A Wavelet Auditory Model and Data Compression

JOHN J. BENEDETTO*,‡ AND ANTHONY TEOLIS†,‡

**Department of Mathematics, †Institute for Systems Research and Department of Electrical Engineering, University of Maryland, College Park, Maryland 20742, and ‡Prometheus, Inc., Newport, Rhode Island 02840*

Reprinted from APPLIED AND COMPUTATIONAL HARMONIC ANALYSIS, Volume 1, No. 1, December 1993

APPLIED AND COMPUTATIONAL HARMONIC ANALYSIS 1, 3–28 (1993)

A Wavelet Auditory Model and Data Compression

JOHN J. BENEDETTO*,‡,[1] AND ANTHONY TEOLIS†,‡,[1]

*Department of Mathematics, †Institute for Systems Research and Department of Electrical Engineering, University of Maryland, College Park, Maryland 20742, and ‡Prometheus, Inc., Newport, Rhode Island 02840

Received April 5, 1993

A time-scale representation of (acoustic) signals, motivated by the structure of the mammalian auditory system, is presented. Drawing from the theory of irregular sampling and frames, a theoretical framework is developed in which an iterative algorithm for reconstruction is constructed. Numerical examples are included which illustrate the validity of such a representation as a new and effective method to deal with speech compression problems.

1. INTRODUCTION

We construct and implement a wavelet auditory model (WAM). The front end of the construction follows the formulation of the auditory system as found in Yang *et al.* [39]. The major distinguishing feature of the WAM construction is our irregular sampling component for decoding auditory patterns. This component was first described in [6], along with a mathematical analysis of non-linear operations in auditory systems. The implementation of WAM is made from the point of view of developing a useful speech processing tool. The main application in this paper deals with data compression, and there are ongoing experiments and forthcoming work on noise suppression. Further applications and developments related to the computer implementation herein are found in [37].

In Section 2, we introduce the wavelet transform and some other mathematical preliminaries. Section 3 provides the physical basis for our approach, using current ideas concerning auditory models, and introduces WAM data, indicating the way we shall use irregular sampling. The material on auditory models is based on [1, 11, 14, 18–20, 25, 26, 29, 33, 34, 39, 40]. Section 4 is divided into two parts. In the first part, we give a mathematical model for the causal mammalian cochlear filter bank used in WAM. The analysis depends on a constructive proof of the Paley–Wiener logarithmic integral theorem, as well as the construction of some special functions. In the second part, we discretize WAM data, analogously to the approach in [39], and then show that our discretization is compatible with ideas associated with wavelet frames. The theory of frames and our approach to irregular sampling [7, 9] are the subjects of Section 5. Further, with this background, we quantify the wavelet frame properties of WAM data by estimating frame bounds. These are important for establishing reconstruction formulas.

All of the aforementioned material leads us to WAM implementation in Section 6, and our approach and application to compression in Section 7. The WAM implementation in Section 6 is based on properties of frames, which, in the context of our irregular sampling approach, lead to a reconstruction algorithm; in fact, the theoretical basis for this algorithm is in terms of *local frames* [36]. In Section 7, we describe our compression method in terms of WAM data and a distribution function, which leads to a natural thresholding. Using this method, we describe an experiment to ensure prescribed bit rates. We conduct the experiment with TIMIT speech data and synthesized signals. Finally, we present an analysis and evaluation of our results.

The paper closes with three appendixes. Appendix A and Appendix B are mathematical, dealing with nonlinearities in WAM and cochlear filter design, respectively. Appendix C shows results of WAM processing for a number of additional data.

2. PRELIMINARIES

$L^2(\mathbb{R})$ is the space of complex-valued finite energy signals defined on the real line $\mathbb{R}$. The *norm* of an element $f \in L^2(\mathbb{R})$ is

$$\|f\|_2 \equiv \int |f(t)|^2 dt < \infty,$$

where integration is over $\mathbb{R}$, and the *inner product* of $f, g \in L^2(\mathbb{R})$ is $\langle f, g\rangle = \int f(t)\bar{g}(t)dt$. The *Fourier transform* of $f \in L^2(\mathbb{R})$ is $\hat{f}(\gamma) = \int f(t)e^{-2\pi i t\gamma}dt$, for $\gamma \in \hat{\mathbb{R}}\ (\equiv \mathbb{R})$, where convergence of the integral to $\hat{f}$ is in the L^2-sense.

For $s > 0$, the *L^2-dilation* operator D_s is defined by $D_s g(t) \equiv s^{1/2} g(st)$ for $g \in L^2(\mathbb{R})$. As such, $(D_s g)\hat{}(\gamma) = s^{-1/2}\hat{g}(s^{-1}\gamma) = D_{s^{-1}}\hat{g}(\gamma)$. For $u \in \mathbb{R}$, the *translation* op-

[1] Supported by DARPA Contract DAA-H01-91-CR212.

1063-5203/93 $5.00

erator τ_u is defined by $\tau_u g(t) \equiv g(t-u)$ for $g \in L^2(\mathbb{R})$. As such, $(\tau_u g)\hat{}(\gamma) = e^{-2\pi i u\gamma}\hat{g}(\gamma)$. The *convolution* of $f, g \in L^2(\mathbb{R})$ is defined by

$$f*g(t) \equiv \int f(t-u)g(u)du = \int f(u)g(t-u)du.$$

$f*g$ is an absolutely convergent Fourier transform.

For a fixed $g \in L^2(\mathbb{R})$, the *wavelet transform* of $f \in L^2(\mathbb{R})$ is the function

$$W_g f(t, s) \equiv W_g(t, s) \equiv (f*D_s g)(t) \qquad (2.1)$$

defined on the *time-scale plane* $t \in \mathbb{R}$, $s > 0$. There are modifications of this definition, and there are seminal developments of wavelet theory by Daubechies [13], Mallat, and Meyer [27].

By a straightforward calculation, we obtain

$$W_g(t, s) = \langle f, \theta_{t,s}\rangle, \qquad (2.2)$$

where $f, g \in L^2(\mathbb{R})$,

$$\theta_{t,s}(u) \equiv \tau_t D_s \tilde{g}(u), \qquad (2.3)$$

and $\tilde{g}$ is the *involution* of g defined as $\tilde{g}(u) \equiv \bar{g}(-u)$. If the derivative $\partial_t g$ is an element of $L^2(\mathbb{R})$, we define $W_{\partial_t g} f$ analogously to the definition of $W_g f$ in (2.1). In this case, if $W_{\partial_t g} f$ converges uniformly on time intervals, for each fixed scale $s > 0$, and if a mild "smoothness" condition is satisfied, then

$$\partial_t W_g(t, s) = sW_{\partial_t g}(t, s). \qquad (2.4)$$

These hypotheses for the validity of (2.4) can be weakened; and (2.4) is true generally for the causal filters g and signals f under consideration here.

Notationally, we follow standard notation in mathematical analysis, e.g., [35]. In particular, besides $L^2(\mathbb{R})$, we use other $L^p(\mathbb{R})$ spaces and their corresponding norms denoted by $\|\cdots\|_p$. $L^2[-\Omega, \Omega]$ is the space of finite energy signals defined on the interval $[-\Omega, \Omega]$; and PW_Ω is the *Paley–Wiener space,* defined as

$$PW_\Omega \equiv \{f \in L^2(\mathbb{R}) : \operatorname{supp} \hat{f} \subseteq [-\Omega, \Omega]\},$$

where $\operatorname{supp} \hat{f}$ is the support of $\hat{f}$. $l^2(\mathbb{Z})$ is the space of finite energy sequences. Finally, we write

$$e_{-t}(\gamma) \equiv e^{-2\pi i t\gamma}.$$

3. WAVELET AUDITORY MODEL (WAM)

In the human auditory system, an acoustic signal f_* produces a pattern of displacements W of the basilar membrane at different locations for different frequencies [20]. Displacements for high frequencies occur at the basal end; for low frequencies they occur at the wider apical end inside the spiral. The signal f_* causes a traveling wave on the basilar membrane; the basilar membrane responds to frequencies between 200 and 20,000 Hz. For comparison, telephone speech bandwidth deals with the range 300–4000 Hz. The cochlea analyzes sound in terms of these traveling waves, much like a parallel bank of linear time-invariant "cochlear" filters, in this case, a bank with 30,000 channels. This cochlear analysis is complex and subtle, and there are unanswered questions [1] and reservations [25]. Our model does not attempt to quantify the unsettled issues in cochlear micromechanics. On the other hand, present knowledge of cochlear encoding is sufficient for constructing successful models in a variety of applications, e.g., [11, 18]. In our model, the impulse responses of the aforementioned cochlear filters along the length of the cochlea are related by *dilation,* and, consequently, their transfer functions are invariant except for a frequency translation along the approximately logarithmic axis of the cochlea [33, 34]. This suggests that the initial processing occurring in the cochlea may be modeled as the wavelet transform $W_g f_*(t, s) \equiv W_g(t, s)$, where g is a fixed causal impulse response and $\{D_s g : s > 0\}$ is the bank of cochlear impulse responses. Thus, as the first step in the construction of WAM, we follow the development in [39], and identify the displacements W, due to the stimulus f_*, with the output of the cochlear filter bank having the impulse responses $\{D_s g\}$; i.e., we set $W = W_g f_*(t, s) \equiv W_g(t, s)$. Specifically, we fix $a_0 > 1$ and set $s_m = a_0^m$ for $m \in \mathbb{Z}$, the set of integers. As such, in WAM the signal f_* first produces a discrete pattern of displacements,

$$W_g(t, s_m), \quad m \in \mathbb{Z}, \qquad (3.1)$$

for points (t, s_m) in the time-scale plane. For mammalian models, a typical value for a_0 is $1/a_0 = 0.9445$, e.g., [39]. From our point of view, the value of a_0 is an adaptive parameter which should be chosen to optimize results for specific problems.

The shape of $\hat{g}$ is critical for the effectiveness of the auditory process, e.g., Section 4. It is well known that these filters have asymmetrical "shark-fin" shaped amplitudes in the frequency domain [1, 29]. In particular, the rate of decay (roll-off) of the filter with respect to distance from its characteristic frequency (CF) is much higher on the high frequency side than on the low frequency side. The high frequency edges of the cochlear filters act as abrupt "scale delimiters." Thus, a sinusoidal stimulus creates a response which propagates up to the appropriate scale and dies out beyond it.

The auditory system does not receive a wavelet transform W_g directly, but rather a substantially modified version of it. In fact, in the next step of the auditory process, the output of

each cochlear filter is effectively high-passed by the velocity coupling between the cochlear membrane and the cilia of the hair cell transducers that initiate the electrical nervous activity by a shearing action on the tectorial membrane. Thus, the mechanical motion of the basilar membrane is converted to a receptor potential in the inner hair cells. It is reasonable to approximate this stage by a time derivative, obtaining the output $\partial_t W_g(t, s)$.

At the next step in the auditory process, an instantaneous sigmoidal non-linearity R is applied, followed by a low pass filter with impulse response h. These operations model the threshold and saturation that occur in the hair cell channels, and the leakage of electrical current through the membranes of these cells [28, 34]. The cochlear output

$$C_{h,R}(t, s) \equiv (R \circ \partial_t W_g(\cdot, s)) * h(t), \tag{3.2}$$

where "$\circ$" is composition and convolution is with respect to time, is a planar auditory nerve pattern sent to the brain along the scale-ordered array of auditory channels $(t, \cdot)$. Typically, the composition by R can be represented by functions

$$R_T(y) \equiv \frac{e^{Ty}}{1 + e^{Ty}},$$

parameterized by T. Obviously, $\lim_{T\to\infty} R_T = H$, the Heaviside function. Approximations to the Heaviside function are reasonable since the nerve fibers from the inner hair cells to the auditory nervous system fire at positive rates, and since this action cannot process above a certain limit, i.e., the aforementioned saturation. For computational convenience in WAM, we take R to be H and set $h \equiv \delta$, the Dirac δ-measure, even though δ does not give rise to a low pass filter. Thus, $C_{h,R}(t, s)$ in (3.2) is replaced by the *cochlear output,*

$$C(t, s) \equiv H \circ \partial_t W_g(t, s). \tag{3.3}$$

The auditory nerve patterns determined by the cochlear output are now processed by the brain in ways that are not completely understood. One such processing model is the lateral inhibitory network (LIN), e.g., [28], and it will be a component of WAM. This network has been studied with a view to extracting spectral patterns of acoustic stimuli [28, 34]. Scientifically, it reasonably reflects proximate scaling channel behavior, and, mathematically, it is relatively simple. Essentially, LIN detects edges and other discontinuities of $C(t, s)$ along the scaling s-axis of the cochlea. Thus, it can be viewed as a scaling derivative ∂_s; and so the operation of LIN on the cochlear output gives rise to the data $\partial_s C$, which can be written as

$$\partial_s C(t, s) = (\delta \circ \partial_t W_g(t, s)) \partial_s \partial_t W_g(t, s), \tag{3.4}$$

since δ is the distributional derivative of H. Notationally, we let

$$\Gamma_s(f_*) \equiv \{(t, s) : \partial_t W_g(t, s) = 0\}.$$

Formally, the factor $\delta \circ \partial_t W_g(t, s)$ in (3.4) can be written as

$$\forall s > 0, \quad \delta \circ \partial_t W_g(u, s) = \sum_{(t,s)\in\Gamma_s(f_*)} \frac{1}{|\partial_{tt} W_g(t, s)|} \delta_t(u), \tag{3.5}$$

where δ_t is the Dirac δ-measure supported by the point $\{t\}$. The calculation establishing (3.5) is in Appendix A. Realistically, f_* is band-limited so that for each fixed $s > 0$, $\partial_t W_g(t, s)$ is an analytic function. Thus, the sum in (3.5) is countable. Further, by definition of the Dirac δ-measure, (3.4) and (3.5) combine to yield

$$\forall s > 0, \quad \partial_s C(u, s) = \sum_{(t,s)\in\Gamma_s(f_*)} \frac{1}{|\partial_{tt} W_g(t, s)|} \times \partial_s \partial_t W_g(t, s) \delta_t(u). \tag{3.6}$$

The relation between the curvature terms in (3.6) and normalizations is discussed in [6]. For the purpose of the present work, the critical aspect of (3.6) is that data $\partial_s C$ processed by the "brain" depend *only* on those values of $\partial_s \partial_t W_g$ at $(t, s) \in \Gamma_s(f_*)$ for a given scale $s > 0$.

