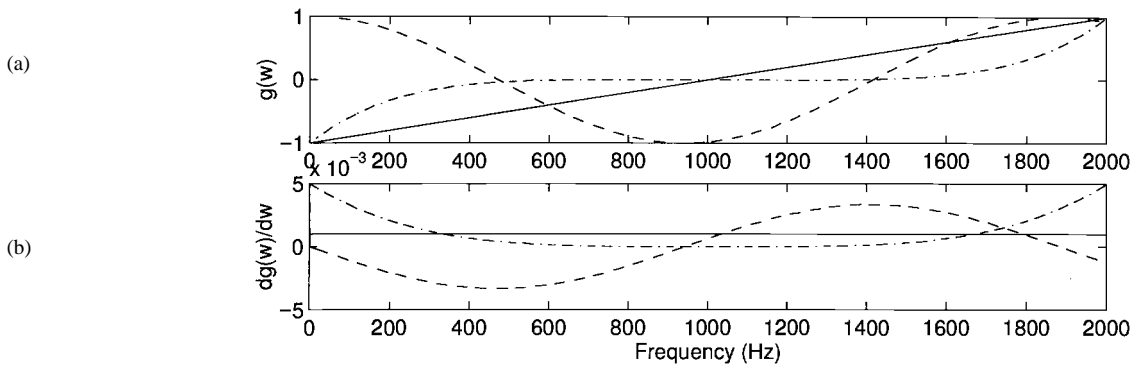


Fig. 3. Example functions (a) $g(\omega)$ and (b) its derivative $\dot{g}(\omega)$ for three different solutions: unique (solid), two solutions (dash), and unstable (dash-dot).

can think of the reciprocal of $|\dot{g}(\omega)|$ as a solution “sensitivity measure” defined by

$$S(\omega) \triangleq \frac{1}{|\dot{g}(\omega)|}. \quad (17)$$

A large value of $S(\omega)$ implies a large change in the solution $\omega(n)$ for a small deviation in $g(\omega)$, i.e., when the model $g(\omega)$ or the measurement $v(n) = (|y_2(n)|/|y_1(n)|)$ is not exact. $S(\omega)$ also provides a means of determining when the solution is ambiguous. Example functions $g(\omega)$ and their derivatives $\dot{g}(\omega)$ are illustrated in Fig. 3, giving conditions for a unique solution (solid) and two solutions (dash) over a frequency interval $[0, 2000]$ Hz. The third case (dash-dot) also provides for a unique solution, but unlike the first case (solid), the solution in the midfrequency region is highly sensitive to perturbations in the function $g(\omega)$ or the measurement $v(n)$.

When $H_1(\omega)$ and $H_2(\omega)$ are the linear functions used in (9), then the condition in (16) is reduced to $U(\omega) = (a_2\omega + b_2)a_1 - (a_1\omega + b_1)a_2 = a_2b_1 - a_1b_2 \neq 0$ which was obtained in Section III through the dual solution. For this piecewise-linear case, $g[\omega(n)] = (a_2\omega(n) + b_2)/(a_2\omega(n) + b_2)$ and the sensitivity function $S(\omega) = (1/|a_2b_1 - a_1b_2|)(a_1\omega + b_1)^2$ that increases with increasing frequency, as well as when the two filters approach multiples of one another, i.e., $U = 0$.

B. Gaussian Filters

In general, determining the solution in (15) is a nonlinear problem and thus perhaps requires iteration. There do exist, however, classes of filters that yield closed-form solutions. One such class is that of Gaussian filters of the form $H_1(\omega) = e^{-a_1(\omega-\omega_1)^2}$ and $H_2(\omega) = e^{-a_2(\omega-\omega_2)^2}$ that are centered at frequencies ω_1 and ω_2 , respectively. For this filter pair, we write

$$\begin{aligned} g[\omega(n)] &= \frac{|H_2[\omega(n)]|}{|H_1[\omega(n)]|} \\ &= \frac{e^{-a_2(\omega-\omega_2)^2}}{e^{-a_1(\omega-\omega_1)^2}} = \frac{|y_2(n)|}{|y_1(n)|}. \end{aligned} \quad (18)$$

Two simplifying cases from the various solution possibilities are: i) equal bandwidths with different center frequencies and ii) equal center frequencies with different bandwidths. Henceforth, we refer to a_1 and a_2 as “bandwidth factors” because they control the filters’ bandwidth.

With equal bandwidth factors, i.e., $a_1 = a_2 = a$, taking the logarithm of both sides of the latter part of (18), we obtain

$$a^{-1} \log \left(\frac{|y_2(n)|}{|y_1(n)|} \right) = (-2\omega\omega_1 + \omega_1^2) + (2\omega\omega_2 - \omega_2^2)$$

and solving for $\omega(n)$

$$\omega(n) = \frac{u(n) - (\omega_1^2 - \omega_2^2)}{-2(\omega_1 - \omega_2)} \quad (19)$$

where $u(n) = a^{-1} \log(|y_2(n)|/|y_1(n)|)$. Unlike the piecewise-linear case, the input carrier and FM sweep is allowed to fall anywhere within the region of filter overlap. The function $U(\omega)$ for the Gaussian pair can be derived as $U(\omega) = -2a(\omega_1 - \omega_2)e^{-a(\omega-\omega_1)^2}e^{-a(\omega-\omega_2)^2}$ so that a solution exists whenever the bandwidth is nonzero and the center frequencies are different, i.e., the filters are not related by a scale factor.⁹ The sensitivity function is given by $S(\omega) = (1/2a|\omega_1 - \omega_2|)e^{-a(\omega_1^2 - \omega_2^2)}e^{-2a\omega(\omega_2 - \omega_1)}$ which increases with decreasing frequency and with decreasing bandwidth or center frequency spacing. Example functions $g(\omega)$ and $S(\omega)$ are illustrated in Fig. 4.

An example of AM-FM separation is shown in Fig. 5. The input sequence $x(n)$ in (1) has an amplitude envelope with a 30 Hz modulation and tapered on each end with a half cycle of a von Hann window, and instantaneous frequency with carrier $\omega_c = 1000$ Hz with modulation frequency $\omega_m = 70$ Hz and modulation index $I = 1$. Two Gaussian filters of equal bandwidth were selected at 900 Hz and 1100 Hz. The sequence $x(n)$ is filtered with the Gaussian-filter pair and the resulting sequences $y_1(n)$ and $y_2(n)$ are used to compute $u(n)$. The frequency and amplitude estimates are shown superimposed on the originals, illustrating accurate estimates except at the edges of the nonzero interval. In these boundary regions, a maximum error is expected since the smoothness of the input amplitude envelope $A(n)$ and instantaneous frequency $\omega(n)$ is minimum in these regions; i.e., the frequency derivative is infinite at the beginning and end of the sine, and the amplitude slope is largest at these time instants. Therefore, the upper bound on the approximation error of (5) is largest in these boundary regions (see Appendix A).

In the second class of Gaussian filters, filter center frequencies are equal, i.e., $\omega_o = \omega_1 = \omega_2$, and bandwidths

⁹More generally, the Gaussian filters can contain an arbitrary amplitude scaling which manifests itself as simply a scaling of $U(\omega)$.

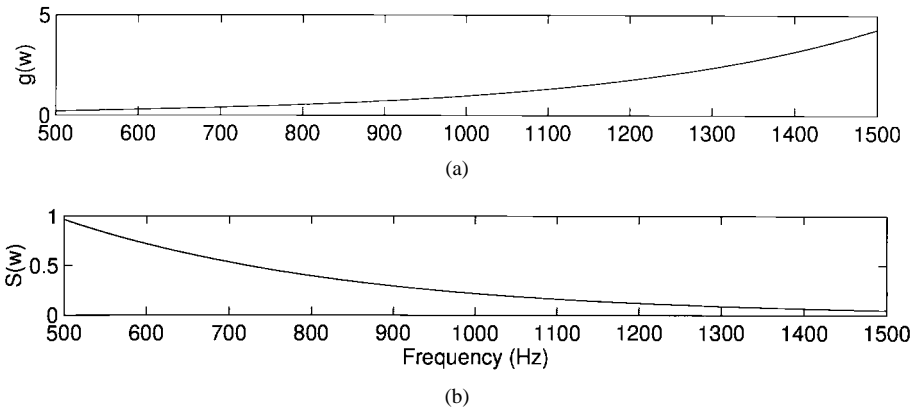


Fig. 4. Functions (a) $g(\omega)$ and (b) sensitivity measure $S(\omega)$ for Gaussian-filter pair of Fig. 5(b).

