1992 International Conference on
ACOUSTICS, SPEECH, and SIGNAL PROCESSING
March 23-26, 1992
San Francisco Marriott, San Francisco, California, USA

ON SEPARATING AMPLITUDE FROM FREQUENCY MODULATIONS USING ENERGY OPERATORS

Petros Maragos,[1] *James F. Kaiser,*[2] *and Thomas F. Quatieri* [3]

[1]Division of Applied Sciences, Harvard University, Cambridge, MA 02138, USA
[2]Bellcore, 445 South St., Morristown, NJ 07962-1910, USA
[3]MIT Lincoln Laboratory, 244 Wood St., Lexington, MA 02173, USA

ABSTRACT

To estimate the amplitude envelope and instantaneous frequency of an AM–FM signal we develop a novel approach that uses nonlinear combinations of instantaneous signal outputs from an energy-tracking operator to separate its output energy product into its amplitude modulation and frequency modulation components. This energy separation algorithm is then applied to search for modulations in speech resonances, which we model using AM–FM signals. Our theoretical and experimental results demonstrate that the energy separation algorithm, due to its low computational complexity and instantaneously-adapting nature, is very useful in detecting modulation patterns in speech and other time-varying signals.

1. INTRODUCTION

Oscillatory signals that have both an amplitude-modulation (AM) and a frequency-modulation (FM) structure are encountered in almost all communication systems. We have used such real-valued AM–FM signals

$$x(t) = a(t)\cos[\phi(t)] = a(t)\cos(\omega_c t + \omega_m \int_0^t q(\tau)d\tau + \theta)$$

to model time-varying amplitude and frequency patterns in speech resonances [6]. Note that $x(t)$ is a cosine of carrier frequency ω_c with a time-varying amplitude signal $a(t)$ and a time-varying *instantaneous angular frequency* signal

$$\omega_i(t) \triangleq \frac{d\phi}{dt}(t) = \omega_c + \omega_m q(t),$$

where $|q(t)| \leq 1$, $\omega_m \in [0, \omega_c]$ is the maximum frequency deviation, and θ is a constant phase offset.

In this paper we develop an efficient approach to estimating the time-varying amplitude envelope $|a(t)|$ and instantaneous frequency $\omega_i(t)$ of an arbitrary AM–FM signal, based on the energy-tracking operator

$$\Psi_c[x(t)] \triangleq \left(\frac{dx}{dt}(t)\right)^2 - x(t)\frac{d^2x}{dt^2}(t) = [\dot{x}(t)]^2 - x(t)\ddot{x}(t)$$

where $\dot{x} = dx/dt$, and its discrete-time counterpart

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

for discrete-time signals $x(n)$, $n = 0, \pm1, \pm2, \ldots$. These *energy* operators were developed by Teager [1, 2] in his work on modeling speech production and were first introduced by Kaiser [3, 4]. When Ψ_c is applied to signals produced by simple harmonic oscillators, it can track the oscillator's energy (per half unit mass), which is equal to the squared product of the oscillation amplitude and frequency. The energy operators are also very useful for analyzing oscillatory signals with time-varying amplitude and frequency. Specifically, we have shown [5, 6] that Ψ_c applied to an AM–FM signal can approximately estimate the squared product of the amplitude a and instantaneous frequency ω_i signals; i.e.,

$$\Psi_c[a(t)\cos(\int_0^t \omega_i(\tau)d\tau + \theta)] \approx [a(t)\omega_i(t)]^2 \qquad (1)$$

assuming that the signals a and ω_i do not vary too fast (time rate of change of value) or too greatly (range of value) in time compared to the carrier frequency ω_c.

In Section 2 we develop an elegant approach for separating the amplitude from the frequency signal in the output energy product of Ψ_c. We call this the *energy separation algorithm* because an oscillator's energy depends on the product of amplitude and frequency and because energy-tracking operators are used. In Section 3 we develop separation algorithms for discrete-time signals. Section 4 discusses the application of these algorithms to track amplitude and frequency modulations in speech resonances, i.e., signals resulting from bandpass filtering speech vowels around their formants.

2. ENERGY SEPARATION

First we present some closed-formula solutions for exact estimation of the constant amplitude and frequency of a cosine and then show that the same equations approximately apply to an AM–FM signal with time-varying amplitude and frequency. To simplify notation, we henceforth drop the subscripts from the continuous and discrete energy operator symbols and use Ψ for both.