Because of the discrete pattern of displacements defined in (3.1) and the analyticity mentioned above, we define

$$\Gamma(f_*) \equiv \{(t_{m,n}, s_m) : \partial_t W_g(t_{m,n}, s_m) = 0\}; \tag{3.7}$$

and, because of the observation about (3.6), we define *WAM data* as

$$\Lambda(f_*) \equiv \{\partial_s \partial_t W_g(t, s) : (t, s) \in \Gamma(f_*)\}. \tag{3.8}$$

At this point, a given signal f_* has been processed to produce the discrete irregularly spaced planar set $\Gamma(f_*)$ taking values $\Lambda(f_*)$. The final component of WAM is the synthesis of f_* from $\Lambda(f_*)$ in terms of irregular sampling reconstruction formulas, e.g., Section 5. In this context, it is natural to compress WAM data in such reconstruction procedures, and this is the subject matter of Section 7. WAM is depicted in Fig. 1.

Because of the role of $\Gamma(f_*)$ in WAM and recent work by Mallat *et al.,* it should be pointed out that the wavelet extrema are *not* the essential data for WAM reconstruction. Further, even the zero set $\Gamma(f_*)$ arises in a nonstandard way, dictated by a physically compelling non-linear operation in the auditory process; viz. (3.4)–(3.6) and Appendix A.

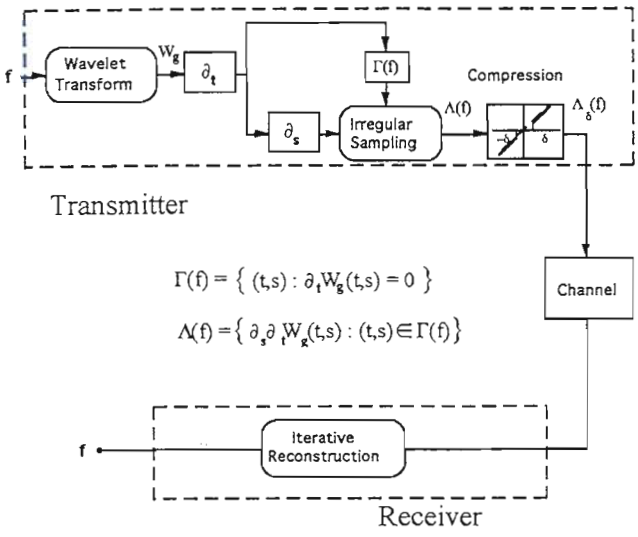

FIG. 1. Schematic of WAM processing.

4. MATHEMATICAL MODELING

4.1. Cochlear Filter Design

As indicated in Section 3, the shape of $|\hat{g}|$ is critical for the effectiveness of the auditory process, and generally $\hat{g}$ has an asymmetrical "shark-fin" shaped amplitude with faster rate of decay on the high frequency side than on the low frequency side.

We begin the construction of $|\hat{g}|$ on $\hat{\mathbb{R}}$ as follows. Take $F(\gamma) \equiv m\gamma 1_{[0,\Omega)}(\gamma)$, where 1_S is the characteristic function of S. Let $\rho \geq 0$ be compactly supported and have the property that $\int \rho(\gamma)d\gamma = 1$. For example, we could take $\rho = n1_{[0,1/n]}$.

It is convenient to deal with smooth functions. As such, we can define ρ as

$$\rho(\gamma) \equiv \frac{\varphi(\epsilon - |\gamma|^2)}{\int \varphi(\epsilon - |\lambda|^2)d\lambda}, \tag{4.1}$$

where $\varphi(\gamma) \equiv e^{-1/\gamma}$ on $[0, \infty)$ and vanishes otherwise. This φ, as well as ρ, is infinitely differentiable, and supp $\rho \subseteq [-\epsilon, \epsilon]$. We now consider the non-negative function

$$A_\rho \equiv F * \rho \tag{4.2}$$

for some such ρ; cf. Example 4.3. A_ρ has compact support and has the desired shape, and A_ρ is smooth if ρ is smooth.

In general, mammalian auditory filters cannot be expected to have zero phase (e.g., the filters in [29] have approximately linear phase), and, clearly, zero phase compactly supported filters cannot be causal filters. On the other hand, all realizable systems, such as our filter bank with "shark-fin" shaped amplitudes, are necessarily causal. In particular, the cochlear filter bank cannot characterize (reconstruct) future utterances in terms of known (present) speech signals. As such, we design causal filters $\hat{g} \in L^2(\hat{\mathbb{R}})$, i.e., supp $g \subseteq [0, \infty)$, for which $\hat{g}$ has the required "shark-fin" shaped amplitude consistent with mammalian auditory models. Our point of view is that such filters provide a realistic mathematical model for the cochlear filters described in Section 3, and are therefore the proper filters for optimizing the reconstruction process inherent in WAM.

The starting point for the design of such causal filters is the Paley–Wiener logarithmic integral theorem [31, Theorem XII]:

THEOREM 4.1. *Let $A \in L^2(\hat{\mathbb{R}}) \setminus \{0\}$ be non-negative on $\hat{\mathbb{R}}$. $A(\gamma) = |\hat{g}(\gamma)|$ a.e. for some causal filter $\hat{g} \in L^2(\hat{\mathbb{R}})$ if and only if*

$$\int \frac{|\log A(\gamma)|}{1 + \gamma^2} d\gamma < \infty. \tag{4.3}$$

Let $A \in L^2(\hat{\mathbb{R}})$ satisfy (4.3), and define

$$\phi(x, \gamma) \equiv \frac{1}{\pi} \int \frac{x \log A(\lambda)}{x^2 + (\gamma - \lambda)^2} d\lambda.$$

Clearly, ϕ is harmonic in the half-plane $x > 0$. If θ is a conjugate harmonic function of ϕ, then it is unique up to an additive constant; and we construct a particular θ in (4.7). The functions ϕ and θ satisfy the Cauchy–Riemann equations, and $K(z) \equiv \phi(x, \gamma) + i\theta(x, \gamma)$, $z = x + i\gamma$, is an analytic function in the half-plane $x > 0$.

We let

$$p(\gamma) \equiv \frac{1}{\pi} \frac{1}{1 + \gamma^2}$$

and consider the L^1-dilations (by $1/x$),

$$p_{1/x}(\gamma) \equiv \rho(x, \gamma) \equiv \frac{1}{\pi} \frac{x}{x^2 + \gamma^2}, \quad x > 0.$$

Thus, $\lim_{x \to 0} p_{1/x} = \delta$ distributionally, in fact, in the $\sigma(M_b, C_0)$ topology, where C_0 is the space of continuous functions vanishing at $\pm\infty$ and M_b is the space of bounded Radon measures on $\hat{\mathbb{R}}$; see, e.g., [5].

By the definition of ϕ we have

$$\phi(x + i\gamma) = p_{1/x} * (\log A)(\gamma), \quad x > 0, \tag{4.4}$$

and because of the approximate identity $p_{1/x}$, a classical calculation yields

$$\lim_{x \to 0^+} \phi(x + i\gamma) = \log A(\gamma) \quad \text{a.e.;} \tag{4.5}$$

see, e.g., [21, 30, 35].

The harmonic function

$$\kappa(x, \gamma) \equiv \frac{-1}{\pi} \frac{\gamma}{x^2 + \gamma^2}, \quad x > 0, \tag{4.6}$$

is a conjugate harmonic function of ρ and so the Cauchy–Riemann equations, $\partial_x \rho = \partial_\gamma \kappa$ and $\partial_\gamma \rho = -\partial_x \kappa$, are valid in the half plane $x > 0$. Using (4.3), the equations

$$\partial_x \phi = (\partial_x \rho) *_\lambda \log A$$

and

$$\partial_\gamma \phi = (\partial_\gamma \rho) *_\lambda \log A, \quad x > 0,$$

follow from (4.4), where "$*_\lambda$" designates convolution in the second variable of ρ. Thus, we define

$$\theta \equiv \kappa *_\lambda \log A, \quad x > 0. \tag{4.7}$$

The function

$$G(z) \equiv e^{K(z)}, \quad z = x + i\gamma,$$

is analytic in the half-plane $x > 0$, and provides the solution asserted in Theorem 4.1 in the following sense. By (4.5), we formally compute

$$G(i\gamma) = A(\gamma) e^{i\theta(0,\gamma)} \quad \text{a.e.}, \tag{4.8}$$

and note, by (4.7), that

$$\theta(0, \gamma) = -\frac{1}{\pi} \int \frac{\log A(\lambda)}{\gamma - \lambda} \, d\lambda \tag{4.9}$$

is formally the Hilbert transform $\mathcal{H}(-\log A)$ of $-\log A$; see e.g., Appendix B. It turns out that condition (4.3) allows us to assert the existence of a causal filter $\hat{g} \in L^2(\hat{\mathbb{R}})$ for which $\hat{g}(\gamma) = G(i\gamma)$ a.e. The actual filter design is a consequence of (4.8) and (4.9), and is formulated in the following result.

THEOREM 4.2. *Let $A \in L^2(\hat{\mathbb{R}}) \setminus \{0\}$ be non-negative on $\hat{\mathbb{R}}$, and assume condition* (4.3). *Then the function*

$$\hat{g} = A e^{-i\mathcal{H} \log A} \tag{4.10}$$

is a causal filter in $L^2(\hat{\mathbb{R}})$; i.e., $g \in L^2(\mathbb{R})$ and supp $g \subseteq [0, \infty)$.

The formal calculations preceding Theorem 4.2 are justified in Appendix B.

EXAMPLE 4.3. The cochlear filters for WAM use Theorem 4.2 and $A_\rho \equiv F * \rho$ defined in (4.2) in the following way. Let $d(\gamma) \equiv e^{-|\gamma|/(\log^2|\gamma|)}$, and pick γ_* so that $d'(\gamma) < 0$ for all $\gamma \geq \gamma_*$. Then we define A as

$$A(\gamma) = \begin{cases} \dfrac{A_\rho(0)}{d(\gamma_*)} d(\gamma - \gamma_*), & \gamma \leq 0, \\ A_\rho(\gamma), & \gamma \in (0, \Omega), \\ \dfrac{A_\rho(\Omega)}{d(\gamma_*)} d(\gamma - \Omega + \gamma_*), & \gamma \geq \Omega. \end{cases}$$

Clearly, $A \in L^2(\hat{\mathbb{R}})$ and (4.3) is valid. Thus, the causal cochlear filter $\hat{g}$ can be defined by (4.10) in Theorem 4.1.

See also Fig. 2.

4.2. Discretization

We begin by choosing g as in Example 4.3, and recall from Section 2 that $\theta_{t,s} \equiv \tau_t D_s \tilde{g}$ and $W_g(t, s) = \langle f_*, \theta_{t,s} \rangle$. We have

$$\partial_s \partial_t W_g(t_0, s_0) = \lim_{h \to 0} \frac{1}{h} (\partial_t \langle f_*, \theta_{t_0, s_0} \rangle - \partial_t \langle f_*, \theta_{t_0, s_0 - h} \rangle). \tag{4.11}$$

Letting $(t_{m,n}, s_m) \in \Gamma(f_*)$ and $h = s_m - s_{m-1}$ in (4.11) we have the approximation

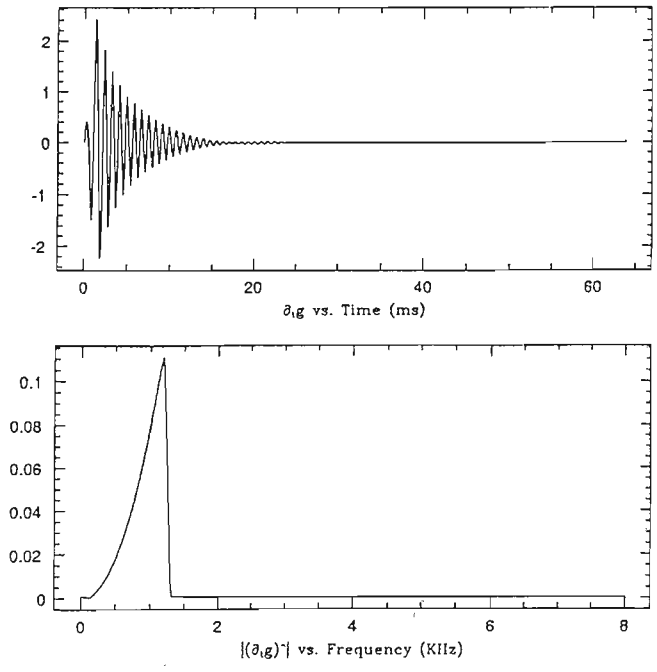

FIG. 2. Causal filter response: ∂g and $|(\partial g)\hat{}|$.

$$\partial_s \partial_t W_g(t_{m,n}, s_m)$$
$$\approx \frac{\partial_t W_g(t_{m,n}, s_m) - \partial_t W_g(t_{m,n}, s_{m-1})}{s_m - s_{m-1}}$$
$$= \frac{-\partial_t W_g(t_{m,n}, s_{m-1})}{s_m - s_{m-1}} = \frac{-s_{m-1} W_{\partial g}(t_{m,n}, s_{m-1})}{s_m - s_{m-1}}, \quad (4.12)$$

where we have used the fact that $(t_{m,n}, s_m) \in \Gamma(f_*)$ and Eq. (2.4). Using (4.12) and writing the approximation there as an equality, we have

$$\partial_s \partial_t W_g(t_{m,n}, s_m) = \frac{1}{a_0 - 1} \langle f_*, \tau_{t_{m,n}} D_{s_{m-1}}(\partial \tilde{g}) \rangle \in \Lambda(f_*). \quad (4.13)$$

Because of (4.13) and the frame-theoretic point of view of Section 5, we define

$$\psi_{m,n} = \frac{1}{a_0 - 1} \tau_{t_{m,n}} D_{s_{m-1}}(\partial \tilde{g}) \quad (4.14)$$

and the mapping

$$L: H \to l^2(\mathbb{Z}^2)$$
$$f \mapsto \{\langle f, \psi_{m,n} \rangle\}, \quad (4.15)$$

where H is a Hilbert subspace of $L^2(\mathbb{R})$ containing the class of acoustic signals to be analyzed.