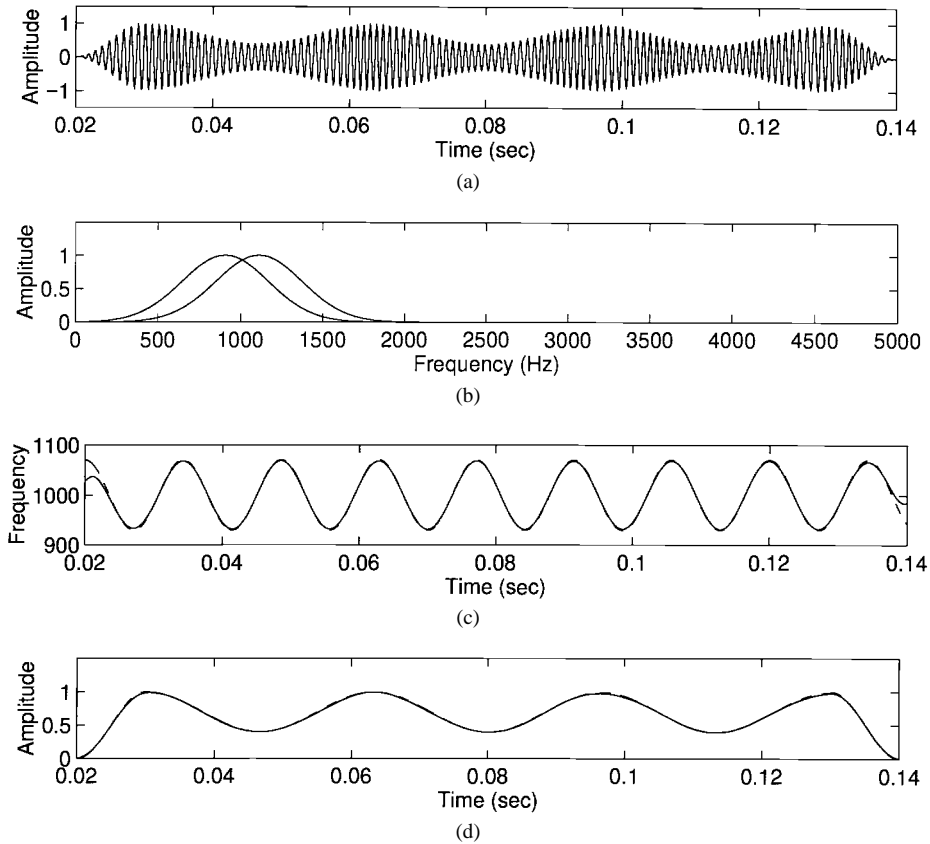


Fig. 5. Example of AM-FM estimation. (a) Original sequence. (b) Gaussian filters. (c) Superimposed original (solid) and estimated (dash) frequency. (d) Superimposed original (solid) and estimated (dash) amplitude.

are unequal, i.e., $a_1 \neq a_2$, for which case two solutions are possible. Taking the logarithm of both sides of the latter part of (18), we obtain

$$\log\left(\frac{|y_2(n)|}{|y_1(n)|}\right) = (a_1 - a_2)(\omega - \omega_o)^2$$

and solving for $\omega(n)$

$$\omega(n) = \omega_o \pm \sqrt{u(n)} \quad (20)$$

where $u(n) = (a_1 - a_2)^{-1} \log(|y_2(n)|/|y_1(n)|)$. The expression (20) is meaningful provided $u(n) \geq 0$. This condition, however, is always satisfied, which is proven by observing that when $a_1 \geq a_2$ then $(|y_2(n)|/|y_1(n)|) \geq$

1. Therefore, $\log(|y_2(n)|/|y_1(n)|) \geq 0$ so that $u(n) \geq 0$. Likewise, when $a_1 \leq a_2$ then $(|y_2(n)|/|y_1(n)|) \leq 1$. Therefore, $\log(|y_2(n)|/|y_1(n)|) \leq 0$ so that $u(n) \geq 0$. The two solutions of (20) reflect the parabolic shape of $g(\omega) = e^{(a_1 - a_2)(\omega - \omega_o)^2}$ about the center frequency ω_o . The corresponding sensitivity measure is given by $S(\omega) = (1/2|\omega - \omega_o||a_1 - a_2|)e^{-(a_1 - a_2)(\omega - \omega_o)^2}$ giving infinite sensitivity at $\omega = \omega_o$ where the slope of $g(\omega)$ is zero (see Fig. 6). Although there are generally two solutions, when $\omega(n)$ is known to fall to the right or to the left of the center frequency ω_o then a unique solution can be found.¹⁰

¹⁰Even when $\omega(n)$ straddles ω_o , by changing the sign in (20) whenever

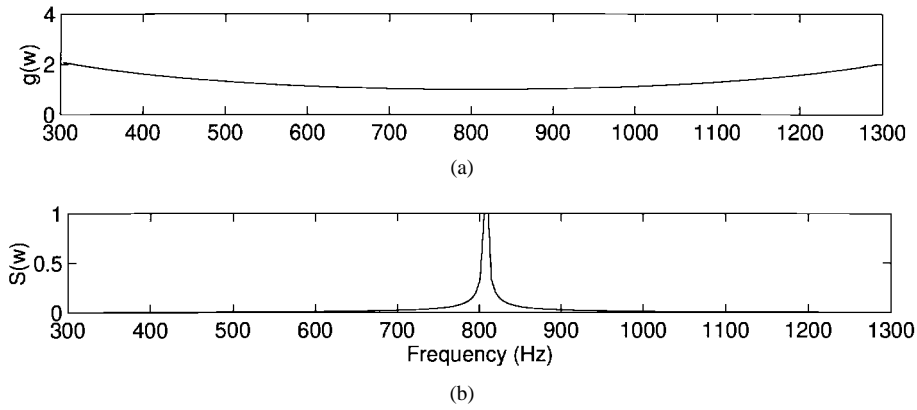


Fig. 6. Functions (a) $g(\omega)$ and (b) sensitivity measure $S(\omega)$ for Gaussian-filter pair of Fig. 7(a).

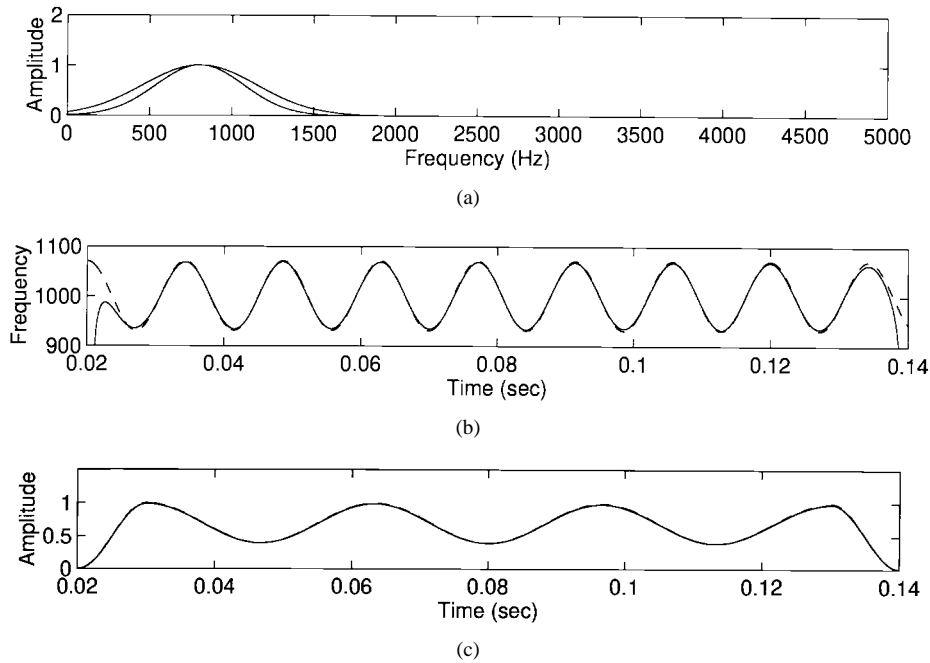


Fig. 7. Example of AM-FM estimation with Gaussian filters at same center frequency. (a) Gaussian filters. (b) Superimposed original (solid) and estimated (dash) frequency. (c) Superimposed original (solid) and estimated (dash) amplitude.

An example of AM-FM separation is shown in Fig. 7. The input sequence $x(n)$ is the same as in the previous example with $\omega_c = 1000$ Hz, $\omega_m = 70$ Hz, and a 30 Hz AM. Two Gaussian filters centered at 800 Hz were selected with bandwidth factors $a_1 = 0.013$ and $a_2 = 0.010$. In this case, the condition $u(n) > 0$ is satisfied so that the correct solution is given by $\omega(n) = \omega_o + \sqrt{u(n)}$. The frequency and amplitude estimates are shown superimposed on the originals, with maximum error, as before, at the beginning and end of the sine input.

V. ERROR EVALUATION

In this section, a flavor is obtained for the performance of the piecewise-linear and Gaussian-filter pairs with respect to filter bandwidth, modulation frequency, and noise addition. Comparisons are made with the Hilbert transform method $u(n) \approx 0$, we can obtain the modulation component $\pm\sqrt{u(n)}$ to within a sign factor.

[8]–[10] and energy separation algorithm [11] for AM-FM estimation.