P. Maragos was supported by the National Science Foundation under Grant MIPS-86-58150 with matching funds from Xerox and Sun, and in part by the ARO Grant DAAL03-86-K-0171 to the Brown-Harvard-MIT Center for Intelligent Control Systems. T.F. Quatieri was sponsored in part by the MIT Lincoln Laboratory Innovative Research Program and in part by the Naval Submarine Medical Research Laboratory, Department of the Navy under Air Force Contract F19628-90-C-0002.

COSINE: Consider a cosine $x(t)$ with constant amplitude A and frequency $\omega_c \geq 0$ and its derivative

$$x(t) = A\cos(\omega_c t + \theta) \;\; ; \;\; \dot{x}(t) = -A\omega_c \sin(\omega_c t + \theta)$$

Then applying Ψ yields

$$\Psi[x(t)] = A^2\omega_c^2 \quad ; \quad \Psi[\dot{x}(t)] = A^2\omega_c^4$$

Hence the constant frequency and the absolute amplitude of the cosine can be obtained from the equations

$$\omega_c = \sqrt{\frac{\Psi[\dot{x}(t)]}{\Psi[x(t)]}} \quad ; \quad |A| = \frac{\Psi[x(t)]}{\sqrt{\Psi[\dot{x}(t)]}}$$

AM–FM SIGNALS: Let $x(t) = a(t)\cos[\phi(t)]$ be an AM–FM signal. Its derivative is

$$\dot{x}(t) = \underbrace{\dot{a}(t)\cos[\phi(t)]}_{y_1} - \underbrace{a(t)\omega_i(t)\sin[\phi(t)]}_{y_2}$$

To make (1) a valid approximation and the analysis of $\Psi(x), \Psi(\dot{x})$ tractable, we henceforth assume that

(CA1): a and q are bandlimited with highest frequencies ω_a and ω_f respectively, and $\omega_a, \omega_f \ll \omega_c$.

(CA2): $\omega_a^2 + \omega_m\omega_f \ll (\omega_c + \omega_m)^2$.

Further, if we define the *order of magnitude of a signal* z to be $O(z) = O(z_{max})$ where $z_{max} = \max_t |z(t)|$, then $O(y_1) \approx O(a\omega_a)$ and $O(y_2) \approx O(a\omega_i)$. (See [5, 7] for details and proofs.) Since $O(\omega_a) \ll O(\omega_i)$, by ignoring y_1 we obtain the approximation

$$\Psi[\dot{x}(t)] \approx \Psi[a(t)\omega_i(t)\sin\phi(t)] \approx a^2(t)\,\omega_i^4(t) \tag{2}$$

By combining (2) and (1) we obtain

$$\sqrt{\frac{\Psi[\dot{x}(t)]}{\Psi[x(t)]}} \approx \omega_i(t) \quad ; \quad \frac{\Psi[x(t)]}{\sqrt{\Psi[\dot{x}(t)]}} \approx |a(t)|$$

This is the *continuous energy separation algorithm (CESA)*. At each time instant it estimates the instantaneous frequency and the amplitude envelope by using only the two instantaneous output values from the energy operator applied to the signal and its derivative. Although we derived the CESA by assuming bandlimited modulating signals, there are also other special cases of AM–FM signals, where the CESA will still yield approximately correct solutions. Examples include FM/Linear (chirp) signals $A\cos(\omega_c t + \omega_m t^2/2L + \theta)$, $0 \leq t \leq L$, whose amplitude and linear instantaneous frequency can be estimated via the CESA provided that $(\omega_m/L) \ll (\omega_c + \omega_m)^2$.

For the validity of the CESA it is assumed that $\Psi(x), \Psi(\dot{x}) \geq 0$. There are large classes of signals satisfying this condition [7]; e.g., all AM–FM signals whose modulation amounts do not exceed 50% and $\omega_a, \omega_f < \omega_c/10$. During our discrete simulations of the CESA on noise-free AM–FM and bandpass filtered speech signals we rarely encountered negative $\Psi(x)$ values, which appeared to be due to round-off errors. Note also that at times t_0 when $\Psi[x(t_0)] = 0$, we have $a(t_0) \approx 0$ if ω_i is assumed always positive. At such rare time instants we need additional information to estimate $\omega_i(t_0)$. For example, in our discrete simulations we interpolated $\omega_i(t_0)$ from its immediate neighbors.