Each function $\psi_{m,n}$ corresponds to an element $(t_{m,n}, s_m) \in \Gamma(f_*)$. In particular, $\{\psi_{m,n}\}$ depends on a given acoustic signal f_*. This is not amenable to a *global* theory of frames, but such a theory is not essential for our purposes. Also, we can rewrite each $\psi_{m,n}$ as

$$\psi_{m,n} = \frac{1}{a_0 - 1} D_{s_{m-1}} \tau_{s_{m-1} t_{m,n}}(\partial \tilde{g}). \quad (4.16)$$

5. FRAMES AND IRREGULAR SAMPLING

Let $\mathcal{H}$ be a Hilbert space contained in $L^2(\mathbb{R})$, and with norm $\|\cdots\| \equiv \|\cdots\|_2$ induced from $L^2(\mathbb{R})$.

DEFINITION 5.1. (a) A sequence $\{\theta_n\} \subseteq \mathcal{H}$ is a *frame* for $\mathcal{H}$ if there exist *frame bounds* $A, B > 0$ such that

$$\forall f \in \mathcal{H}, \quad A\|f\|^2 \leq \sum |\langle f, \theta_n \rangle|^2 \leq B\|f\|^2, \quad (5.1)$$

where summation is over $\mathbb{Z}$. The theory of frames is due to Duffin and Schaeffer [16]; cf. [13, 15, 22, 38].

(b) The *frame operator* of the frame $\{\theta_n\}$ is the function $S: \mathcal{H} \mapsto \mathcal{H}$ defined as $Sf = \Sigma \langle f, \theta_n \rangle \theta_n$.

The following result exhibits some fundamental properties of frames; e.g., [8, 12, 16].

THEOREM 5.2. (a) *If $\{\theta_n\} \subseteq \mathcal{H}$ is a frame with frame bounds A, B, then S is a topological isomorphism with inverse S^{-1}, $\{S^{-1}\theta_n\}$ is a frame with frame bounds B^{-1} and A^{-1}, and*

$$\forall f \in \mathcal{H}, \quad f = \sum \langle f, S^{-1}\theta_n \rangle \theta_n = \sum \langle f, \theta_n \rangle S^{-1}\theta_n \quad (5.2)$$

in $\mathcal{H}$.

(b) *If $\{\theta_n\} \subseteq \mathcal{H}$, let $L: \mathcal{H} \mapsto l^2(\mathbb{Z})$ be defined as $Lf = \{\langle f, \theta_n \rangle\}$, cf. (4.15). If $\{\theta_n\}$ is a frame then $S = L^*L$, where L^* is the adjoint of L.*

(c) *$\{\theta_n\} \subseteq \mathcal{H}$ is a frame for $\mathcal{H}$ with frame bounds A and B if and only if the mapping L is a well-defined topological isomorphism onto a closed subspace of $l^2(\mathbb{Z})$. In this case,*

$$\|L\| \leq B^{1/2} \quad \text{and} \quad \|L^{-1}\| \leq A^{-1/2},$$

where L^{-1} is defined on the range $L(\mathcal{H})$.

The theory of frames allows us to prove *irregular sampling theorems* of the following form; e.g., [7, 9].

THEOREM 5.3. *Suppose $\Omega > 0$ and $\Omega_1 > \Omega$, and let $\{t_n\} \subseteq \mathbb{R}$ have the property that $\{e_{-t_n}\}$ is a frame for $\mathcal{H} \equiv L^2[-\Omega_1, \Omega_1]$ with frame operator S. Further, let $\theta \in L^2(\mathbb{R})$ have the properties that $\hat{\theta} \in L^\infty(\mathbb{R})$, supp $\hat{\theta} \subseteq [-\Omega_1, \Omega_1]$, and $\hat{\theta} = 1$ on $[-\Omega, \Omega]$. Then*

$$\forall f \in PW_\Omega, \quad f = \sum c_n(f) \tau_{t_n} \theta \quad \text{in } L^2(\mathbb{R}), \quad (5.3)$$

where

$$c_n(f) = \langle S^{-1}(\hat{f}\, 1_{[-\Omega_1, \Omega_1]}), e_{-t_n} \rangle.$$

The characterization of sequences $\{t_n\}$ which generate frames for $L^2[-\Omega, \Omega]$ of the form $\{e_{-t_n}\}$ is due to Jaffard [24], and their role in sampling theory is explained in [7, 8].

EXAMPLE 5.4. Let $\mathcal{H} = L^2(\mathbb{R})$ and let $\theta \in L^2(\mathbb{R})$. The *wavelet system* $\{\theta_{m,n}: (m, n) \in \mathbb{Z}^2\}$ is defined by

$$\forall m, n \in \mathbb{Z}, \quad \theta_{m,n}(t) \equiv 2^{m/2} \theta(2^m t - n).$$

We can consider wavelet systems $\{\theta_{m,n}\}$ which are *wavelet frames,* in which case (5.1) is satisfied for $\{\theta_{m,n}\}$, or *wavelet orthonormal bases,* which are special cases of wavelet frames.

Orthogonal wavelet systems have the fundamental *vanishing moment property;* see, e.g., [4, 27]. In fact, it is elementary to prove that *if the elements of a wavelet system $\{\theta_{m,n}\}$ are mutually orthogonal, and θ, $\hat{\theta} \in L^1 \cap L^2$, then $\hat{\theta}(0) = 0$;* cf. [8, Sect. 5.12].

EXAMPLE 5.5. (a) In light of the relationship between (4.14) and the dilation translation structure of wavelet systems, it is relevant that the "wavelet" $\psi \equiv \partial\tilde{g}$ of (4.14) can satisfy the vanishing moment property of Example 5.4. In fact, we can prove the following result for the causal filter $\hat{g}$ defined in Section 4: *if* $\tilde{g}, \partial\tilde{g} \in L^1(\mathbb{R})$ *and if* $\partial\tilde{g}(t)$ *exists for each* $t \in \mathbb{R}$ *then* $(\partial\tilde{g})\hat{}(0) = 0$; see, e.g., [5, p. 151].

(b) The following calculation illustrates to what extent $\{\psi_{m,n}\}$, defined in (4.14), can be considered a wavelet frame for some sufficiently robust Hilbert space $\mathcal{H} \subseteq L^2(\mathbb{R})$; cf. the beginning of Section 6 and the critical observation in Section 4 that $\psi_{m,n}$ depends on f_*.

We first compute

$$\sum |\langle f, \psi_{m,n}\rangle|^2 = \frac{1}{(a_0 - 1)^2} \sum |\langle \hat{f}\overline{D_{s_{m-1}^{-1}}(\partial\tilde{g})\hat{}}, e_{-t_{m,n}}\rangle|^2.$$

Then, in the spirit of the frame hypothesis of Theorem 5.3, we assume that for each $m \in \mathbb{Z}$, $\{-e_{-t_{m,n}} : n \in \mathbb{Z}\}$ is a frame for $\mathcal{H} = L^2[-\Omega, \Omega]$ with frame bounds A_m, B_m. Thus,

$$\sum_m A_m \|\hat{f} D_{s_{m-1}^{-1}}(\partial\tilde{g})\hat{}\|^2 \leq (a_0 - 1)^2 \sum_m \sum_n |\langle f, \psi_{m,n}\rangle|^2 \leq \sum_m B_m \|\hat{f} D_{s_{m-1}^{-1}}(\partial\tilde{g})\hat{}\|^2.$$

Consequently, if we suppose that

$$0 < A \leq \frac{1}{(a_0 - 1)^2} A_m \leq \frac{1}{(a_0 - 1)^2} B_m \leq B < \infty,$$

for some A, B, then by a simple calculation and Plancherel's theorem, we have

$$A(\inf_\gamma \sum_m |D_{s_{m-1}^{-1}}(\partial\tilde{g})\hat{}|^2)\|f\|^2 \leq \sum |\langle f, \psi_{m,n}\rangle|^2 \leq B\|\sum_m |D_{s_{m-1}^{-1}}(\partial\tilde{g})\hat{}(\gamma)|^2\|_\infty \|f\|^2. \quad (5.4)$$

The inequalities in (5.4) lead to frame properties of $\{\psi_{m,n}\}$ if

$$G(\gamma) \equiv \sum_m |D_{s_{m-1}^{-1}}(\partial\tilde{g})\hat{}(\gamma)|^2 \quad (5.5)$$

is bounded above and bounded below away from 0. In any case, the function in (5.5) must be quantified to obtain effective frame decompositions by means of Theorem 5.2; and it should be noted that a_0 plays a role in (5.5) which manifests itself in our numerical work. See Fig. 3.

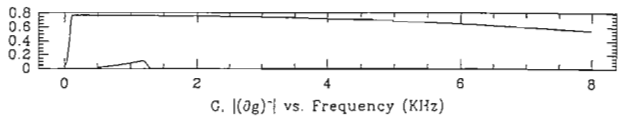

FIG. 3. The function G generated by $(\partial\tilde{g})\hat{}$.

6. WAM IMPLEMENTATION

The basic idea of WAM can now be formulated by combining the calculation of Section 4.2 with Theorems 5.2 and 5.3. First, WAM data have the form

$$\partial_s \partial_t W_g(t_{m,n}, s_m) = \langle f, \psi_{m,n}\rangle \quad (6.1)$$

because of (4.13) and (4.14). Second, if $\{\psi_{m,n}\}$ were a frame for $\mathcal{H}$, with frame operator S and L defined by (4.15), then

$$\forall f \in \mathcal{H}, \quad f = S^{-1}L^*(Lf), \quad (6.2)$$

by Theorem 5.2. In particular, f can be reconstructed by knowledge of the discrete WAM data Lf. Third, in light of the irregular sampling theorem, Theorem 5.3, a frame hypothesis on $\{\psi_{m,n}\} \subseteq \mathcal{H}$ and Eqs. (5.2) yield the "irregular sampling frame decomposition formula,"

$$\forall f \in \mathcal{H}, \quad f = \sum c_{m,n}\tau_{t_{m,n}}(D_{s_{m-1}}\partial\tilde{g}). \quad (6.3)$$

Notwithstanding the difficulties of justifying frame properties of $\{\psi_{m,n}\}$ for a sufficiently large space $\mathcal{H}$ (as indicated at the end of Section 4.2 and in Example 5.5(b)), Eqs. (5.3) and (3.8)–(6.3) motivate the WAM implementation we now present. In fact, recalling that $\{\psi_{m,n}\}$ depends on f_*, we really only need to obtain (6.2) and/or (6.3) iteratively for $f = f_*$; and it is not unreasonable to suppose that $\mathcal{H}$, however small, at least contains the acoustic signal f_* which was used to generate $\{\psi_{m,n}\}$.

Assuming the frame setup of the previous paragraph, and letting A, B be frame bounds for $\{\psi_{m,n}\}$, we have

$$\left\| I - \frac{2}{A + B} S \right\| \leq \frac{B - A}{A + B} < 1,$$

so that by the Neumann expansion,

$$S^{-1} = \frac{2}{A + B} \sum_{j=0}^{\infty} \left(I - \frac{2}{A + B} S \right)^j, \quad (6.4)$$

where I is the identity operator; see, e.g., [7, Algorithm 50; 8, Sect. 6.6]. Applying (6.4) to Sf_* yields

$$f_* = \sum_{j=0}^{\infty} (I - \lambda S)^j (\lambda S) f_*, \quad (6.5)$$

where $\lambda = 2/(A + B)$.

PROPOSITION 6.1. *The signal* f_* *may be recovered from WAM data as*

$$f_* = \lambda \sum_{j=0}^{\infty} L^*(I - \lambda LL^*)^j Lf_*,$$

where Lf_* *is WAM data defined by* (4.13)–(4.15), *and where* $L^*c = \sum c_{m,n}\psi_{m,n}$ *for* $c = \{c_{m,n}\}$.

Proof. Since $\langle Lf, c\rangle \equiv \langle f, L^*c\rangle$ and

$$\langle Lf, c\rangle = \sum \bar{c}_{m,n}\langle f, \psi_{m,n}\rangle = \langle f, \sum c_{m,n}\psi_{m,n}\rangle,$$

we obtain the formula for L^*c.