A. Filter Bandwidth

As the bandwidth of the filters $H_i(\omega)$ narrows, one expects the performance of the technique to decline because the conditions under which the separation was derived become less valid; the “localization” of the filter responses $h_i(n)$ is reduced with decreasing filter bandwidth. Furthermore, the sensitivity $S(\omega)$ increases with decreasing bandwidth. To obtain a feeling for the change in performance with decreasing filter bandwidth, a piecewise-linear filter pair configuration was first constructed, similar to that of Fig. 2. To characterize each filter, we first define a frequency vector $\bar{\Omega} = [\omega_p, \omega_i, \omega_n]$ Hz, where ω_p is the location at which the positive-sloped region intersects the frequency axis, ω_n is the location at which the negative-sloped region intersects the frequency axis, and ω_i is the location at which the two regions intersect. The filter

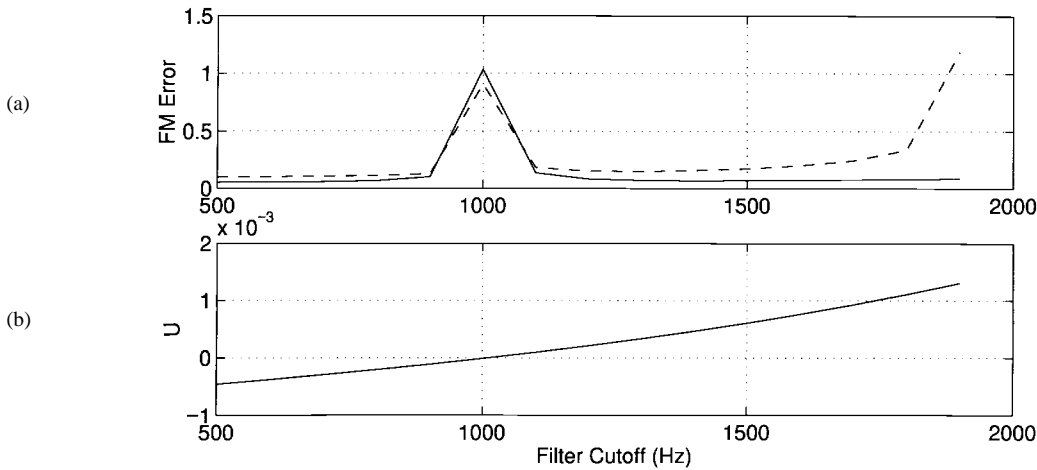


Fig. 8. FM estimation with piecewise-linear filter of varying bandwidth. (a) Mean-squared FM error (in Hz) for a 2000 Hz (dash) and 3000 Hz (solid) carrier as function of varying frequency cutoff ω_l . (b) Denominator term $U = a_2b_1 - a_1b_2$ as a function of varying frequency cutoff ω_l .

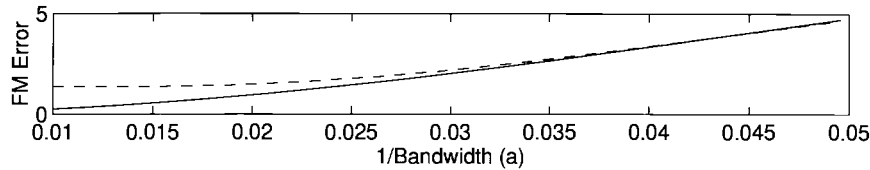


Fig. 9. Example of mean-squared FM error (in Hz) for 2500 Hz carrier (solid) and 250 Hz carrier (dash) with decreasing Gaussian filter bandwidth.

$H_1(\omega)$ has a fixed frequency vector $\bar{\Omega}_1 = [1000, 4000, 4500]$ Hz, and the filter $H_2(\omega)$ has a variable frequency vector $\bar{\Omega}_2 = [\omega_l, 4000, 4500]$ Hz, i.e., $H_2(\omega)$ is identical to $H_1(\omega)$ but where ω_l is allowed to vary. The variable ω_l of $H_2(\omega)$ is swept through the range of values 500 Hz to 2000 Hz in increments of 100 Hz. The input signal $x(n)$ is identical to that of Fig. 5 except for the carrier frequency and a 50 Hz FM. Fig. 8(a) gives the mean-squared error in the FM estimate as a function of the variable ω_l for a carrier of 2000 Hz and 3000 Hz. As predicted from (6), there is a general upward trend in the FM error as ω_l increases.¹¹ There is also an abrupt increase in the FM error as the varying cutoff frequency ω_l of filter $H_2(\omega)$ approaches $\omega_p = 1000$ Hz of $H_1(\omega)$. As the frequencies merge at 1000 Hz, the denominator term $U = a_2b_1 - a_1b_2$ of (10) becomes zero, as illustrated in Fig. 8(b), resulting in an unstable solution because the two filters become close to identical.¹² In this case, the sensitivity measure $S(\omega) = (1/|a_2b_1 - a_1b_2|)(a_1\omega + b_1)^2$ approaches infinity. Similar trends are obtained for AM estimation error. The observed increase in error with decreasing carrier is a function of the particular filter-pair configuration; AM-FM estimation for a low carrier can be improved with an alternate configuration as illustrated in a following section.

Error evaluation was also performed with a Gaussian-filter pair configuration, similar to that of Fig. 5(b), where the bandwidth was made variable. The bandwidth factor a is swept

¹¹In addition, the error increases more rapidly for the 2000 Hz carrier because the regions of linearity in the filter pair are exceeded sooner.

¹²The varying cutoff ω_l did not exactly meet the 1000 Hz cutoff of $H_2(\omega)$ because of the discrete increments with which the frequency changes in this simulation.

through the range of values $a = 0.01$ to 0.05 corresponding to decreasing bandwidth. The input sequence $x(n)$ is of the form of Fig. 5. Fig. 9 gives the mean-squared error in the FM estimate as a function of bandwidth for the two different carrier frequencies of 2500 Hz and 250 Hz. In each case, the Gaussian-filter pair is centered at the carrier with a 200 Hz separation of the two filter peaks. As predicted, there is a general upward trend in the FM error as the bandwidth decreases. A similar error trend occurs for the AM estimate.

B. Source Sensitivity

In the next experiment set, the modulation frequency ω_m is varied and the filter pair is kept fixed. Both piecewise-linear and Gaussian filter-pair configurations, similar to those in Fig. 2 and 5, respectively, are tested for the input carriers of 2500 Hz and 250 Hz. For the 2500 Hz carrier, the piecewise-linear filters are characterized by frequency offsets $\bar{\Omega}_1 = [500, 4000, 4500]$ Hz and $\bar{\Omega}_2 = [1000, 4000, 4500]$ Hz. The Gaussian filters have a fixed bandwidth factor $a = 0.01$, and center frequencies 2400 Hz and 2600 Hz. For the 250 Hz carrier, the linear-filter pair configuration has the form¹³ shown in Fig. 10, with frequency vectors $\bar{\Omega}_1 = [0, 50, 1000]$ Hz and $\bar{\Omega}_2 = [0, 50, 2000]$ Hz, where the negative-going filter slopes are used for transduction. The Gaussian filters are centered at 150 Hz and 350 Hz. Fig. 11(a) shows the mean-squared FM error with FM increasing from 30 Hz to 170 Hz for the 2500 Hz carrier. The FM error increases as the modulation frequency increases because the input "smoothness" condition for the approximation of (13) becomes less valid (see

¹³Interestingly, this shape is similar to that of auditory filters for very low characteristic frequency.

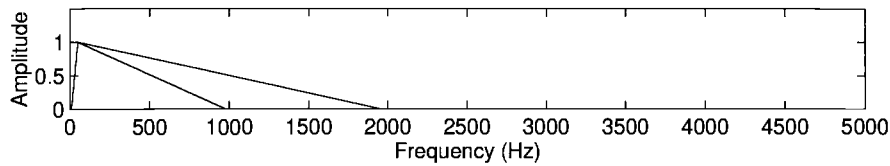


Fig. 10. Filters $H_1(\omega)$ (lower) and $H_2(\omega)$ (upper) with piecewise-linear spectra used for AM-FM estimation with low-frequency carrier.

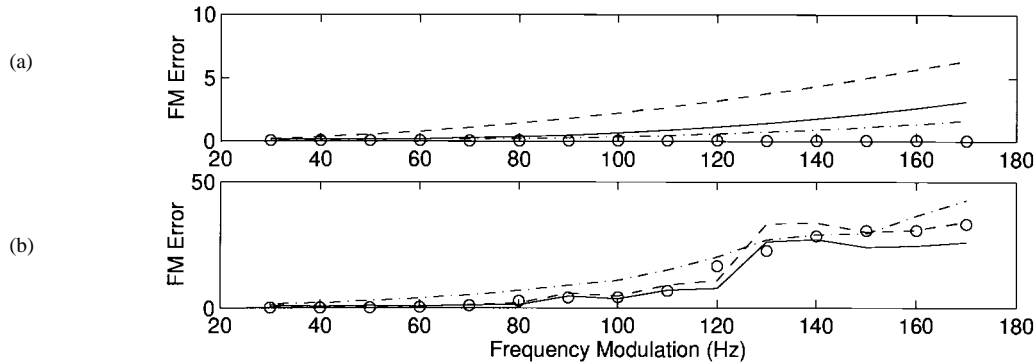


Fig. 11. Example of mean-squared FM estimation error (in Hz) with varying FM modulation frequency for FM linear transduction (circles), Gaussian transduction (solid), energy separation (dash-dot), and Hilbert transform (dash) methods. (a) 2500 Hz carrier. (b) 250 Hz carrier.

Appendix A). Fig. 11(b) shows the FM error for the piecewise-linear and Gaussian transduction filters with the 250 Hz carrier, illustrating a similar increasing error trend.