3. DISCRETE ALGORITHMS

By using the discrete energy operator and approximating derivatives with differences, we derive two algorithms for discrete-time AM–FM signals

$$x(n) = a(n)\cos[\phi(n)] = a(n)\cos(\Omega_c n + \Omega_m \int_0^n q(k)dk + \theta)$$

to estimate their instantaneous frequency

$$\Omega_i(n) = \frac{d\phi}{dn}(n) = \Omega_c + \Omega_m q(n)$$

where $0 \leq \Omega_m \leq \Omega_c$, $|q(n)| \leq 1$, and to estimate their amplitude envelope. For mathematical tractability, we analyze only two classes of such signals: (i) *AM-FM/Cosine* where $q(n) = \cos(\Omega_f n)$, and (ii) *AM-FM/Linear* where $q(n) = n/N$ with $n = 0, 1, .., N$. Note that the continuous-time frequencies ω_c, ω_m, and ω_i have been replaced by their discrete-time counterparts Ω_c, Ω_m, and Ω_i. All discrete-time frequencies are assumed to be in $[0, \pi]$.

3.1. DESA–2

As we showed in [5, 6, 7],

$$\Psi[a(n)\cos(\int_0^n \Omega_i(m)dm + \theta)] \approx a^2(n)\sin^2[\Omega_i(n)] \tag{3}$$

under the assumptions

(DA1): a has bandwidth $\Omega_a \ll \Omega_c$.

(DA2): $8\sin^2[(\Omega_a + \Omega_f)/2] \ll [\sin^2(\Omega_i)]_{max}$.

(DA3): (a) $\Omega_f \ll 1$ for AM–FM/Cosine; or

(b) $(\Omega_m/N) \ll [\sin^2(\Omega_i)]_{max}$ for AM–FM/Linear.

In general $[\sin^2(\Omega_i)]_{max} = \sin^2(\Omega_c + \Omega_m)$ if $\Omega_c + \Omega_m \leq \pi/2$. Now consider the symmetric difference of x

$$s(n) = [x(n+1) - x(n-1)]/2$$

As we showed in [7], $s(n) \approx -a(n)\sin[\Omega_i(n)]\sin[\phi(n)]$, and (if $\Omega_m \ll 1$) its envelope has an effective bandwidth $\leq \Omega_a + \Omega_f$. Hence, by (DA1)–(DA3),

$$\Psi[s(n)] \approx a^2(n)\sin^4[\Omega_i(n)] \tag{4}$$

From (3) and (4), we obtain the following formulas for estimating the time-varying frequency and amplitude envelope:

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

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

This is the DESA–2 algorithm, where '2' implies the approximation of first-order derivatives by differences between samples whose time indices differ by 2. The frequency estimation part assumes that $0 < \Omega_i(n) < \pi/2$ because the computer's implementation of $\arcsin(u)$ function assumes that $|u| \leq \pi/2$. Thus the DESA–2 can be used to estimate instantaneous frequencies $\leq 1/4$ the sampling frequency f_s. Using $2f_s$ as sampling frequency allows the DESA–2 to estimate frequencies up to $f_s/2$. Note that, if $x(n) = A\cos(\Omega_c n + \theta)$, then the DESA–2 formulas yield the exact constant frequency and amplitude.

3.2. DESA–1

An alternative discrete algorithm results if we replace derivatives with backward and forward differences

$$y(n) = x(n) - x(n-1) \quad ; \quad z(n) = x(n+1) - x(n) = y(n+1)$$

By working as for the DESA–2, we showed in [7] that

$$\begin{aligned} \Psi[y(n)] &\approx 4a^2(n)\sin^2[\Omega_i(n-0.5)/2]\sin^2[\Omega_i(n-0.5)] \\ \Psi[z(n)] &\approx 4a^2(n)\sin^2[\Omega_i(n+0.5)/2]\sin^2[\Omega_i(n+0.5)] \end{aligned}$$

By averaging these two results and assuming that the shifts by $+1/2$ and $-1/2$ sample (which correspond to a small error in Ω_i) approximately cancel out, we obtain

$$\frac{\Psi[y(n)] + \Psi[z(n)]}{2} \approx 4a^2(n)\sin^2[\Omega_i(n)/2]\sin^2[\Omega_i(n)] \quad (5)$$

Thus, the action of Ψ on asymmetric differences is partially 'symmetrized' by averaging its action on two opposite differences. Then combining (3) and (5) yields

$$\begin{aligned} 1 - \frac{\Psi[y(n)] + \Psi[y(n+1)]}{4\Psi[x(n)]} &= G(n) \\ \arccos[G(n)] &\approx \Omega_i(n) \\ \sqrt{\frac{\Psi[x(n)]}{1 - G^2(n)}} &\approx |a(n)| \end{aligned}$$

We call this the DESA–1 algorithm, where '1' implies the approximation of derivatives with a single sample difference. The frequency estimation part assumes that $0 < \Omega_i(n) < \pi$. Thus, the DESA–1 algorithm can estimate instantaneous frequencies up to 1/2 the sampling frequency.