Because of (6.5) and the fact that $S = L^*L$, it is sufficient to prove that

$$\lambda \sum_{j=0}^{\infty} L^*(I - \lambda LL^*)^j Lf_* = \sum_{j=0}^{\infty} (I - \lambda L^*L)^j(\lambda L^*L)f_*. \tag{6.6}$$

The $j = 0$ terms are clearly the same in (6.6). Assume that

$$\lambda L^*(I - \lambda LL^*)^j Lf_* = (I - \lambda L^*L)^j(\lambda L^*L)f_*. \tag{6.7}$$

Then, using (6.7), we compute

$$\begin{aligned}
&\lambda L^*(I - \lambda LL^*)^{j+1}Lf_* \\
&= \lambda L^*(I - \lambda LL^*)^j Lf_* - \lambda L^*(I - \lambda LL^*)^j \lambda LL^*Lf_* \\
&= \lambda(I - \lambda L^*L)^j L^*Lf_* - \lambda(I - \lambda L^*L)^j L^*L(\lambda L^*Lf_*) \\
&= \lambda(I - \lambda L^*L)^j(I - \lambda L^*L)L^*Lf_* \\
&= \lambda(I - \lambda L^*L)^{j+1}L^*Lf_*,
\end{aligned}$$

and the result follows by induction. □

ALGORITHM 6.2. Suppose we are given WAM data $c_0 \equiv Lf_*$, and set $f_0 = 0$. We define

$$\begin{aligned}
h_n &\equiv L^*c_n, \\
c_{n+1} &\equiv c_n - \lambda Lh_n, \\
f_{n+1} &\equiv f_n + h_n.
\end{aligned}$$

An elementary induction argument shows that

$$\forall n, f_{n+1} = L^*\Big(\sum_{j=0}^{n} (I - \lambda LL^*)^j\Big)c_0.$$

Consequently, by Proposition 6.1, we have

$$\lim_{n\to\infty} \lambda f_n = f_*.$$

As such, since $c_0 = Lf_*$ is WAM data, we use Algorithm 6.2 to reconstruct f_* from Lf_*.

7. COMPRESSION

We apply our WAM processing to the area of speech compression. For speech compression problems, the goal is to represent speech signals in a way which minimizes storage and transmission bandwidth requirements under the constraint that sufficiently high "quality" approximations of the original speech signal can be recovered from the representation. The meaning of the "quality" of a reconstruction is a criterion which is difficult to specify precisely. In vague terms we would like our representation to preserve pertinent perceptual information in the speech signal, e.g., timbre, emotional state of the speaker, inflections, etc. Intelligibility is a less stringent criterion by which we shall judge our reconstructions. In this case, we require only that listeners be able to determine the textual content of the original speech signal purely from audition of the reconstruction.

In the following, we detail a speech compression scheme based on the WAM representation in the context of a voice communication system. As such, we assume inherent constraints on signal bandwidth and allowable bit rates due to realistic device limitations.

7.1. Approach

In this section we describe the general method and setup by which we use our WAM processing to achieve compression for speech signals.

Let f_* be an acoustic signal on the interval I of duration $|I|$ and let L be the WAM discretization operator such that

$$Lf_* = \Lambda(f_*) = \{\langle f_*, \psi_{m,n}\rangle\}$$

is the set of *WAM coefficients;* cf. (4.15) and (3.8). Recall that the elements of the WAM system $\{\psi_{m,n}\}$ are of the form

$$\psi_{m,n} = \frac{1}{a_0 - 1}\tau_{t_{m,n}}D_{s_{m-1}}(\partial\tilde{g}),$$

e.g., (4.14).

It is the WAM coefficients which must be transmitted. For transmission in digital form, it is necessary that the WAM coefficients be quantized. The quantization strategy which we employ is one which maps positive values uniformly along some interval. This uniform mapping corresponds to representing each coefficient with a fixed number of bits. The fixed number of bits which we allocate for the representation of each coefficient is denoted by b_c (bits/coef). The quantization function Q_{b_c} is defined as

$$Q_{b_c}(x) = \begin{cases} 0, & x \leq 0, \\ \left[\frac{x}{M} 2^{b_c}\right], & x \in (0, M), \\ 2^{b_c} - 1, & x \geq M, \end{cases} \tag{7.1}$$

where

$$M \equiv \|Lf_*\|_\infty \equiv \sup_{m,n} \{|\langle f_*, \psi_{m,n}\rangle|\}, \tag{7.2}$$

and where $[x]$ is the largest integer less than or equal to x. Note that we have chosen to neglect the negative coefficients.

We specify the inherent constraint on the amount of information which we can transmit per unit time as a maximum allowable bit rate of b_r bps (bits per second). For convenience, we do not fix this quantity explicitly. Instead, we specify a corresponding coefficient rate c_r, and vary the bit allocation b_c to meet the information rate constraint b_r, through the simple relation $b_r = c_r b_c$. With the coefficient rate fixed and specified, the maximum number of coefficients n_c that we are able to transmit for the function f_* of duration $|I|$ is

$$n_c = c_r |I|.$$

Thus, given the acoustic signal f_* of duration $|I|$ and a fixed coefficient rate, the maximum number of coefficients with which f_* may be represented, while still satisfying the information rate constraint, is given by n_c. With respect to WAM data, this maximum number of coefficients n_c can further be related to a value for a threshold δ. To see this, we introduce the *distribution function,*

$$\lambda(\delta) \equiv \text{card } \Lambda_\delta(f_*) \equiv \text{card}\{\langle f_*, \psi_{m,n}\rangle \geq \delta\}, \tag{7.3}$$

for $\delta \in [0, M]$. It should be noted that we could define the distribution function on $[-M, M]$ (or even some larger interval); however, this is unnecessary, since we have chosen to neglect the negative coefficients. The distribution function λ: $[0, M] \mapsto \mathbb{N}$ is monotonically decreasing and continuous from the left. We may associate with λ an "inverse" λ^{-1} defined as

$$\lambda^{-1}(n) \equiv \inf\{x \in [0, M]: \lambda(x) < n\},$$

where $n \in \mathbb{N}$. If a threshold value δ is chosen as

$$\delta = \lambda^{-1}(n_c),$$

then the WAM thresholded data $\Lambda_\delta(f_*)$ have a cardinality card $\Lambda\lambda^{-1}(n_c) \leq n_c$. Consequently, the total bit requirement for representing the acoustic signal f_* of duration $|I|$ in b_c bits/coefficient is no greater than $n_c b_c$ bits. This, in turn, guarantees that the WAM encoding of the signal f_* is compatible with the bit rate constraint; i.e.,

$$b_c \text{ card } \Lambda_{\lambda^{-1}(n_c)} \leq b_r |I|.$$

7.2. Performance Evaluation

7.2.1. Experiment

We have implemented our WAM processing in software and applied it to a variety of test signals including both synthesized and real speech data. The real speech data are signals taken from the extensive TIMIT data base and specifically includes words taken from the sentence "She had your dark suit in greasy wash water all year," as spoken by one male and one female speaker. Synthesized signals include single and multicomponent sine waves, chirps, and more general frequency/amplitude modulated signals. In this section we detail the processing and WAM output for an example signal taken from this set. The example signal we present here is a female saying the word "water." Complete results of WAM processing for the remaining signals of the test set can be viewed in Appendix C.

For evaluation purposes we fix a coefficient rate at $c_r = 4800$ coefficients per second (coef/sec), and vary the number of bits b_c with which each coefficient is represented. Specifically, we take values of $b_c = 1, 2, 4$. Since these quantities are related to the overall bit rate b_r by the relation $b_r = c_r b_c$, the corresponding bit rates are $b_r = 4.8, 9.6, 19.2$ Kbps. For each value of b_c, reconstructions are computed via Algorithm 6.2 based on the thresholded WAM data $\Lambda_{\lambda^{-1}(n_c)}$. Recall that the set $\Lambda_{\lambda^{-1}(n_c)}$ is completely determined from the signal f_*, the specified coefficient rate c_r, and the coefficient bit allocation b_c, as described in Section 7.1.

Our experiment can be summarized as follows. Fix $c_r = 4800$ coef/sec. For each f_* in our test set and for each value of b_c, we perform the following steps:

(i) Compute the WAM representation of f_*:

$$\Lambda(f_*) = \{\langle f_*, \psi_{m,n}\rangle\}.$$

(ii) Determine the maximum number of coefficients n_c with which f_* can be represented and still meet the coefficient rate constraint,

$$n_c = c_r \cdot |I(f_*)|,$$

where $|I(f_*)|$ is the duration of f_*.

(iii) Compute the distribution function λ.

(iv) Threshold the WAM representation $\Lambda(f_*)$ by $\delta = \lambda^{-1}(n_c)$, yielding the truncated representation

$$\Lambda_\delta(f_*) \equiv \{\langle f_*, \psi_{m,n}\rangle \geq \delta\}.$$

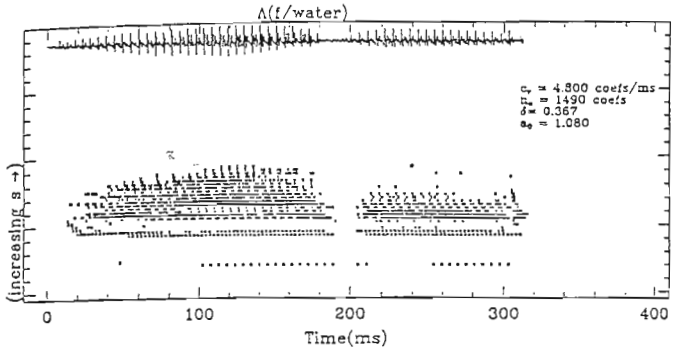

FIG. 4. Female spoken "water" and its thresholded WAM representation.

(v) Quantize the thresholded WAM representation, yielding the sequence

$$Q_{b_c}(\Lambda_\delta(f_*)).$$

(vi) Generate a reconstruction of f_* using Algorithm 6.2, where the initial data are

$$c_0 = Q_{b_c}(\Lambda_\delta(f_*)).$$

In the following we explain in detail this procedure for a particular example signal, and, where applicable, we compute values for parameters and show pictures of the resulting processing. Recall that the particular signal on which we have chosen to illustrate this process is the acoustic signal "water" as spoken by a female.

Figure 4 depicts the acoustic signal "water" as spoken by a female and its associated (thresholded) WAM representation. In this figure, the time signal is superimposed at the top of time-scale plane for easy reference. Each "x" in the figure, corresponding to a point in the time-scale plane (t, s), represents a particular coefficient in the WAM representation of f_*. From this figure it can be seen that the duration of the word "water" is roughly 310 ms. For a fixed coefficient rate of $c_r = 4.8$ coef/ms, this translates into a maximum allowable number of coefficients $n_c = c_r|I| \approx 4.8 \cdot (310) = 1488$. The λ coefficient distribution for the female spoken "water" is shown in Fig. 5. From this graph it is easy to read off an appropriate threshold value of $\delta = \lambda^{-1}(n_c) \approx 0.4$ (actually 0.367). Now we can say precisely that what is depicted in Fig. 4 corresponds to elements of the WAM thresholded data $\Lambda_\delta(f_*)$, where f_* is the female spoken word "water" and $\delta \approx 0.4$. Thus, in this figure there are roughly 1488 (actually 1490) x's.

Since it comes from the set $\Lambda_\delta(f_*)$, each "x" in the time-scale plane of Fig. 4 has an associated coefficient which has a particular positive value greater than the threshold δ. For digital communication these values are then quantized according to the function Q_{b_c} given in (7.1). To perform the

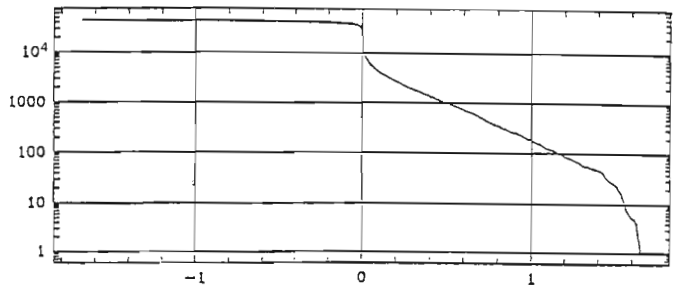

FIG. 5. λ distribution of female spoken "water."

quantization, the value M in (7.2) and the value of b_c must be known. M is signal dependent and may be read from the λ distribution of the signal as $M = \lambda^{-1}(0)$. For the particular spoken word "water" the λ distribution of Fig. 5 reflects a value for M of about 1.65. As an experimental parameter, b_c is varied through the values 1, 2, and 4 bits per coefficient.

Once the thresholded coefficients have been quantized according to Q_{b_c}, they are passed through Algorithm 6.2 for reconstruction. Figure 6 displays the time-signal reconstructions obtained from a single iteration, i.e., f_1, of the algorithm for the three values of $b_c = 1$, 2, and 4. Similarly, Fig. 7 displays the magnitudes of the Fourier transforms of the same reconstructions.

7.2.2. *Compression Ratios*

When dealing with data and schemes for data compression it is natural to introduce a measure of compression. In rough terms, a "compression ratio" measures the relative decrease in complexity of data in a raw form as compared to the complexity of its new compressed form.

For speech it is customary to deal directly with bit rates instead of compression ratios. This is because the bandwidth of speech is a fixed constant which, for some practical purposes, may be taken to be $\Omega = 4000$ Hz. Consider the "raw" form of an analog speech signal f_* to be a sampled version

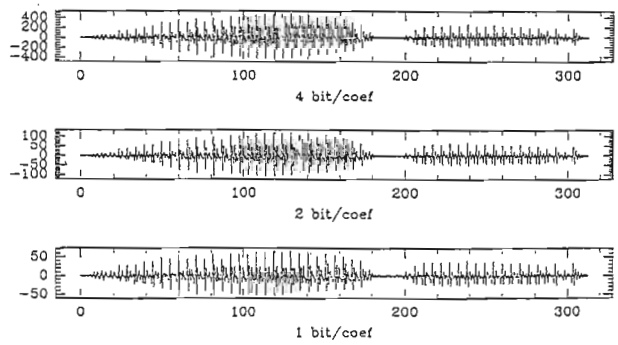

FIG. 6. Time domain reconstructions of female spoken "water" as the bit allocation per coefficient b_c varies (1, 2, and 4 bits per coefficient).