A comparison with two standard AM-FM estimation methods, the discrete energy separation algorithm (DESA) [11], based on the Teager energy operator [12], and the Hilbert transform-based method [8]–[10], [15], is also shown in Fig. 11. The energy separation algorithm in discrete time is given by

$$a(n) = \frac{2\Psi[x(n)]}{\sqrt{\Psi[x(n+1) - x(n-1)]}}$$

$$\omega(n) = \arcsin \left(\sqrt{\frac{\Psi[x(n+1) - x(n-1)]}{4\Psi[x(n)]}} \right)$$

where the Teager energy operator $\Psi[x(n)]$ is given by the three-point function

$$\Psi[x(n)] \triangleq x^2(n) - x(n-1)x(n+1).$$

Prior to these operations a short five-point FIR smoothing is applied to $\Psi[x(n)]$ and $\Psi[x(n+1) - x(n-1)]$ to reduce estimation error [11]. Given a real AM-FM continuous-time signal $x(t) = a(t) \cos[\phi(t)]$, an alternative approach to estimate its envelope $|a(t)|$ and instantaneous frequency $\omega(t) = \dot{\phi}(t)$ is to use the Hilbert transform of $x(t)$. Specifically, if $X(\omega)$ is the Fourier transform of $x(t)$, its Hilbert transform is the signal $\hat{x}(t)$ with Fourier transform $\hat{X}(\omega) = -j \operatorname{sgn}(\omega)X(\omega)$. The related complex-valued *analytic signal* is $x(t) + j\hat{x}(t) = r(t)e^{j\theta(t)}$. Thus the Hilbert transform can provide an envelope and instantaneous frequency $r(t) = \sqrt{x^2(t) + \hat{x}^2(t)}$ and $\dot{\theta}(t) = (d/dt) \arctan[\hat{x}(t)/x(t)]$, respectively. In discrete time, the amplitude is given by samples of $r(t)$ and the

instantaneous frequency is approximated¹⁴ by a first-backward difference on samples of the unwrapped phase of $\dot{\theta}(t) = (d/dt) \arctan[\hat{x}(t)/x(t)]$, i.e.,

$$a(n) = r(n)$$

$$\dot{\theta}(n) = \Theta(n) - \Theta(n-1)$$

where $\Theta(n)$ represents a discretized unwrapped phase obtained by tracking 2π jumps in the principal phase function (calculated modulo 2π) of $\theta(n)$ [8]. The Hilbert transform was designed with the Parks–McClellan minimax error-based algorithm [8], constrained to give smooth transition bands near $\omega = 0$ and $\omega = \pi$ to avoid aliasing in the Hilbert transform time-domain response. For the bandpass test signals, this design reduces error in the discrete-time phase derivative.

The three methods of AM-FM estimation, based on filter transduction, the energy separation algorithm, and Hilbert transform, generally yield different estimates. As seen in Fig. 11, the FM transduction methods overall give the lowest FM mean-squared error for the low carrier. The Hilbert transform method, however, is very close to these in performance. For the high carrier, the linear transduction method gives the least error, while the Gaussian transduction method gives an error between that from the energy separation algorithm and Hilbert transform method. It was observed that the mean-squared FM error for the transduction methods is dominated by the error at the waveform edges where there is maximum change in frequency and amplitude; this is consistent with the local error bound of Appendix A. The energy separation algorithm, on the other hand, gives the least observed

¹⁴ Alternatively, the phase derivative can be approximated using samples of $\dot{\theta}(t) = [x(t)\hat{x}(t) + \dot{x}(t)\hat{x}(t)]/r^2(t)$ where samples of the derivatives $\dot{x}(t)$ and $\dot{\hat{x}}(t)$ are approximated by first-backward differences. Using this method, empirical results comparable to first-differencing unwrapped phase samples were obtained.

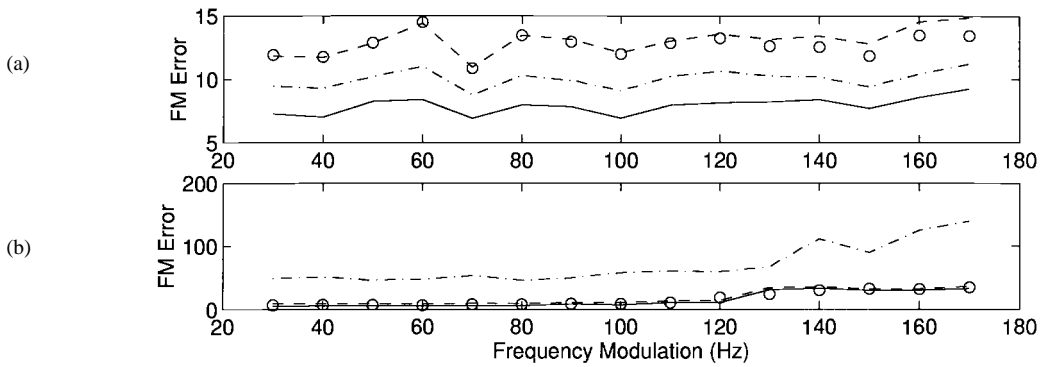


Fig. 12. Example of mean-squared FM estimation error (in Hz) with varying FM modulation frequency in 30 dB SNR Gaussian noise, for FM linear transduction (circles), Gaussian transduction (solid), energy separation (dash-dot), and Hilbert transform (dash) methods. (a) 2500 Hz carrier. (b) 250 Hz carrier.

instantaneous error at these boundary points because, for each output sample, it uses an extremely short window (five samples in duration), which allows it to nearly instantaneously adapt during signal transitions. Consequently, a different error criterion more severely penalizing maximum instantaneous error would likely change the relative performance of these methods.¹⁵

C. Robustness in Noise

A comparison of the transduction methods with the energy separation algorithm and Hilbert transform technique was also carried out in a noise background. Because the transduction-filter methods implicitly remove much of the input noise, in order to test the methods in a comparable noise level, the noise background was generated by bandpass filtering white-Gaussian noise with a Gaussian filter of bandwidth equal to the smallest of the two Gaussian transduction filters. This bandpass noise was added to the signal $x(n)$ and the resulting sum used as input to the AM-FM estimation algorithms. Specifically, the AM-FM signal of the previous example with a carrier of 2500 Hz and 250 Hz was corrupted by the additive bandpass Gaussian noise at a signal-to-noise ratio (SNR)¹⁶ of approximately 30 dB. The results are shown in Fig. 12 for the same four methods of Fig. 11. For the low carrier, the two transduction methods with piecewise-linear and Gaussian filters give performance which is comparable to the Hilbert transform technique, the energy separation algorithm showing the greatest noise sensitivity. For the high carrier, the Gaussian transduction filters give the best performance, while the linear transduction filter and Hilbert transform methods give comparable performance with the greatest error. Similar trends occur for an SNR of 20 dB, but with larger error.

In the transduction methods, one can displace filter pairs to invoke transduction along different filter regions. For example, displacing the Gaussian filters relative to the 2500 and 250 Hz

carrier by 100 Hz to the left gave a noticeable error increase, while displacing to the right by 100 Hz made little change, consistent with the sensitivity measure $S(\omega)$ for this Gaussian pair. Performance with such change in filter location, both with and without a noise background, requires further study.

VI. AM-FM SEPARATION WITH AUDITORY-LIKE FILTERS

Motivation for the AM-FM estimation algorithms of this paper is the hypothesis that the auditory system uses FM-to-AM transduction in measuring sine-wave FM. Although the auditory system may not indeed exploit FM transduction to separate AM and FM, nevertheless, it is of interest to determine whether AM-FM separation can be performed with gammatone filters, used as simplified models of auditory filters, as well as with measured auditory filters. Such filters also serve to demonstrate the approach when a closed-form solution does not exist, and thus where an iterative solution is useful.

A. Gammatone Filters

A gammatone-filter impulse response, used to model both cochlear and neurophysiological tuning curves [5], [16], [17], is the product of a gammatone distribution and a tone, and in continuous time is expressed as $h(t) = At^{N-1}e^{-bt} \cos(\omega_0 t + \theta)$. Fig. 13(a) shows the discrete Fourier transform magnitude of a pair of gammatone filters.¹⁷ For two closely spaced discrete-time gammatone filters, $h_1(n)$ and $h_2(n)$, the function $g(\omega) = (|H_2[\omega(n)]|/|H_1[\omega(n)]|)$ was not found amenable to a closed-form solution for $\omega(n)$. Several other proposed expressions for the magnitude of the auditory-filter transfer functions were also considered. For example, an alternative to gammatone filters is the rounded exponential (roex) filter [19] given by $H(\omega) = (1-r)(1+p\omega)e^{-p\omega} + r$. Although simpler in form than the gammatone filter, the roex transduction filter also was not found to provide a closed-form solution to AM-FM separation. Nevertheless, in these cases, one can use an

¹⁵In addition to different error criteria, one might explore different filter configurations as, for example, the complementary piecewise-linear filter pair of Armstrong's balanced frequency discriminator. The filters of this paper were selected for their resemblance to auditory filters.

¹⁶SNR is defined as the ratio of the variance of the signal and bandpass noise in dB. The variance of the input signal is obtained by averaging power over its nonzero region, while the noise variance is obtained by averaging power of the bandpass-filtered white noise.

¹⁷Discrete-time gammatone filters were obtained from the auditory toolbox developed by Slaney [18]. Filter order depends on its characteristic frequency and ranges from $N = 2$ to $N = 32$. In approximately simulating auditory filters, bandwidth of the gammatone filters increases logarithmically with increasing characteristic frequency.