In [7] we compared the (mean absolute and rms) errors of the two DESAs in estimating the amplitude and frequency of synthetic AM–FM signals. On the average (for AM and FM amounts of 5%–50%) both DESAs yielded very small errors in the order of 1% or less. While the DESA–1 yielded slightly smaller errors than the DESA–2, the latter is slightly faster and leads to a simpler mathematical analysis. All the experiments in this paper were done using the DESA–1. Figure 1 demonstrates that the DESA performs quite well in approximately estimating the time-varying amplitude and frequency of AM–FM signals despite the large amounts of modulation.

4. SPEECH RESONANCE ANALYSIS

By 'speech resonances', also called 'formants', we loosely refer to the oscillator systems formed by local cavities of the vocal tract emphasizing certain frequencies and de-emphasizing others during speech production. Teager's experimental work provided evidence that speech resonances can change rapidly within a single pitch period, possibly due to the rapidly-varying and separated speech airflow in the vocal tract [1, 2]. It is also known that time variations of the elements of simple harmonic oscillators can result in amplitude or frequency modulation of the simple oscillator's cosine response. The above evidence motivated in our work [6, 7] the modeling of a single speech resonance within a pitch period by an exponentially-damped AM–FM signal $r^n a(n)\cos[\phi(n)]$, where the instantaneous frequency $\Omega_i(n) = \Omega_c + \Omega_m q(n)$ models the deviation of a time-varying formant from its center value Ω_c, $r \in (0,1)$ is related to the rate of energy dissipation, and the time-varying amplitude and frequency modulating signals a, q are cosines or some similar oscillatory signals.

Figure 2(a) shows a segment $s(n)$ of a speech vowel /E/ sampled at 30 kHz. A speech resonance signal $x(n)$ was extracted around a formant at $f_c = 3400$ Hz by convolving $s(n)$ with a bandpass Gabor filter's impulse response $h(t) = \exp(-\alpha^2 t^2)\cos(2\pi f_c t)$ properly discretized and truncated [6] with $\alpha = 1000$. Fig. 2(b) shows the estimated (via DESA) amplitude signal, where we see a strong AM modulation and two pulses per pitch period. These multiple pulses are almost identical to the *'energy pulses'* [2] observed in $\Psi[x(n)]$, which provide evidence that speech resonances exhibit a structure that cannot originate from a linear time-invariant resonator model because applying Ψ to its impulse response $\ell(n) = Ar^n\cos(\Omega_c n + \theta)$ yields a decaying exponential $\Psi[\ell(n)] = A^2 r^{2n}\sin^2(\Omega_c)$ without any multi-pulse structure. The estimated instantaneous frequency in Fig. 2(c) oscillates around its center value with a deviation that can reach 200 Hz. It contains some isolated narrow spikes, which are usually caused either by amplitude valleys or by the onset of a new pitch pulse. We eliminate some of these spikes by post-smoothing the frequency signal via a median filter. Excluding these narrow spikes, in vowels the instantaneous frequency and amplitude envelope profiles follow simple oscillatory, roughly sinusoidal, patterns. We have seen similar numbers (2-4) of amplitude pulses and formant frequency oscillations within a pitch period in many other of our experiments with signals from speech vowels. In a few cases of signals from low formants of vowels we observed only one energy pulse per pitch period. This may be partially explained by a large amount of damping, or by a low amount of modulation for the specific speaker/sound/formant combination. Usually, we observed stronger modulations in higher formants.

Our AM–FM model of a *single* resonance does not explicitly take into consideration the facts that actual speech vowels are quasi-periodic and usually consist of multiple resonances. Both of these phenomena may affect the DESA estimates. The pitch periodicity induces narrow spikes in the amplitude and (mainly) the frequency signal around the onset of each pitch pulse [7]. The effect of neighboring formants that have not been completely rejected by the bandpass filter is to cause 'parasitic' FM and AM modulations, which have a smaller order of magnitude than the main observed modulations [7].