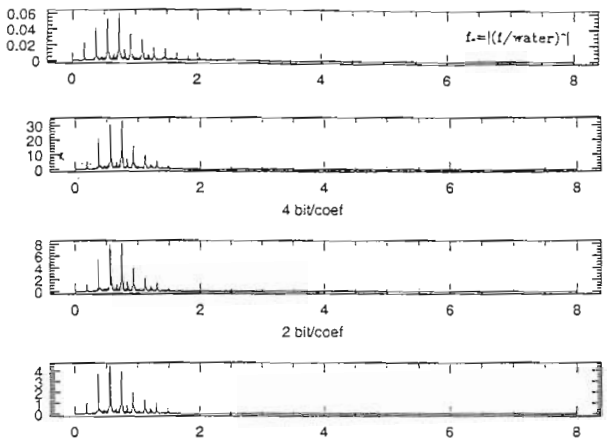

FIG. 7. Frequency domain reconstructions of female spoken "water" as the bit allocation per coefficient b_c varies (1, 2, and 4 bits per coefficient).

with 8 bits per sample and a uniform sampling period of $T = 1/(2\Omega) = 1/8000$ seconds. In this case the required "raw" bit rate is 64 Kbps. Since the hypotheses of the classical sampling theorem are satisfied, it is possible to reconstruct (modulo slight errors due to quantization) the original speech signal f_*. Any representation which allows for recovery of the original speech signal f_* and requires a bit rate less than 64 Kbps is a compressed version of f_*. Consequently, a compression ratio of 10:1 is achieved by a particular speech compression scheme if that scheme yields a bit rate of 6.4 Kbps.

Many speech compression schemes have been introduced which have a required bit rate of less than 64 Kbps. The development of frequency channel vocoders by Dudley (in 1928) was a major early effort in compression. Linear predictive coding (LPC) was introduced by Atal and Schroeder [2, 3]. LPC10 and its variants deal effectively with low bit rates ($\leq$2.4 Kbps) for certain quality criterion. Codebook excited linear predictive coding (CELP), discrete cosine transform (DCT) coders, and various subband coding schemes produce intelligible and good quality reconstructions with bit rates in the range of 4 to 16 Kbps.

For comparison, recall that our experiments deal with WAM representations which require bit rates in the range of 4.8 to 19.2 Kbps. It should be noted that this range is by no means the limit of WAM compression, and we expect WAM compression to be able to achieve good quality reconstructions at much lower bit rates.

7.2.3. Analysis

Examining the results of our experiment we can make the following observations:

(i) In general, reconstructions are "good" for all values of $b_c = 1, 2,$ or 4. We use the term "good" in the sense that both the time and frequency magnitude reconstructions are judged to be close to their original counterparts. Moreover, all reconstructions are intelligible.

(ii) Frequency magnitude reconstructions degrade less severely than time reconstructions as b_c varies from 4 to 1 bit per coefficient.

(iii) Strong frequency components (peaks) are replicated faithfully.

All of the observations suggest that the proposed WAM speech compression scheme is a promising one. In particular, observations (ii) and (iii) are in accord with the notion that the ear is coding only perceptually relevant features [17, 32]. Specifically, we can interpret these observations to indicate that the WAM representation and compression scheme preserve frequency magnitude information with greater accuracy than phase information.

The results of the experiment also indicate that the WAM representation is highly robust to quantization effects. Allocating just a single bit per coefficient ($b_c = 1$) still allows for good quality reconstructions. It is here that we see one benefit of non-orthogonal highly redundant systems. Since the WAM system $\{\psi_{m,n}\}$ generated by a particular acoustic signal f_* is highly linearly dependent, the operator L^* has a large kernel. Modeling quantization as coefficient noise, we have $\tilde{c} \equiv Q_{b_c}(c) \approx c + w$, where w is some random noise with appropriate statistics. Now $L^*\tilde{c} = L^*c + L^*w$, and since the kernel of L^* is large, much of the noise (since it is random) occurs in the kernel; thus $L^*\tilde{c} \approx L^*c$. This type of argument for coefficient noise robustness is developed in [15].

7.3. Discussion

There are other variables, trade-offs, and issues for evaluating the WAM compression scheme. We list some of them here.

- Besides coefficient quantization, time quantization must also be addressed. Each WAM coefficient is associated with an element in the WAM system $\{\psi_{m,n}\}$; see, e.g., (4.14). Since the sequence $\{t_{m,n}\}$ is signal dependent, the reconstruction process (receiver) must have knowledge of these values. In our reconstructions we have assumed complete knowledge of the sequence $\{t_{m,n}\}$.
- Some synthesized examples do not seem to reconstruct in time as well as some TIMIT speech data. In particular, frequency modulated signals seem to present some difficulty for time reconstructions, e.g., the fm echo signal in Appendix C. On the other hand, frequency magnitude reconstruction is still good for such signals.
- There is a slight trade-off between reconstruction accuracy and algorithm speed, i.e., the number of iterations used in the reconstruction Algorithm 6.2. All the reconstructions presented in this paper are the result of a *single* iteration of Algorithm 6.2. From our experiment, it is evident, then,

that one iteration already provides "good" reconstruction. This observation can be related to the fact that the function G of (5.5) is almost constant along the frequency band of interest.

- Essentially, we have circumvented the issue of windowing of the signal by taking the window to be of a length equal to the duration of the signal. In practice, windowing plays a finer role; e.g., in narrowband speech compression systems going back to Dudley, (in 1939) speech coders divide the speech signal into intervals of duration 10 to 25 ms.
- The dilation parameter a_0 effectively changes the frequency support of each filter in the filter bank $\{(D_{a_0^k}\partial\tilde{g})\hat{}\}$. In particular, increasing the value of a_0 decreases the bandwidth of each filter in the bank. A decrease in bandwidth necessarily implies that a particular filter will respond to a smaller band in frequency. Thus, keeping f the same, an increase in the value of a_0 will cause the bands of activity in the original a_0 representation to become compressed along the s-axis in the new increased a_0 representation. Thus, with δ fixed, larger a_0 lead to smaller WAM data sets. On the other hand, larger values of a_0 cause the function G of (5.5) to have greater variations from constant value. This condition necessarily implies a spread between possible frame bounds, i.e., movement away from tightness. This, in turn, can be related to a slowing of the rate of convergence of the reconstruction Algorithm 6.2.

We conclude this section with a list of some of the key features of the WAM speech compression process which has been developed here.

Quantization Robustness. The WAM reconstruction Algorithm 6.2 exhibits a high degree of robustness to quantization of coefficients. This property makes the WAM compression scheme well suited for communication.

Embedded Compressed Representations. An appealing property of our WAM compression is that the compressed representations Λ_δ form a decreasing continuum of sets with respect to the threshold parameter δ. In other words, if $\delta_1 < \delta_2$ then $\Lambda_{\delta_2} \subseteq \Lambda_{\delta_1}$. Thus, the compression is hierarchical in the sense that representations with small information content Λ_{δ_2} are embedded in ones with higher information content Λ_{δ_1}.

Robust Compression. It has been experimentally observed that as the WAM representation is compressed through the application of smaller thresholds δ, reconstructions based on initializing Algorithm 6.2 with the data Λ_δ degrade in a robust way. Because the compressed representations are naturally embedded (see above), the increase of the threshold δ from δ_1 to δ_2 ($\delta_1 < \delta_2$) corresponds to the removal of some elements of Λ_{δ_1}. This in turn corresponds to removing the least significant components of the reconstruction based on Λ_{δ_1}. This ensures that a small change in δ will not lead to a catastrophic change in the reconstruction.

Information Limit Transmission. An interesting inherent feature of the WAM compression scheme and representation is that, in a communication setting, all of the available bandwidth can be used for the transmission of information about the underlying signal. Suppose a fixed information transmission rate limit, e.g., the bit rate constraint b_r. Further, suppose that a threshold of $\delta = 0$ yields a finite WAM representation $\Lambda_0(f_*)$ from which it is possible to reconstruct perfectly the original signal f_*. Clearly, the cardinality of the representation $\Lambda_0(f_*)$ depends on the information content of the underlying signal f_*. For example, if $f_* \equiv 0$ then card $\Lambda_\delta = 0$ for all values of δ. For signals with low enough information content, i.e., $b_c \cdot \text{card } \Lambda_0 < b_r \cdot |I(f_*)|$, the information rate constraint poses no problem. For more complex signals, though, a threshold of 0 does not suffice to meet the information constraint. Suppose we have such a signal. In this case, an appropriate threshold must be found via the distribution function in (7.3). Since the thresholded representations are embedded (see above), the WAM compression scheme can be viewed as a method to remove the least significant coefficients from the set Λ_δ to meet the information rate constraint. Removing the least significant coefficients insures that the least amount of information will be lost. The significance of this is that (i) the WAM compression scheme yields the best thresholded representation that meets the information constraint; and (ii) if the information constraint b_r were increased, the new WAM compressed representation would contain the old WAM compressed representation. This is a consequence of the fact that the information constraint b_r can be related directly to the threshold δ by the relation

$$b_c \text{ card } \Lambda_\delta / |I(f_*)| \leqslant b_r,$$

and that card Λ_δ is a decreasing function of δ. If $b_1 < b_2$ are two information rate constraints, the two corresponding thresholded representations generated by the WAM compression scheme will be Λ_{δ_1} and Λ_{δ_2}, and $\delta_1 \geqslant \delta_2$ so that $\Lambda_{\delta_1} \subseteq \Lambda_{\delta_2}$. Consequently, over different signals (or pieces of signals) it is always possible to transmit at a rate up to the information limit b_r by adjusting the threshold δ.

Perceptually Significant Reconstruction. In all of our experiments the frequency magnitude of reconstructions closely matches that of the frequency magnitude of the original signal while phase differences are more severe. This observation suggests that the WAM representation is preserving frequency magnitude information more precisely than phase information.

8. CONCLUSION

We have presented the construction and implementation of a wavelet auditory model (WAM). From the WAM model comes a theoretical representation of acoustic signals based on irregular sampling and the theory of frames. From this

representation, we have developed an algorithm for the reconstruction of signals from their WAM representations. Further, we have developed a scheme for the compression of speech signals based on thresholded versions of the WAM representation. This scheme has been applied to many examples, including both real speech data and synthetic signals. Results for WAM compression are discussed in Section 7.

APPENDIX A. COCHLEAR SAMPLING

Let R be an instantaneous sigmoidal non-linear operator. In the high gain limit and with natural mathematical hypotheses, we verify that the derivative, $R'(\partial_t W_g(u, s))$, is the sum of the Dirac δ-measures centered at the extrema of the wavelet transform W_g and scaled by the values of the curvature of W_g around these points. This statement is expressed mathematically by Eq. (3.5).

Suppose $\lim_{T\to\infty} R_T = H$, distributionally, where R_T is defined as in Section 3 or by some other reasonable approximant of H. Suppose, further, that ϕ is a strictly monotonic continuously differentiable function on $\mathbb{R}$, and that $\phi(t_0) = 0$. Then, if ψ is a smooth compactly supported function, we have

$$\begin{aligned}(R_T'\circ\phi)(\psi) &\equiv \int R_T'(\phi)(t)\psi(t)dt \\ &= \int R_T'(u)\psi(\phi^{-1}(u))\frac{1}{|\phi'(\phi^{-1}(u))|}du \\ &\equiv \int R_T'(u)\Psi(u)du,\end{aligned}$$

where the left hand expression is defined by the first integral, where we made the substitution $u = \phi(t)$, and where $\Psi(u)$ is defined as $\psi(\phi^{-1}(u))|\phi'(\phi^{-1}(u))|^{-1}$. Since $H' = \delta$, distributionally, we see that

$$\lim_{T\to\infty}(R_T'\circ\phi)(\psi) = \Psi(0). \tag{A.1}$$

Because $\phi(t_0) = 0$, we have

$$\Psi(0) = \psi(t_0)\frac{1}{|\phi'(t_0)|} \equiv \left(\frac{1}{|\phi'(t_0)|}\delta_{t_0}\right)(\psi), \tag{A.2}$$

where δ_{t_0} is the Dirac δ-measure at t_0 and where the right hand side denotes the distributional operation of the measure $(1/|\phi'(t_0)|)\delta_{t_0}$ on the test function ψ.

Since (A.1) and (A.2) are valid for a sufficiently large class of test functions, and since

$$(\delta\circ\phi)(\psi) \equiv \lim_{T\to\infty}(R_T'\circ\phi)(\psi),$$

we conclude that

$$\delta\circ\phi = \frac{1}{|\phi'(t_0)|}\delta_{t_0}. \tag{A.3}$$

If $s > 0$ is a fixed scale, and if $\phi(t)$ is defined as $\partial_t W_g(t, s)$ and has a sequence of zeros, then (A.3) allows us to obtain (3.5) assuming there is strict monotonicity in neighborhoods of the zeros.