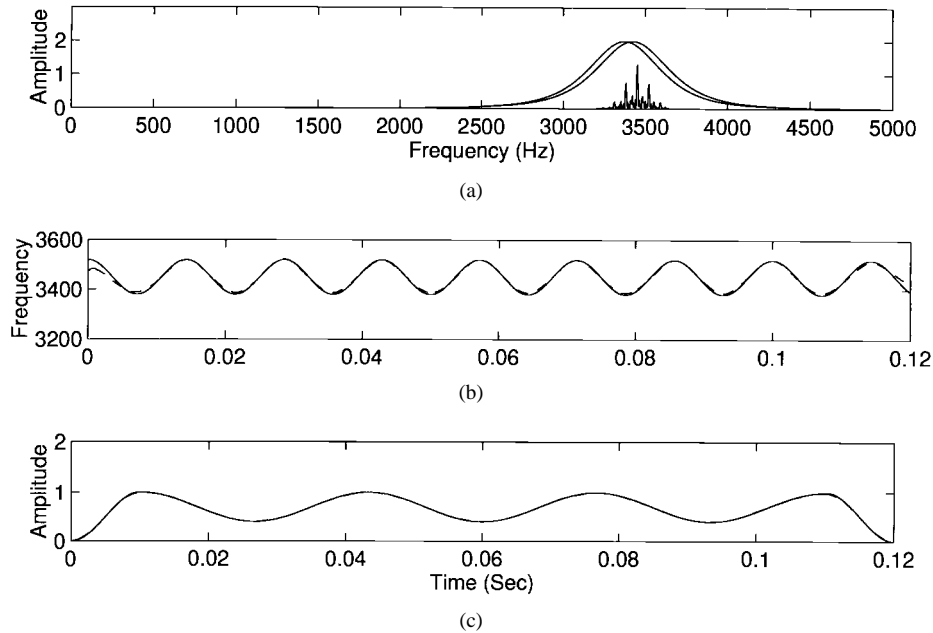


Fig. 13. Example of AM-FM estimation using gammatone filters for 3450 Hz carrier. (a) Frequency response of gammatone-filter pair and AM-FM sine input. (b) Superimposed original (solid) and estimated (dash) frequency. (c) Superimposed original (solid) and estimated (dash) amplitude.

iterative solution procedure, provided that $g(\omega)$ meets our monotonicity condition for invertibility.

One approach to recovering $\omega(n)$ using such filter pairs is to first compute $g(\omega)$ from $H_1(\omega)$ and $H_2(\omega)$ using a large discrete Fourier transform (DFT).¹⁸ Given such a fine frequency sampling of $g(\omega)$, i.e., $g_k = g(k\Delta\omega)$ with small $\Delta\omega$, $\omega(n)$ is obtained through a table look-up procedure followed by iterative refinement. Specifically, for each $v(n) \triangleq (|y_2(n)|/|y_1(n)|)$, we first obtain the g_k closest to $v(n)$, denoted by g_k^* . An estimate of $\omega(n)$ is expressed as $\hat{\omega}(n) \triangleq (g_k^*)^{-1}$ which can be refined by iteration. In Newton's method [20], in particular, we first form the function $f(\omega) = g(\omega) - v(n)$ from which we obtain the derivative $\dot{f}(\omega) = (d/d\omega)f(\omega)$. The refined frequency estimate is then obtained as $\omega_{k+1} = \omega_k - [f(\omega_k)/\dot{f}(\omega_k)]$ where $f(\omega)$ and $\dot{f}(\omega_k)$ are estimated for noninteger frequencies by interpolation.

FM estimation by table look-up, followed by ten passes of the Newton iterative refinement, was applied to the AM-FM sine of the form of Fig. 5 with a carrier of 3450 Hz and 70 Hz FM. Two gammatone transduction filters with characteristic frequencies at 3370 Hz and 3420 Hz [see Fig. 13(a)] were applied. The function $g(\omega)$ was obtained by computing $H_1(\omega)$ and $H_2(\omega)$ with an 8192-point DFT, and was found to be monotonic, and thus invertible, within the overlapping frequency range. This monotonicity property was found empirically to hold generally for closely spaced gammatone-filter pairs. The AM-FM estimates are shown in Figs. 13(b) and (c). The sensitivity measure corresponding to $g(\omega)$ in this case was found to be minimum in the region near 3450 Hz. Indeed, the Newton iteration used in Fig. 13 did not noticeably reduce the

error in the AM-FM estimates when compared with the coarser initial table-lookup procedure. However, as the carrier moves away from 3450 Hz, the sensitivity to quantization introduced by the table look-up procedure increases, and as a consequence the Newton iteration gives a noticeable error reduction.

Finally, from our results with decreasing bandwidth, we expect gammatone filters of small characteristic frequency to be more prone to estimation error in the separation process due to their smaller bandwidth. An example is shown in Fig. 14 with filter characteristic frequencies 334 Hz and 368 Hz. The input signal has a 350 Hz carrier and 25 Hz FM. The function $g(\omega)$ and the sensitivity measure $S(\omega)$, similar to those at 3370 Hz and 3420 Hz of the above high-frequency gammatone-filter pair, have minimum sensitivity about midway between the two characteristic frequencies.¹⁹

B. Measured Auditory Filters

In previous sections, the transduction mechanism was illustrated with piecewise-linear, Gaussian, or gammatone-filter spectral shape. In this section, FM-to-AM transduction is demonstrated with filters derived from measured auditory-nerve tuning curves. The tuning curves represent the threshold of auditory-nerve discharge at different characteristic frequencies, with sound pressure level as a function of frequency, and were obtained from measurements on cats [21], warped to fall within the human frequency range. Auditory-filter impulse responses were obtained by first inverting each tuning curve to form a spectral magnitude. Because the magnitude values are measured on a logarithmic scale, the values were then interpolated to 512 uniform samples over a 5000 Hz bandwidth. A zero-phase function was attached to the spectrum and a 1024-

¹⁸Zero phase is attached to the filters $H_1(\omega)$ and $H_2(\omega)$. This phase function is motivated in Section VI-B, where this same zero-phase selection is made for measured auditory filters.

¹⁹Although $S(\omega)$ has this property for our two gammatone-filter examples, this property does not hold for arbitrary gammatone-filter pairs. Monotone decreasing and increasing sensitivity measures were also observed.

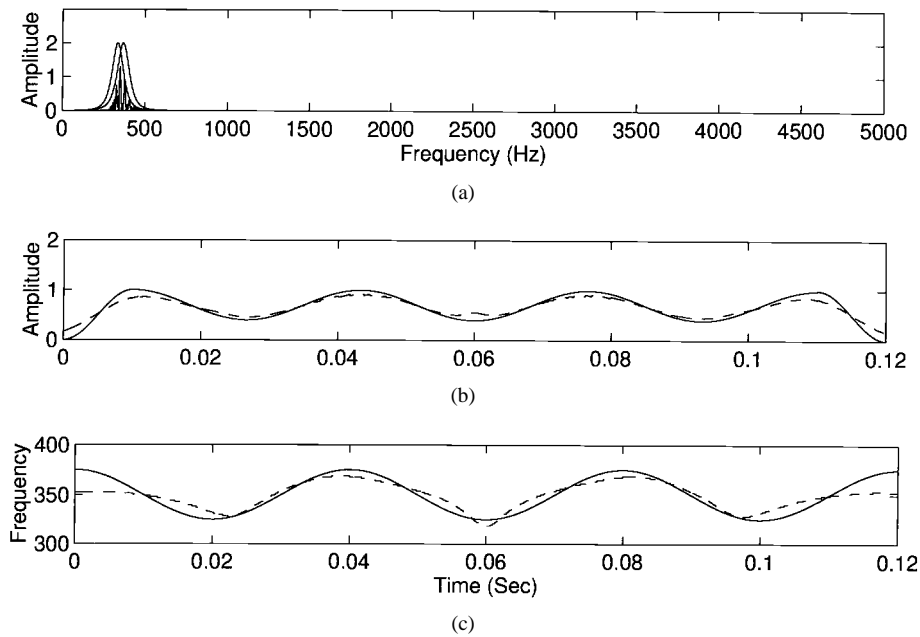


Fig. 14. Example of AM-FM estimation using gammatone filters for 350 Hz carrier. (a) Frequency response of gammatone-filter pair and AM-FM sine input. (b) Superimposed original (solid) and estimated (dash) frequency. (c) Superimposed original (solid) and estimated (dash) amplitude.

point inverse DFT applied to obtain an impulse response.²⁰ The filter characteristic frequencies are nonuniformly spaced, being finely spaced in the low-frequency end, e.g., 10 Hz spacing, and more widely spaced in the high-frequency end, e.g., 200 Hz spacing.

In one example, AM-FM separation was performed using a measured auditory-filter pair with characteristic frequencies of 998 Hz and 1056 Hz. The input sine has carrier of 1050 Hz with a 60 Hz FM. The filter frequency responses are shown in Fig. 15(a) where the measurements are seen to be more skewed than those of the gammatone filters.²¹ The function $g(\omega)$, computed with an 8192-point DFT, was found to be monotonic over the region in which the two filters overlap, and thus is invertible. The table look-up procedure was applied, and the resulting AM-FM estimates are given in Fig. 15(b) and (c). Increasing the modulation frequency above 60 Hz quickly resulted in significant error. In other examples where the characteristic frequencies of the auditory-filter pair were lowered, the allowed FM for accurate estimation decreases, as expected because the filter bandwidth decreases.

VII. TWO-DIMENSIONAL GENERALIZATION

In this section, a two-dimensional (2-D) generalization of the AM-FM separation method is described. Such an approach may enhance methods that rely on 2-D channel filters for the measurement of 2-D signal frequency that give image orientation, roughness, and flow patterns [22]. One expression

²⁰The zero-phase response is localized about the time origin, and is amenable to small transduction approximation error. Furthermore, we have shown empirically that attaching a minimum-phase construction to the frequency response increases error in the AM and FM estimates (see Appendix A). Discarding the phase, however, does not of course imply that the auditory system makes no use of the phase of auditory filters.