We have also applied the DESA to synthetic speech vowel signals produced by linear resonators with constant formants and excited by periodic impulse trains. As Fig. 3 shows, the estimated amplitude envelopes consisted of exponentially-decaying segments and the estimated instantaneous frequency signals were roughly constant, both interrupted by discontinuity jumps close to onsets of pitch pulses. In contrast to the linear synthetic case, the DESA has uncovered oscillatory pulses in the amplitude and frequency signals of real speech resonances, which indicates the existence of modulations.

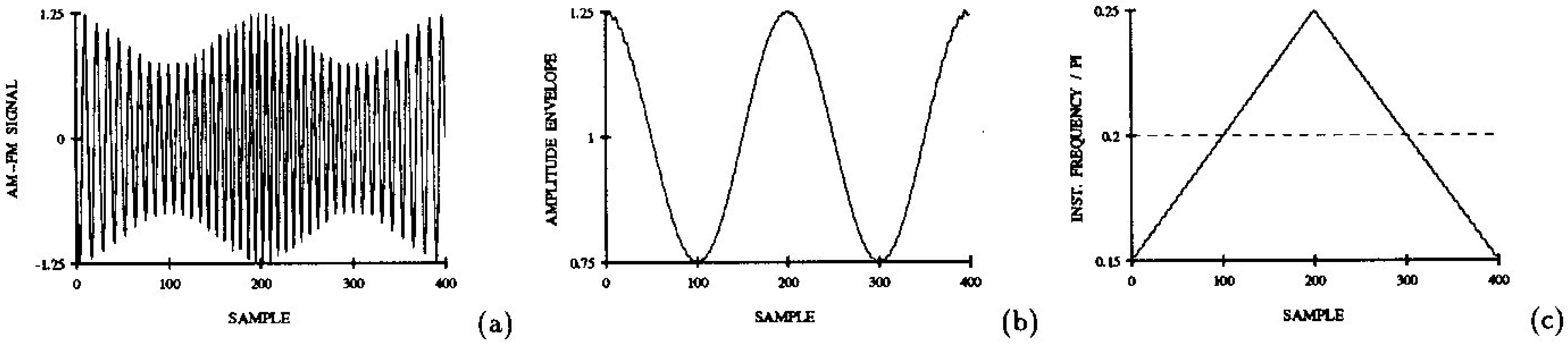

Figure 1: (a) AM–FM signal $a(n)\cos[0.2\pi(n-100)+\pi(n-100)^2/4000]$ for $n = 0, \ldots, 200$ and $a(n)\cos[0.25\pi(n-200)-\pi(n-200)^2/4000+\pi]$ for $n = 201, \ldots, 400$, where $a(n) = [1+0.25\cos(\pi n/100)]$. (b) Estimated amplitude envelope using DESA–1. (c) Estimated instantaneous frequency (as fraction of π); dotted line shows average value of true $\Omega_i(n)/\pi$.