APPENDIX B. CAUSALITY

The *principal value distribution* $P(\gamma) = pv(1/\gamma)$ is a first order (Schwartz) distribution well defined as

$$P(\phi) = \lim_{\epsilon\to 0}\int_{|\gamma|\geq\epsilon}\frac{\phi(\gamma)}{\gamma}d\gamma$$

for continuously differentiable compactly supported test functions ϕ on $\hat{\mathbb{R}}$. It is easy to prove that the distributional derivative of the Heaviside function H is

$$\hat{H}(\gamma) = \frac{1}{2\pi i}pv\left(\frac{1}{\gamma}\right) + \frac{1}{2}\delta(\gamma).$$

The *Hilbert transform* $\mathcal{H}K$ of K is

$$\mathcal{H}K(\gamma) \equiv \frac{1}{\pi}K*pv\left(\frac{1}{\cdot}\right)(\gamma), \tag{B.1}$$

and $\mathcal{H}$ is well defined on $\hat{\mathbb{R}}$. If $\hat{f} = K = K_r + iK_i$ and supp $f \subseteq [0, \infty)$, it is an elementary but important fact that

$$K_r = \mathcal{H}K_i \quad \text{and} \quad K_i = -\mathcal{H}K_r.$$

Now suppose that $G(x + i\gamma) \equiv e^{\phi(x,\gamma)}e^{i\theta(x,\gamma)}$ is analytic in the half-plane $x > 0$, where $A \in L^2(\hat{\mathbb{R}})$ is non-negative,

$$\phi(x, \gamma) \equiv \int \log A(\lambda)d\mu_{x,\gamma}(\lambda),$$

$$d\mu_{x,\gamma}(\lambda) = \frac{1}{\pi}\frac{x}{x^2 + (\gamma - \lambda)^2}d\lambda,$$

and

$$\int\frac{\log A(\gamma)|}{1 + \gamma^2}d\gamma < \infty.$$

We show, à la Paley–Wiener, that $G \in H^2$ of the right half-plane; i.e., that

$$\sup_{x>0}\int |G(x + i\gamma)|^2 d\gamma < \infty. \tag{B.2}$$

Since the exponential function is convex we can combine Jensen's inequality, $e^{\int K d\mu} \leq \int e^K d\mu$, for the positive measure $\mu \equiv \mu_{x,\gamma}$, and Holder's inequality to compute

$$|G(x+i\gamma)|^2 \leq \frac{1}{\pi}\left(\int A(\lambda)^2 d\lambda\right)^{1/2}\left(\int\left(\frac{x}{x^2+(\gamma-\lambda)^2}\right)^2 d\lambda\right)^{1/2}.$$

Thus, $|G|$ is uniformly bounded on any half-plane $x \geq x_1 > 0$ and tends to zero uniformly in γ as $x \to \infty$. Further, by Jensen's inequality, Holder's inequality, and Fubini's theorem, we compute

$$\begin{aligned}\int |G(x+i\gamma)|^2 d\gamma &\leq \int\left|\frac{1}{\pi}\int \frac{xA(\lambda)}{x^2+(\gamma-\lambda)^2} d\lambda\right|^2 d\gamma \\ &\leq \int \frac{1}{\pi^2}\int \frac{xA(\lambda)^2}{x^2+(\gamma-\lambda)^2} d\lambda \int \frac{x}{x^2+(\gamma-\nu)^2} d\nu d\gamma \\ &= \frac{1}{\pi}\int\left(\int \frac{xA(\lambda)^2}{x^2+(\gamma-\lambda)^2} d\gamma\right) d\lambda = \int A(\lambda)^2 d\lambda,\end{aligned}$$

and (B.2) is obtained.

For this situation, we can invoke the (first) classical Paley–Wiener theorem [30, p. 20; 31, p. 8] to infer that there is a causal function $g \in L^2(\mathbb{R})$ for which $\hat{g} = G(i\gamma)$ a.e.

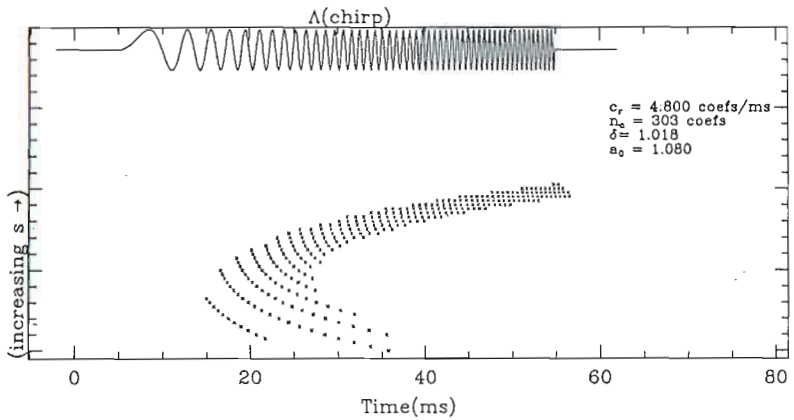

FIG. 8. Synthesized chirp signal and its thresholded WAM representation.

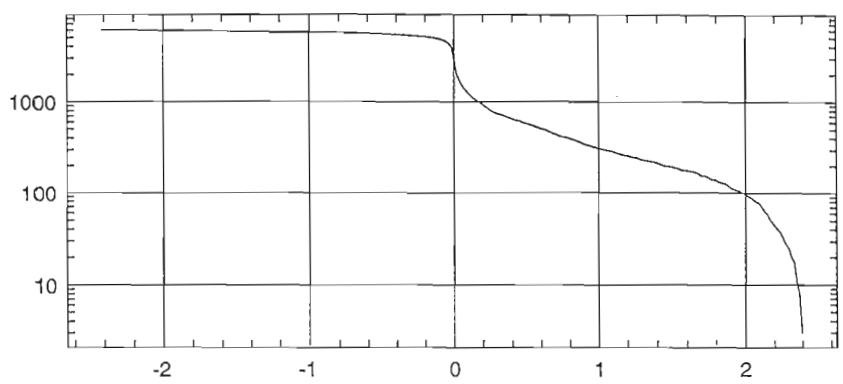

FIG. 9. λ distribution of the synthesized chirp signal.

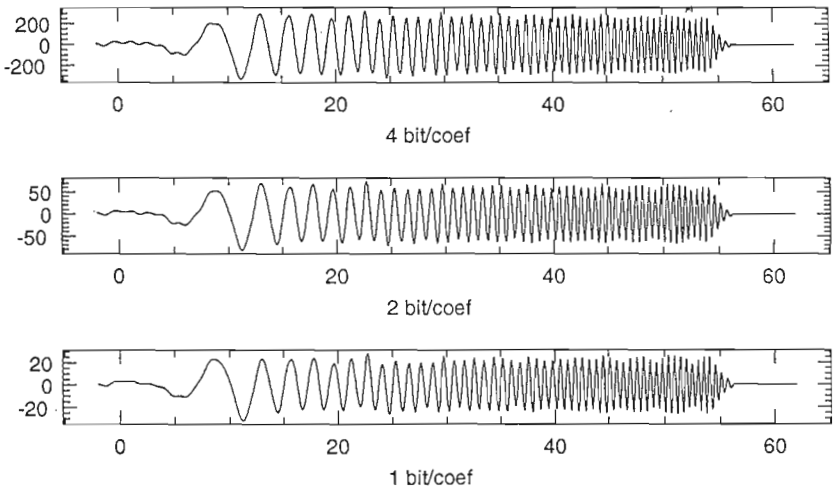

FIG. 10. Time domain reconstructions of synthesized chirp signal as the bit allocation per coefficient b_c varies (1, 2, and 4 bits per coefficient).

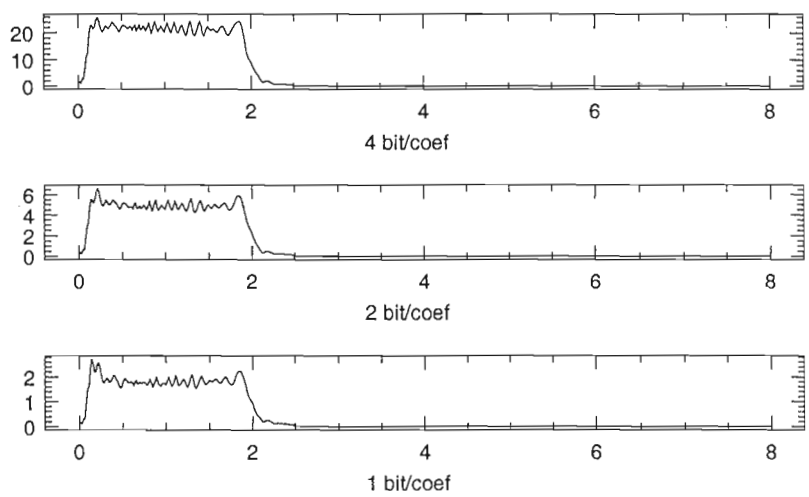

FIG. 11. Frequency domain reconstructions of synthesized chirp signal as the bit allocation per coefficient b_c varies (1, 2, and 4 bits per coefficient).

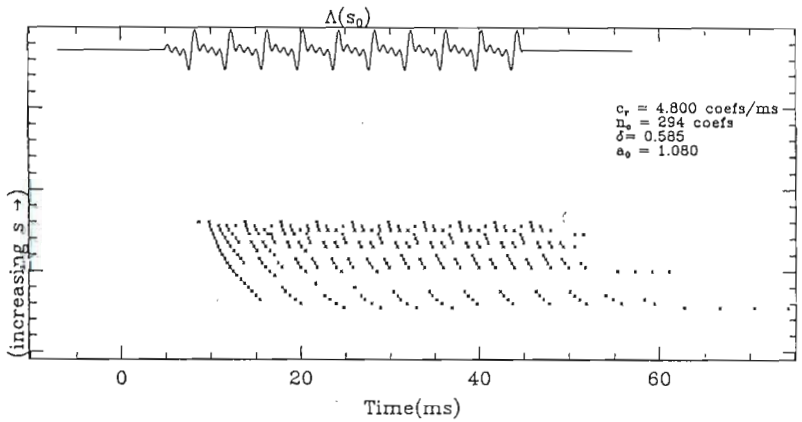

FIG. 12. Synthesized four component sine wave and its thresholded WAM representation.

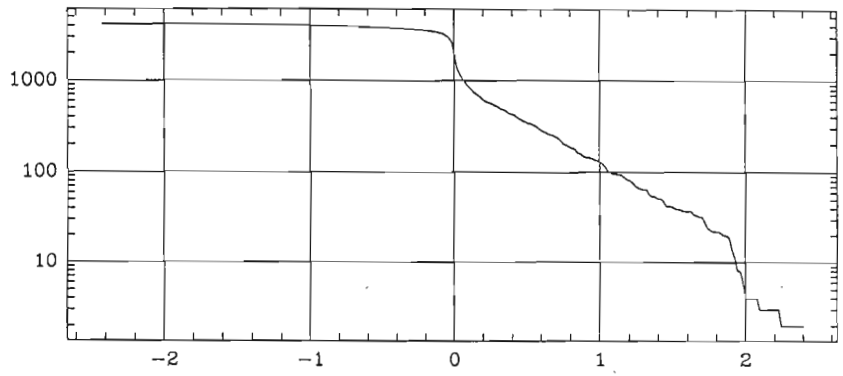

FIG. 13. λ distribution of the synthesized four component sine wave.

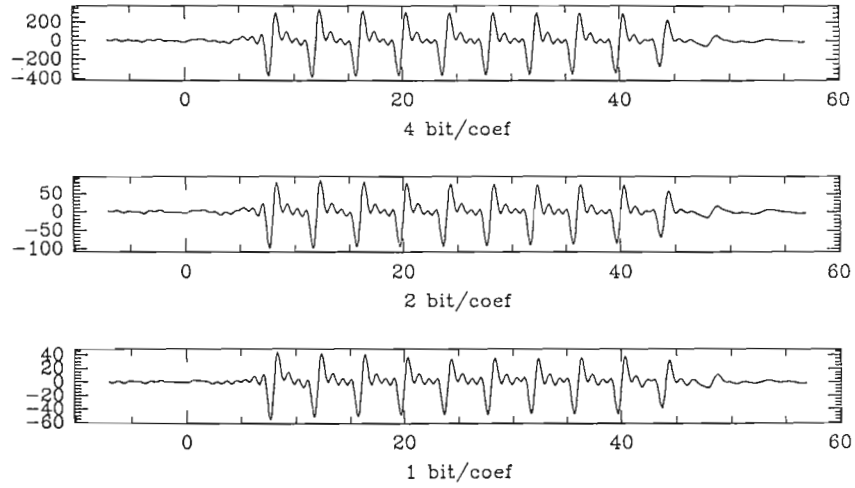

FIG. 14. Time domain reconstructions of synthesized four component sine wave as the bit allocation per coefficient b_c varies (1, 2, and 4 bits per coefficient).

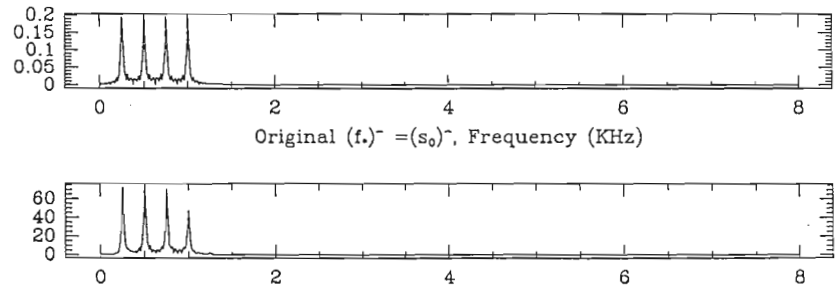

FIG. 15. Frequency domain reconstruction of synthesized four component sine wave with bit allocation per coefficient $b_c = 4$ bits per coefficient.

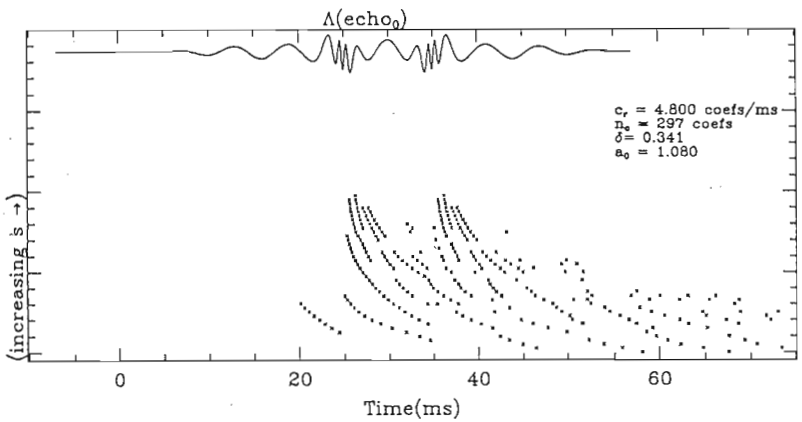

FIG. 16. Synthesized FM echo signal and its threshold WAM representation.