²¹Auditory filters may be more accurately modeled by a chirped gammatone. The addition of a chirp is responsible for the spectral skewness [17].

for a 2-D discrete-time AM-FM sine wave is given by

$$x(n, m) = A(n, m) \cos[\omega_{c_1} n + I_1 \sin(\omega_{m_1} n) + \omega_{c_2} m + I_2 \sin(\omega_{m_2} m)] \quad (21)$$

where ω_{c_1} and ω_{m_1} are the carrier and modulation frequencies for the first dimension, ω_{c_2} and ω_{m_2} are the carrier and modulation frequencies for the second dimension, $A(n, m)$ is the time-varying amplitude, and I_1 and I_2 are the indices of modulation. The instantaneous frequency $\omega_1(n)$ is defined as the phase derivative, i.e., $\omega_1(n) = \omega_{c_1} + I_1 \omega_{m_1} \cos(\omega_{m_1} n)$, and likewise for the instantaneous frequency $\omega_2(m)$. Consider the class of 2-D filters $h(n, m)$ with frequency response that is zero for $\omega_1 < 0$ and $\omega_2 < 0$, i.e., $H(\omega_1, \omega_2) = 0$ for $\omega_1 < 0$ and $\omega_2 < 0$. The output of a filter $h(n, m)$ to the input sequence $x(n, m)$ can be approximated by $y(n, m) = x(n, m)H[\omega_1(n), \omega_2(m)]$ under conditions analogous to the 1-D case of Section II.

For 2-D frequency responses $H_1(\omega_1, \omega_2)$ and $H_2(\omega_1, \omega_2)$, generalizing (13), the following equation pair is written:

$$\begin{aligned} |y_1(n, m)| &= A(n, m) |H_1[\omega_1(n), \omega_2(m)]| \\ |y_2(n, m)| &= A(n, m) |H_2[\omega_1(n), \omega_2(m)]| \end{aligned} \quad (22)$$

and thus

$$\frac{|H_2[\omega_1(n), \omega_2(m)]|}{|H_1[\omega_1(n), \omega_2(m)]|} = \frac{|y_2(n, m)|}{|y_1(n, m)|} = v(n, m)$$

Defining

$$g[\omega_1(n), \omega_2(m)] \triangleq \frac{|H_2[\omega_1(n), \omega_2(m)]|}{|H_1[\omega_1(n), \omega_2(m)]|}. \quad (23)$$

An estimate of $\omega_1(n)$ and $\omega_2(m)$ can be obtained by solving $g[\omega_1(n), \omega_2(m)] = v(n, m)$. The amplitude envelope can then be obtained through (22). The amplitude of two filter outputs, however, may not be sufficient to estimate the 2-D AM and

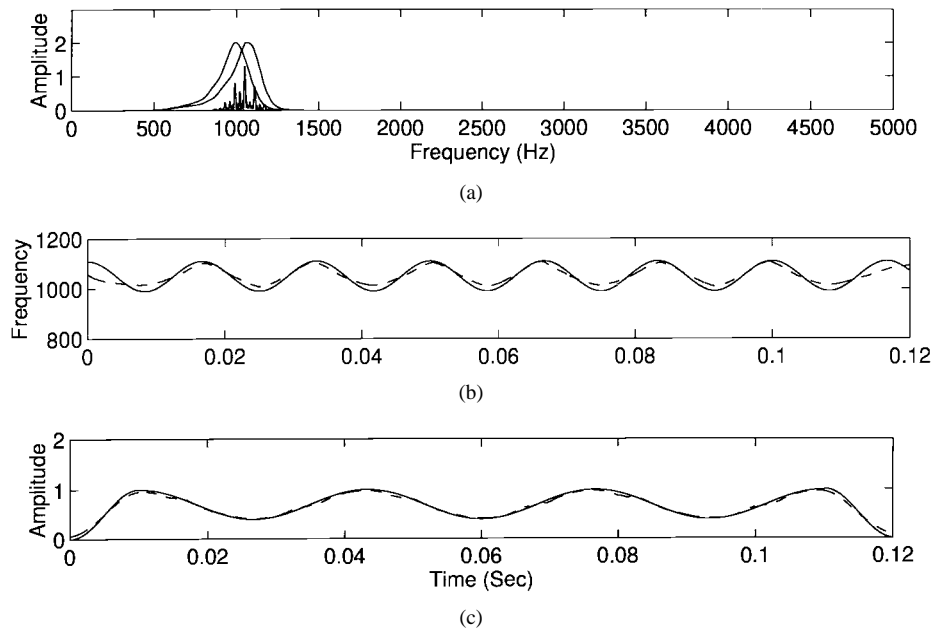


Fig. 15. Example of AM-FM estimation using measured auditory filters for 1050 Hz carrier. (a) Frequency response of measured auditory-filter pair and AM-FM sine input. (b) Superimposed original (solid) and estimated (dash) frequency. (c) Superimposed original (solid) and estimated (dash) amplitude.

FM; a third filter output amplitude may be required as shown in the following example.

Consider two 2-D separable Gaussian filters of the form $H_1(\omega_1, \omega_2) = e^{-a(\omega_1 - \mu_1)^2} e^{-a(\omega_2 - \mu_2)^2}$ and $H_2(\omega_1, \omega_2) = e^{-a(\omega_1 - \nu_1)^2} e^{-a(\omega_2 - \nu_2)^2}$. Then with some tedious algebra, the relation $\log[g(\omega_1, \omega_2)] = \log[|y_2(n, m)|/|y_1(n, m)|]$ can be written as

$$(\nu_1 - \mu_1)\omega_1(n) + (\nu_2 - \mu_2)\omega_2(m) = p(n, m)$$

which represents one equation in two unknowns. Thus, a third filter output amplitude $|y_3(n, m)| = A(n, m)|H_3[\omega_1(n), \omega_2(m)]|$ with $H_3(\omega_1, \omega_2) = e^{-a(\omega_1 - \mu_3)^2} e^{-a(\omega_2 - \mu_3)^2}$ when paired with $H_2(\omega_1, \omega_2)$, leads to the relation

$$(\nu_2 - \mu_2)\omega_1(n) + (\nu_3 - \mu_3)\omega_2(m) = q(n, m)$$

which can be solved for $\omega_1(n)$ and $\omega_2(m)$ through

$$\begin{bmatrix} (\nu_1 - \mu_1) & (\nu_2 - \mu_2) \\ (\nu_2 - \mu_2) & (\nu_3 - \mu_3) \end{bmatrix} \begin{bmatrix} \omega_1(n) \\ \omega_2(m) \end{bmatrix} = \begin{bmatrix} p(n, m) \\ q(n, m) \end{bmatrix} \quad (24)$$

where $p(n, m)$ and $q(n, m)$ are functions of $\log[|y_2(n, m)|/|y_1(n, m)|]$ as well as of the filters' center frequencies and bandwidth. This relation can be solved whenever the matrix on the right side of (24) is invertible, thus imposing a certain structure on the placement of the Gaussian filters and their relation to one another. For example, the center frequencies of $H_1[\omega_1(n), \omega_2(m)]$ and $H_2[\omega_1(n), \omega_2(m)]$ cannot both fall on a 45° line, i.e., $\nu_1 - \mu_1 \neq 0$ and $\nu_2 - \mu_2 \neq 0$. Additionally, if the three Gaussian filters fall on a radial line, i.e., $(\nu_2 - \mu_2) = a(\nu_1 - \mu_1)$ and $(\nu_3 - \mu_3) = b(\nu_1 - \mu_1)$, then imposing the condition that the determinant of the matrix of (24) not equal zero, a solution does not exist when $b = a^2$. Conditions for a unique solution are therefore generally more complex than in the 1-D case.

VIII. DISCUSSION

This paper has described an approach to AM-FM estimation based on FM-to-AM transduction. The generalization from piecewise-linear filters focused first on the class of Gaussian filters for which a closed-form solution to AM-FM separation exists. For gammatone and measured auditory filters, where a closed-form solution was not possible, AM-FM separation was achieved by a table look-up procedure and iterative refinement. Properties of the approach were described, including solution stability, error evaluation, and comparative performance with two standard methods.

Further properties and generalizations of the method have yet to be addressed. These include a more thorough analysis of the relation of temporal resolution to transduction filter parameters, a better understanding of the importance of filter phase in the transduction process, further analysis and reduction of the dependence of "smoothness" assumptions on $A(n)$ and $\omega(n)$ under which $y(n) = x(n)H(\omega)$, and an extension to a more general class of filters and inputs. Design of filter pairs that minimize solution sensitivity measure $S(\omega)$ with respect to place of transduction, further generalizations to two dimensions, and improved robustness by increasing the number of transduction filters are other fascinating areas being examined. For example, robustness of the FM transduction method in noise was improved by averaging AM-FM estimates derived from multiple piecewise-linear filter pairs; with 25 filter pairs about a 3 dB improvement in SNR was obtained in estimating a 70 Hz modulation around a 2000 Hz carrier at a 20 dB SNR.