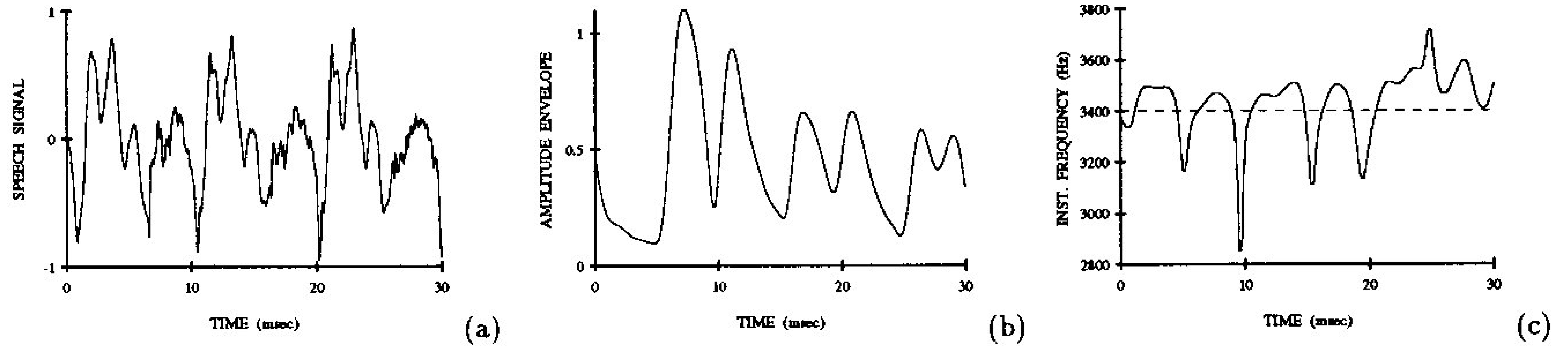

Figure 2: (a) Signal from speech vowel. (b) Estimated amplitude envelope of a resonance around 3400 Hz using DESA–1. (c) Estimated instantaneous frequency, smoothed by an 11-point median filter; dotted line shows center formant value.

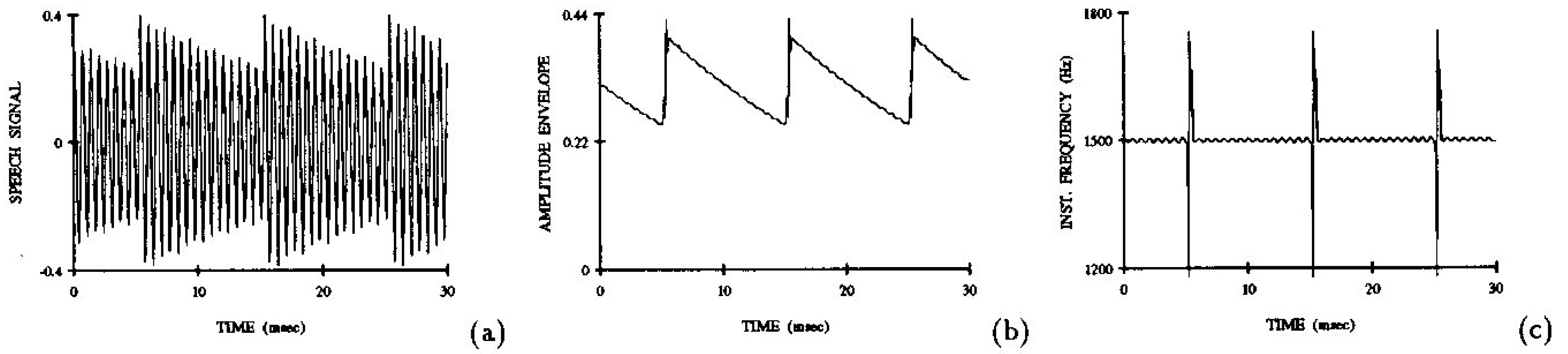

Figure 3: (a) Signal from synthetic speech vowel with one formant at 1500 Hz and pitch frequency of 100 Hz. (b) Estimated amplitude envelope using DESA–1. (c) Estimated instantaneous frequency.

REFERENCES

[1] H. M. Teager, "Some Observations on Oral Air Flow During Phonation", *IEEE Trans. Acoust., Speech, Signal Processing,* ASSP–28, pp. 599-601, Oct. 1980.

[2] H. M. Teager and S. M. Teager, "Evidence for Nonlinear Sound Production Mechanisms in the Vocal Tract", in *Speech Production and Speech Modelling,* NATO Adv. Study Inst. Series D, Vol. 55, Bonas, France, July 1989; Kluwer Acad. 1990.

[3] J. F. Kaiser, "On a simple algorithm to calculate the 'energy' of a signal", in *Proc. IEEE ICASSP-90,* Albuquerque, NM, pp. 381–384, Apr. 1990.

[4] J. F. Kaiser, "On Teager's Energy Algorithm and Its Generalization to Continuous Signals", in *Proc. 4th IEEE Digital Signal Processing Workshop,* Mohonk (New Paltz), NY, Sep. 1990.

[5] P. Maragos, J. F. Kaiser, and T. F. Quatieri, "On Detecting Amplitude and Frequency Modulations Using Energy Operators", Tech. Report 91–6, Harvard Robotics Lab, April 1991.

[6] P. Maragos, T. F. Quatieri, and J. F. Kaiser, "Speech Nonlinearities, Modulations, and Energy Operators", in *Proc. IEEE ICASSP-91,* Toronto, May 1991.

[7] P. Maragos, J. F. Kaiser, and T. F. Quatieri, "Energy Separation in Signal Modulations with Application to Speech Analysis", Tech. Report 91–17, Harvard Robotics Lab, Nov. 1991.