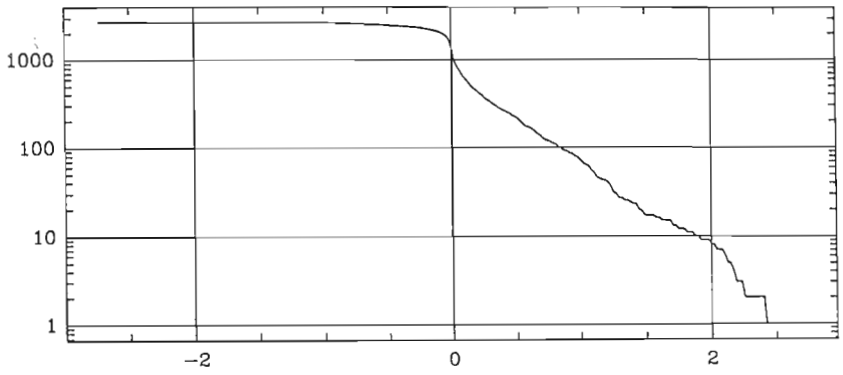

FIG. 17. λ distribution of the synthesized FM echo signal.

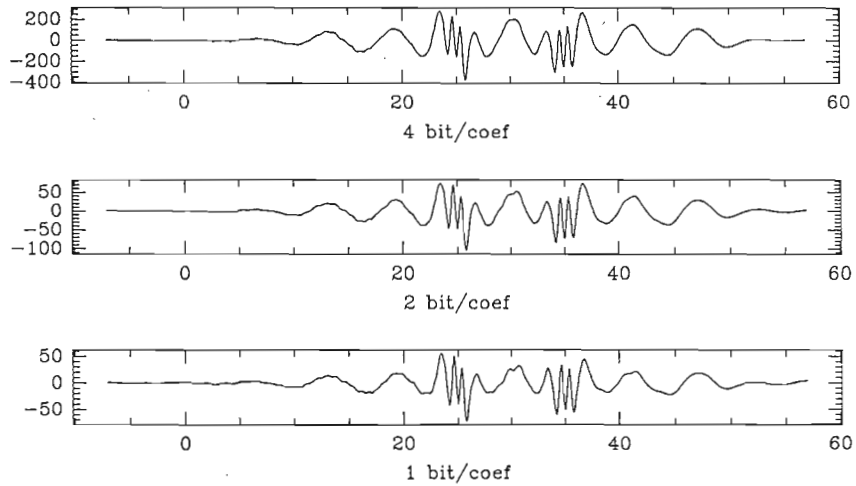

FIG. 18. Time domain reconstructions of synthesized FM echo signal as the bit allocation per coefficient b_c varies (1, 2, and 4 bits per coefficient).

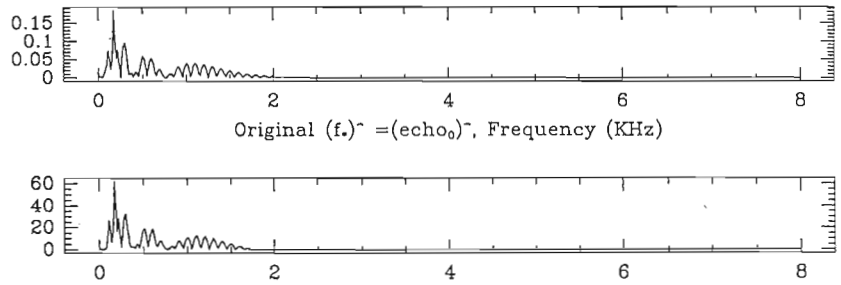

FIG. 19. Frequency domain reconstruction of synthesized FM echo signal with bit allocation per coefficient b_c = 4 bits per coefficient.

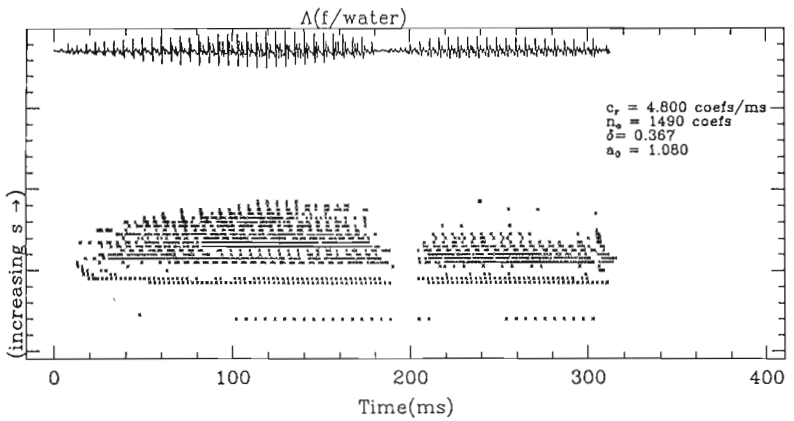

FIG. 20. Female spoken word "water" and its thresholded WAM representation.

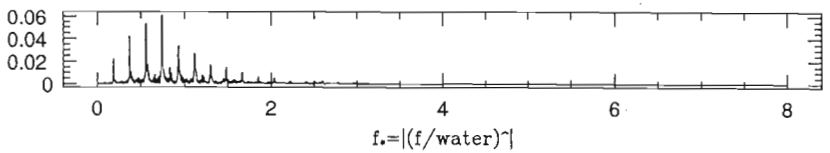

FIG. 21. Magnitude of the Fourier transform of the female spoken word "water."

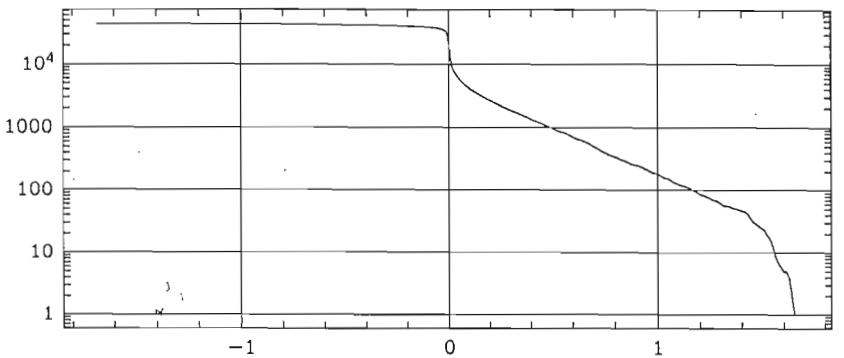

FIG. 22. λ distribution of the female spoken word "water."

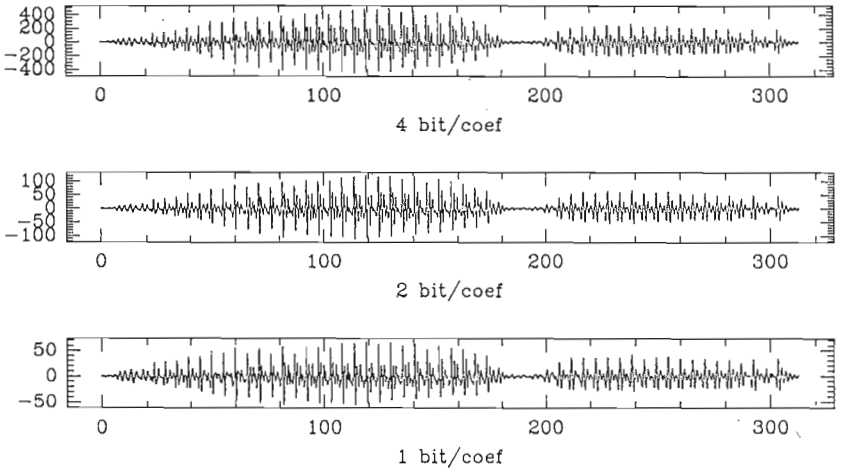

FIG. 23. Time domain reconstructions of the female spoken word "water" as the bit allocation per coefficient b_c varies (1, 2, and 4 bits per coefficient).

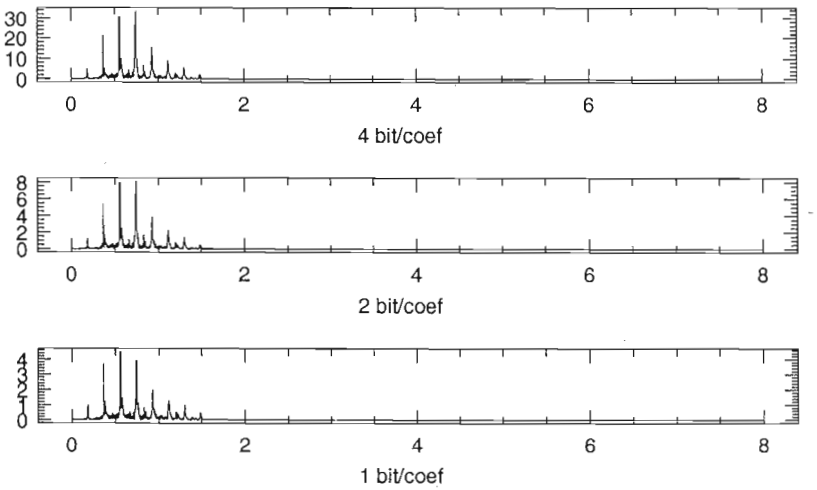

FIG. 24. Magnitude of the Fourier transform of reconstructions of the female spoken word "water" as the bit allocation per coefficient b_c varies (1, 2, and 4 bits per coefficient).

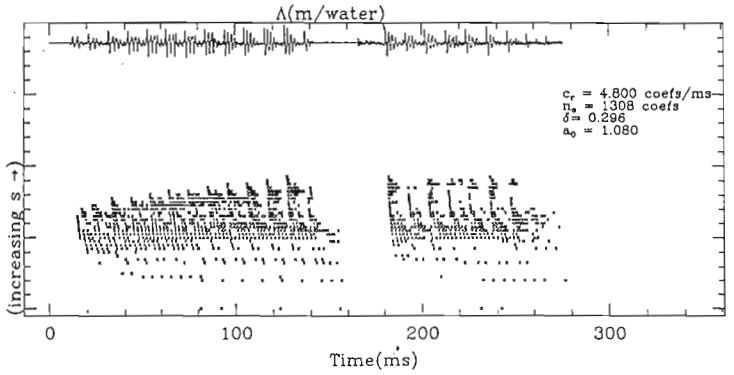

FIG. 25. Male spoken word "water" and its thresholded WAM representation.

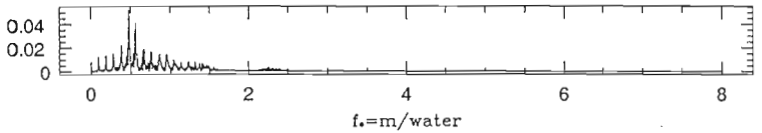

FIG. 26. Magnitude of the Fourier transform of the male spoken word "water."

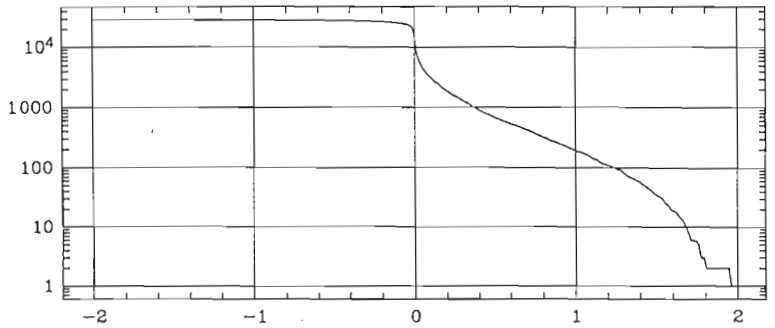

FIG. 27. λ distribution of the male spoken word "water."

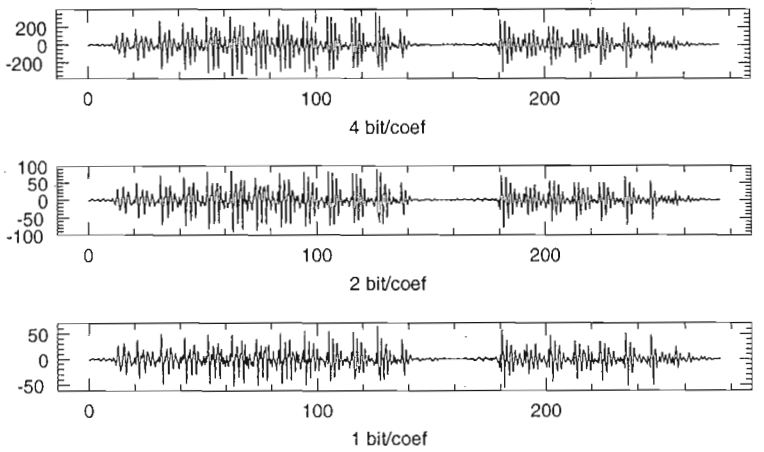

FIG. 28. Time domain reconstructions of the male spoken word "water" as the bit allocation per coefficient b_c varies (1, 2, and 4 bits per coefficient).