The approach of this paper was motivated by the hypothesis that the auditory system exploits an FM transduction mechanism [1]. It is interesting to conjecture, therefore, on how the algorithmic results of this paper might be interpreted in the context of auditory signal transduction and measurement

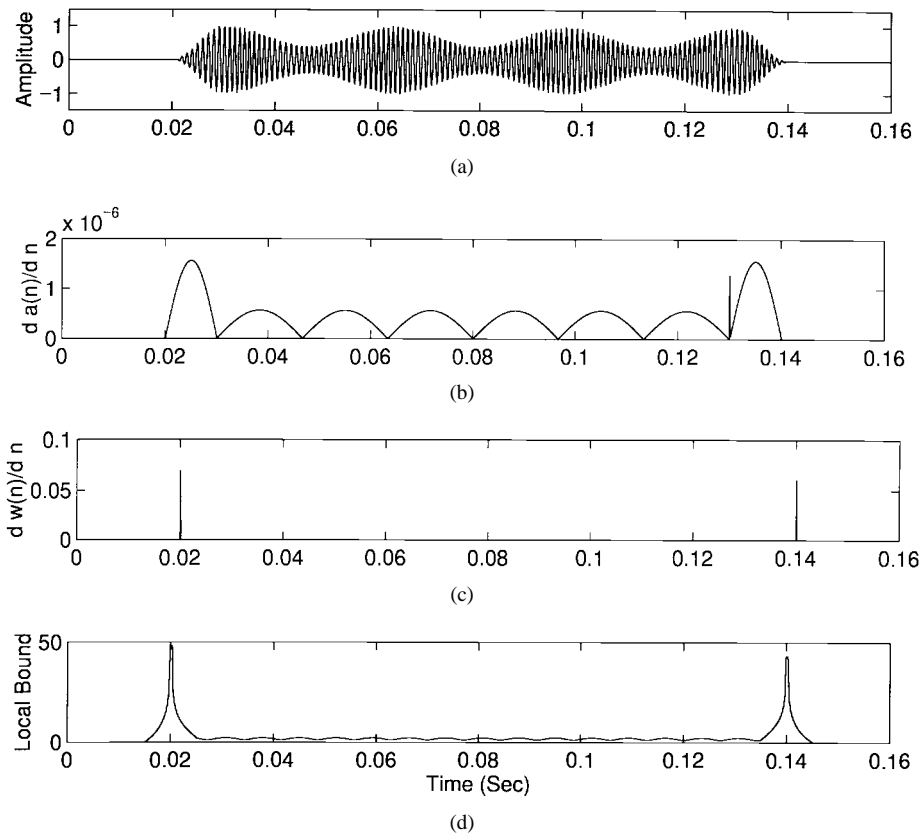


Fig. 16. Example of local error bound of FM-to-AM transduction. (a) Waveform. (b) Amplitude envelope derivative. (c) Frequency modulation derivative. (d) Local error bound.

of modulation. One such speculation is that the amplitude envelope of an auditory filter's output is related to the "energy" required to generate the input sine wave. This interpretation can be argued by setting the offset term $b = 0$ in (8) and observing that the result is proportional to the *product* of the input sine-wave amplitude and frequency. This product when squared can be shown to be the energy, i.e., the sum of potential and kinetic energy, stored by a harmonic oscillator required to generate an input sine [12], which may provide a robust representation for further auditory stages. A second example is the differencing or dividing of two filter amplitude envelope outputs, which may also have relevance in the auditory context since differences across auditory filters may be exploited in enhancing spectral resolution [3], [23]. As such, in further relating the AM-FM estimator to the auditory mechanism, one may wish to compare the accuracy of the estimator with human psychophysical performance.

In spite of these intriguing possibilities, it is important to emphasize, as we have done in the introduction, that the use of FM-to-AM transduction in aural perception is speculative, the motivating experiments [1] indicating that this possibility provides simply one candidate mechanism for auditory FM demodulation. Even if this transduction is exploited, the mechanism may not be a linear one as this paper suggests. For example, the demodulation assumes that the filter shapes are constant. In the cochlea, however, there is a fast-acting automatic gain control (AGC) that provides compression in the main part of the filter passband (the tip of

the filter), while leaving the gain in the low-frequency portion of the filter (the tail) unaffected. The compression may be the result of an almost instantaneous saturating nonlinearity in the cochlear mechanics [24], and could both introduce fluctuations in the system output for a low-frequency sinusoid with constant amplitude and reduce the fluctuations in a AM-FM modulated tone. Incorporating the influence of such a nonlinearity will be essential in understanding the full complexity of transduction by the auditory system.

APPENDIX A ERROR BOUNDS ON FM TRANSDUCTION

In this Appendix, error bounds are given on the approximation (4) that is the output of the transduction filter $H(\omega)$ to a modulated complex sine input²² with amplitude envelope $A(n)$ and instantaneous frequency $\omega(n)$, i.e., $x(n) = A(n)e^{j\phi(n)} = A(n)e^{j[\int_0^n \omega(\tau) d\tau]}$. When no modulation is present, i.e., the amplitude and frequency modulation are constant, then $x(n) = Ae^{j\omega_c n}$, and $x(n)$ is an eigenfunction to the linear filter, the output being $s(n) = Ae^{j\omega_c n} H(\omega_c)$ [8]. For an input $x(n)$ with a time-varying amplitude and frequency, we expect a similar relation when the modulating amplitude and frequency are "slowly-varying" and the filter is of "short duration." Under these conditions, the modulated input will appear as a sine of approximately constant amplitude and frequency, $x(n)$ being

²²The simplifying case considered in the examples throughout this paper is $x(n) = A(n)e^{j[\omega_c n + I \sin(\omega_m n)]}$. The bounds given in this appendix, however, apply to a more general signal class.

referred to as a “pseudoeigenfunction” of $H(\omega)$. The output $y(n)$ can therefore be expressed as

$$\begin{aligned} y(n) &= A(n)e^{j\phi(n)}H(\omega(n)) \\ &\approx s(n) = h(n) * x(n), \end{aligned} \quad (25)$$

In order to state quantitatively the conditions under which the approximation is valid [13], we first define the functions

$$\begin{aligned} \Delta_k(h) &= \left[\sum_n n^{2k} |h(n)|^2 \right]^{1/2} \\ &= \left[\frac{1}{2\pi} \int_{-\pi}^{\pi} \left| \frac{d^k}{d\omega^k} H(\omega) \right|^2 d\omega \right]^{1/2} \end{aligned} \quad (26)$$

and

$$D(z) = \int_{-\infty}^{\infty} |\dot{z}(\tau)| d\tau. \quad (27)$$

The function $\Delta_k(h)$ measures the extent, i.e., the “localization” of $h(n)$, while the function $D(z)$ measures the “smoothness” of the signal $z(n)$. Given the approximation error as $e(n) = |y(n) - s(n)|$, an error bound can be expressed in terms of these functions as

$$e(n) \leq \frac{\pi}{\sqrt{3}} [\Delta_1(h)D(A) + A_{\max}\Delta_2(h)D(\omega)] \quad (28)$$

so that the error bound²³ is a function of the time localization of $h(n)$, equivalently the smoothness of $H(\omega)$, and the *global* smoothness of the envelope and frequency trajectories of $x(n)$. Although this measure involves infinite limits on the summation in $D(A)$ and $D(\omega)$, it does provide a meaningful upper error bound, since the signal is typically of finite extent, and even when large in extent, the bound may be small (especially in a comparative sense), as in the case of a sine wave with small frequency or amplitude modulation.

An error bound can also be obtained as a function of time in terms of the *local* smoothness properties of amplitude and frequency. Specifically [13]

$$\begin{aligned} e(n) &\leq \sum_{m \neq 0} |h(m)| \\ &\cdot \sum_{p=n-m}^n (|\dot{A}(p)| + A_{\max}|m|\dot{\omega}(p)) \end{aligned} \quad (29)$$

where the filter localization is reflected in $|h(n)|$ and the input smoothness in $|\dot{A}(n)|$ and $|\dot{\omega}(n)|$. If $h(n)$ is of short duration, then the inner sum in (29) (the limits on the inner sum are a function of the index p away from the center of mass of $h(n)$ has small weight; thus, a local average is taken on amplitude and frequency, rather than the global average of (29). Derivatives $\dot{A}(p)$ and $\dot{\omega}(p)$ in the above equations are assumed samples of continuous counterparts or approximations derived from first and second differences of discrete functions. Further discussion of the error bounds are found in [13].

²³The constant $(\pi/\sqrt{3}) = \sqrt{\sum_{p \neq 0} (1/p^2)}$ and is a consequence of the infinite sum in (26) [13].

An example of a local error bound is illustrated in Fig. 16 for the same sequence $x(n)$ of Fig. 1(a). The instantaneous frequency has carrier $\omega_c = 1000$ Hz with modulation frequency $\omega_m = 70$ Hz. A Gaussian filter of the form $H(\omega) = e^{-a(\omega-\omega_o)^2}$, with center frequency ω_o at 1000 Hz is used as the transduction filter. Fig. 16 shows that, as predicted from (29), the error bound is largest where the amplitude and frequency derivatives are largest, i.e., at the signal edges.

Observe that both the magnitude and the phase of $H(\omega)$ play a role in determining the validity of the transduction approximation. From the expression for $\Delta_k(h)$, we see that the presence of phase can increase the moments of $h(n)$ about the origin, and thus increase the error bound. Therefore, for the study of this paper, we have constrained the class of filters to be zero-phase, having the property of being localized about the time origin. For example, the minimum-phase counterparts to the zero-phase gammatone and auditory transduction filters used in this paper were found to increase the AM-FM estimation error, even though the magnitude of the transduction approximation is used, phase being discarded. This is consistent with our measurement that the associated impulse responses have larger second moments than their zero-phase counterparts, and thus contribute to a larger error bound. Additional study is needed to further understand the importance of filter phase in the transduction process.

APPENDIX B

BALANCED FREQUENCY DISCRIMINATION

In an early method of FM demodulation introduced by Armstrong [6] for radio broadcasting, FM-to-AM transduction, approximately linear, was applied to a flattened waveform; differencing the output of two complementary filters was used to eliminate undesired bias in the FM estimate. This method of FM estimation is referred to as a *balanced frequency discriminator*.

In a discrete-time analog to the continuous-time balanced frequency discriminator [6], [14], two piecewise-linear filters are used with complementary positive- and negative-sloped regions of the form

$$\begin{aligned} H_1(\omega) &= a(\omega - \omega_c) + b & \omega \in H^+ \\ H_2(\omega) &= -a(\omega - \omega_c) + b & \omega \in H^- \end{aligned} \quad (30)$$

where $H^+ = H^-$ denote the frequency intervals over which $H_1(\omega)$ and $H_2(\omega)$ have positive and negative slopes, respectively, and where ω_c is the carrier frequency of the input sine, which is assumed to have constant amplitude A and time-varying input frequency $\omega(n)$. To eliminate the “bias” term b , the difference in the amplitude of the output of the two filters is taken and reduces to

$$\begin{aligned} \Delta[\omega(n)] &= H_1(\omega(n)) - H_2(\omega(n)) \\ &= A[a(\omega(n) - \omega_c) + b - (-a(\omega(n) - \omega_c) + b)] \\ &= 2Aa(\omega(n) - \omega_c) \\ &= 2Aa\tilde{\omega}(n) \end{aligned} \quad (31)$$

where $\tilde{\omega}(n) = \omega(n) - \omega_c$ is the desired modulation frequency about the carrier ω_c and which represents the message signal

in FM broadcasting. When $A(n)$ is time-varying, then it can be estimated from the sum $H_1(\omega) + H_2(\omega) = 2bA(n)$, which can then be used to separate $\omega(n)$ from $\Delta[\omega(n)]$ of (31).

ACKNOWLEDGMENT

The authors thank the reviewers for their detailed comments on the manuscript.

REFERENCES

- [1] K. Saberli and E. R. Hafter, "A common neural code for frequency- and amplitude-modulated sounds," *Nature*, vol. 374, pp. 537–539, Apr. 1995.
- [2] J. L. Goldstein, "Auditory spectral filtering and monaural phase perception," *J. Acoust. Soc. Amer.*, vol. 41, pp. 458–479, 1967.
- [3] R. H. McEachern, "How the ear really works," *IEEE Int. Symp. Time-Frequency and Time-Scale Analysis*, Victoria, B.C., Oct. 1992, pp. 437–440.
- [4] J. M. Kates, "Accurate tuning curves in a cochlear model," *IEEE Trans. Speech Audio Processing*, vol. 1, pp. 453–462, Oct. 1993.
- [5] R. F. Lyon, "The all-pole gammatone filter and auditory models," in *Proc. Computational Models of Signal Processing in the Auditory System, Forum Acusticum '96*, Antwerp, Belgium.
- [6] E. H. Armstrong, "A method of reducing disturbances in radio signaling by a system of frequency modulation," *Proc. Inst. Radio Eng.*, vol. 24, pp. 689–740, May 1936.
- [7] B. W. Edwards and N. F. Viemeister, "Psychoacoustic equivalence of frequency modulation and quasifrequency modulation," *J. Acoust. Soc. Amer.*, vol. 95, pp. 1510–1513, Mar. 1994.
- [8] A. V. Oppenheim and R. W. Schaffer, *Discrete-Time Signal Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1989.
- [9] L. Cohen, *Time-Frequency Analysis*. Englewood Cliffs, NJ: Prentice-Hall, 1995.
- [10] B. Boashash, "Estimating and interpreting the instantaneous frequency of a signal," *Proc. IEEE*, vol. 80, pp. 519–568, Apr. 1992.
- [11] P. Maragos, J. F. Kaiser, T. F. Quatieri, "Energy separation in signal modulations with application to speech analysis," *IEEE Trans. Signal Processing*, vol. 41, pp. 3024–3051, Oct. 1993.
- [12] J. F. Kaiser, "On a simple algorithm to calculate the 'energy' of a signal," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, Albuquerque, NM, Apr. 1990, pp. 381–384.
- [13] A. C. Bovik, J. P. Havlicek, M. D. Desai, and D. S. Harding, "Limits on discrete modulated signals," *IEEE Trans. Signal Processing*, vol. 45, pp. 867–879, Apr. 1997.
- [14] S. Haykin, *Communication Systems*. New York: Wiley, 1978.
- [15] A. Potamianos and P. Maragos, "A comparison of the energy operator and the Hilbert transform approach to signal and speech demodulation," *Signal Process.*, vol. 37, pp. 95–120, May 1994.
- [16] R. D. Patterson, "Auditory filter shape," *J. Acoust. Soc. Amer.*, vol. 55, pp. 802–809, 1974.
- [17] I. Toshio, "An optimal auditory filter," in *Proc. 1995 IEEE Acoustics, Speech, and Signal Processing Workshop on Applications of Signal Processing to Audio and Acoustics*, Oct. 1995.
- [18] M. Slaney, "Auditory Toolbox: A matlab toolbox for auditory modeling work," Tech. Rep. 45, Apple Computer, Inc., 1993.
- [19] R. D. Patterson, I. Nimmo-Smith, D. L. Weber, and R. Milroy, "The deterioration of hearing with age: frequency selectivity, the critical ratio, the audiogram, and speech threshold," *J. Acoust. Soc. Amer.*, vol. 72, pp. 1788–1803, Dec. 1982.
- [20] R. W. Hamming, *Numerical Methods for Scientists and Engineers*, 2nd ed. New York: McGraw-Hill, 1972.
- [21] M. C. Liberman, "Auditory-nerve response from cats raised in a low-noise chamber," *J. Acoust. Soc. Amer.*, vol. 63, pp. 442–455, Feb. 1978.
- [22] A. C. Bovik, N. Gopal, T. Emmoth, and A. Restrepo, "Localized measurement of emergent image frequencies by Gabor wavelets," *IEEE Trans. Inform. Theory*, vol. 38, pp. 691–712, Mar. 1992.
- [23] X. Yang, K. Wang, and S.A. Shamma, "Auditory representations of acoustic signals," *IEEE Trans. Inform. Theory*, vol. 38, pp. 824–839, Mar. 1992.
- [24] P. M. Sellick, R. Patuzzi, and B. M. Johnstone, "Measurement of basilar membrane motion in the guinea pig using the Mössbauer technique," *J. Acoust. Soc. Amer.*, vol. 72, pp. 1788–180, July 1982.



Thomas F. Quatieri (S'73–M'79–SM'87) was born in Somerville, MA, on January 31, 1952. He received the B.S. degree (summa cum laude) from Tufts University, Medford, MA, in 1973, and the S.M., E.E., and Sc.D. degrees from the Massachusetts Institute of Technology (MIT), Cambridge, in 1975, 1977, and 1979, respectively. He is currently a senior research staff member at MIT Lincoln Laboratory, Lexington.

In 1980, he joined the Sensor Processing Technology Group of MIT Lincoln Laboratory, where he worked on problems in multidimensional digital signal processing and image processing. Since 1983, he has been a member of the Speech Systems Technology Group, Lincoln Laboratory, where he has been involved in digital signal processing for speech and audio applications, underwater sound enhancement, and data communications. He has contributed many publications to journals and conference proceedings, written several patents, and co-authored chapters in numerous edited books, including *Advanced Topics in Signal Processing* (Englewood Cliffs, NJ: Prentice-Hall, 1987), *Advances in Speech Signal Processing* (New York: Marcel Dekker, 1991), and *Speech Coding and Synthesis* (Amsterdam, The Netherlands: Elsevier, 1995). He holds the position of Lecturer at MIT where he has developed the graduate course in digital speech processing, and is active in advising graduate students on the MIT campus.

Dr. Quatieri is the recipient of the 1982 Paper Award of the IEEE Acoustics, Speech and Signal Processing Society for the paper, "Implementation of 2-D Digital Filters by Iterative Methods." In 1990, he received the IEEE Signal Processing Society's Senior Award for the paper, "Speech Analysis/Synthesis Based on a Sinusoidal Representation," published in the IEEE TRANSACTIONS ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING, and in 1994 won this same award for the paper "Energy Separation in Signal Modulations with Application to Speech Analysis," which was also selected for the 1995 IEEE W.R.G. Baker Prize Award. He was a member of the IEEE Digital Signal Processing Technical Committee. From 1983 to 1992, he served on the steering committee for the biannual Digital Signal Processing Workshop, and was Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING in the area of nonlinear systems. He is a member of Tau Beta Pi, Eta Kappa Nu, Sigma Xi, and the Acoustical Society of America.



Thomas E. Hanna was born in Chicago, IL, in 1954. He received the B.A. and B.S. degrees in mathematics and psychology from the University of Illinois, Champaign, in 1978, and the Ph.D. degree in experimental psychology from Indiana University, Bloomington, in 1982.

In 1985, he joined the Auditory and Communication Sciences Department, Naval Submarine Medical Research Laboratory, Groton, CT, where he worked in the areas of auditory detection and recognition for sonar applications. He currently heads the Submarine Systems Department and is involved in work on sonar, diver hearing, and hearing conservation.

Dr. Hanna is a member of the Acoustical Society of America and the Psychonomic Society.



Gerald C. O'Leary received the B.S., M.S., and E.E. degrees from the Massachusetts Institute of Technology (MIT), Cambridge, in 1963, 1964, and 1966, respectively.

From 1966 to 1971, he was a member of the Technical Staff, Advanced Techniques Radar Department, MITRE Corp., Bedford, MA. From 1972 to 1977, he was employed by Signal Processing Systems, Inc., Waltham, MA, developing applications for programmable signal processors. In 1977, he joined the Communications Satellite Division, MIT Lincoln Laboratory, Lexington. In 1978, he joined the Speech Systems Technology Group, Lincoln Laboratory, where he is now the Associate Group Leader.