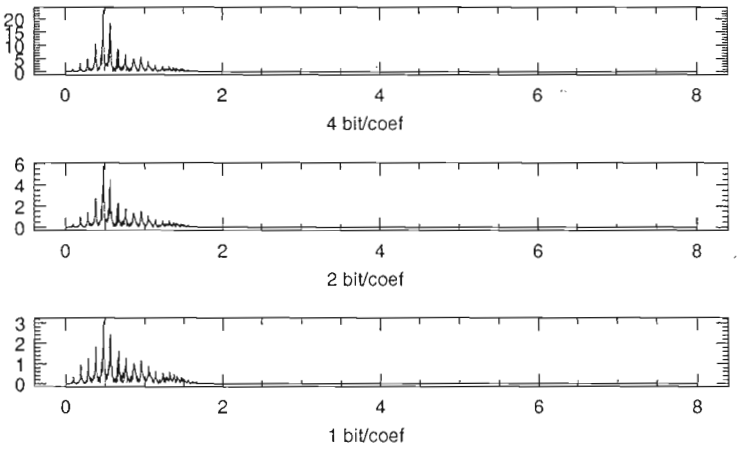

FIG. 29. Magnitude of the Fourier transform of reconstructions of the male spoken word "water" as the bit allocation per coefficient b_c varies (1, 2, and 4 bits per coefficient).

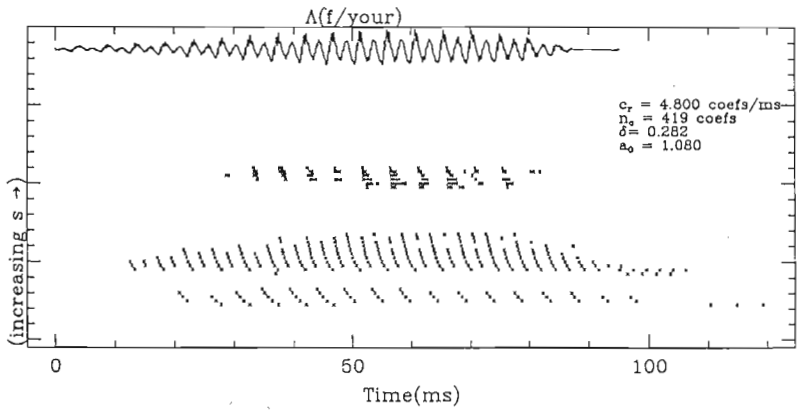

FIG. 30. Female spoken word "your" and its thresholded WAM representation.

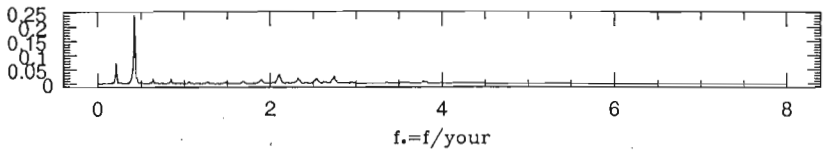

FIG. 31. Magnitude of the Fourier transform of the female spoken word "your."

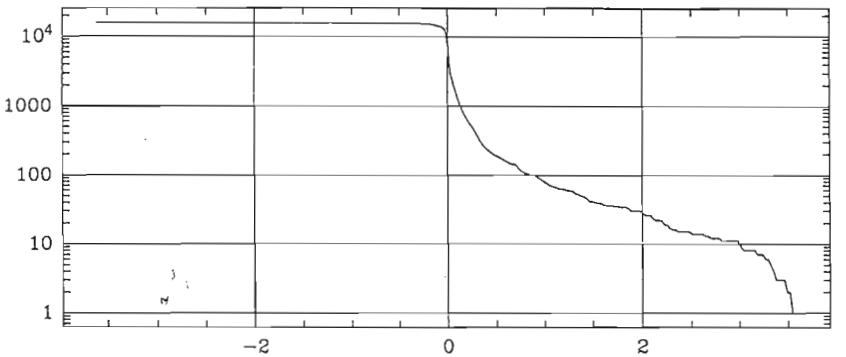

FIG. 32. λ distribution of the female spoken word "your."

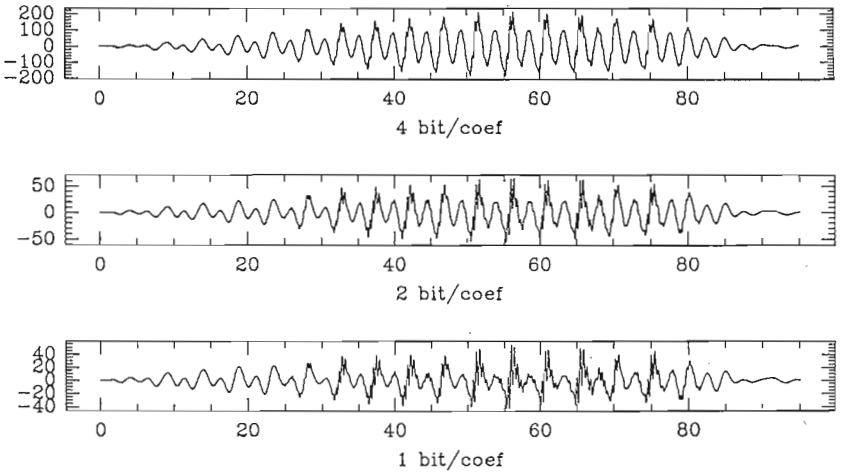

FIG. 33. Time domain reconstructions of the female spoken word "your" as the bit allocation per coefficient b_c varies (1, 2, and 4 bits per coefficient).

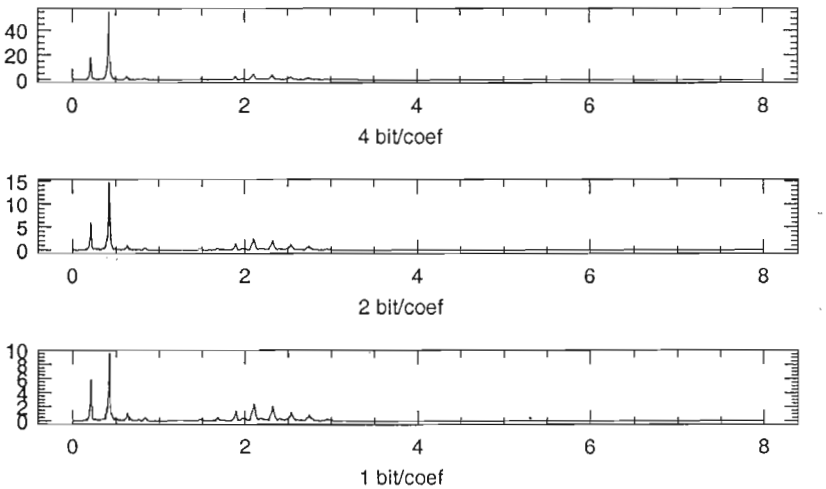

FIG. 34. Magnitude of the Fourier transform of reconstructions of the female spoken word "your" as the bit allocation per coefficient b_c varies (1, 2, and 4 bits per coefficient).

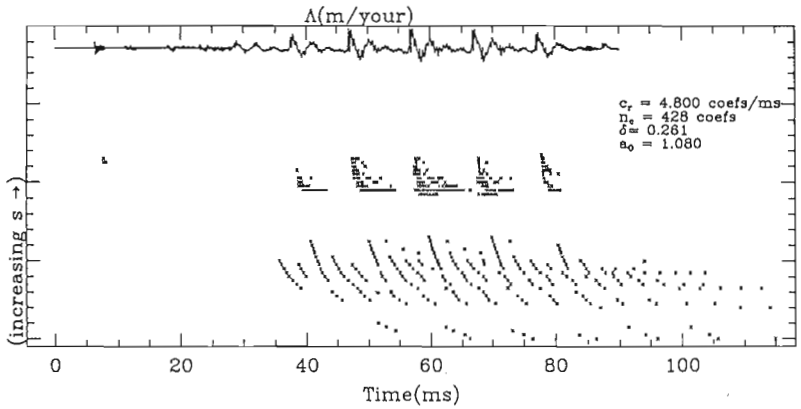

FIG. 35. Male spoken word "your" and its thresholded WAM representation.

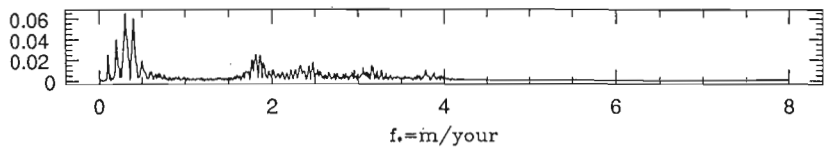

FIG. 36. Magnitude of the Fourier transform of the male spoken word "your."

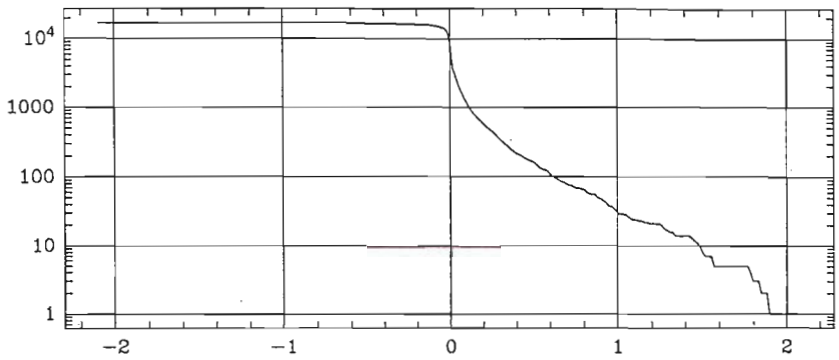

FIG. 37. λ distribution of the male spoken word "your."

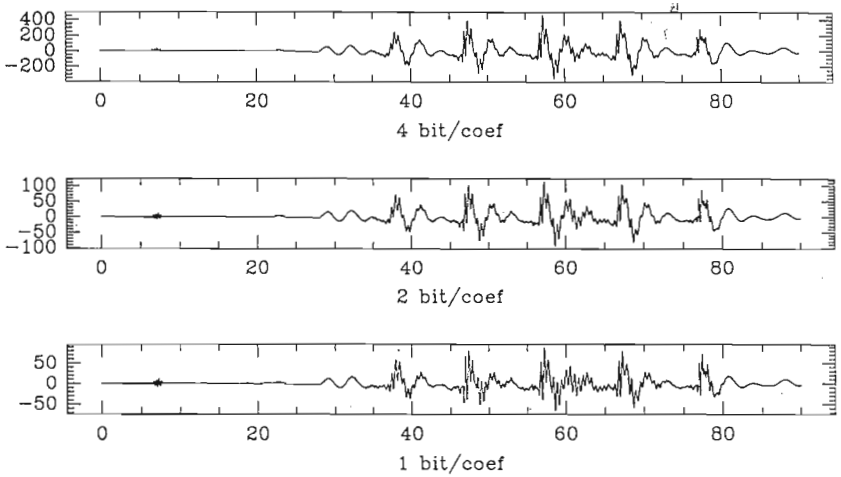

FIG. 38. Time domain reconstructions of the male spoken word "your" as the bit allocation per coefficient b_c varies (1, 2, and 4 bits per coefficient).

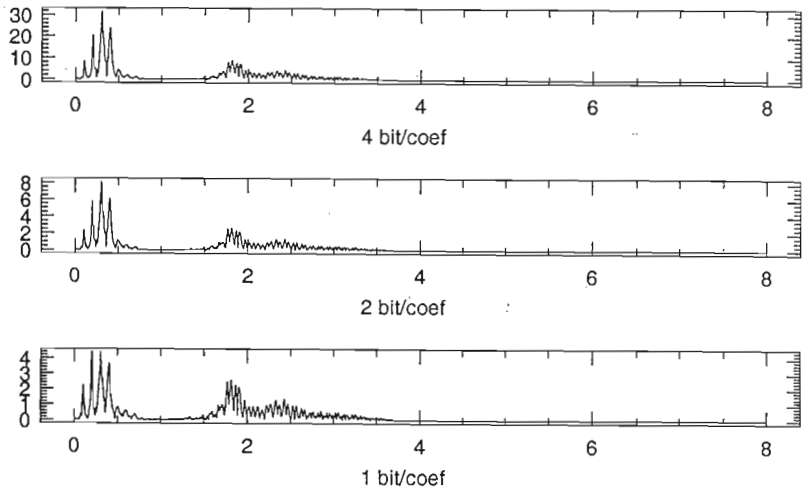

FIG. 39. Magnitude of the Fourier transform of reconstructions of the male spoken word "your" as the bit allocation per coefficient b_c varies (1, 2, and 4 bits per coefficient).

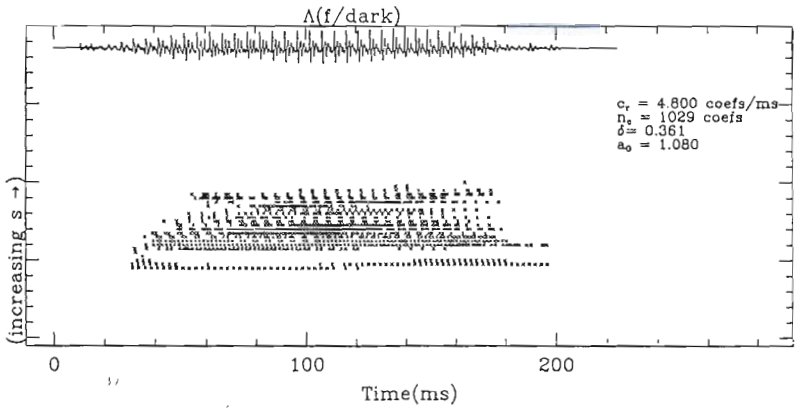

FIG. 40. Female spoken word "dark" and its thresholded WAM representation.

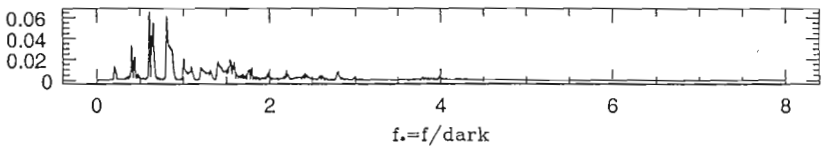

FIG. 41. Magnitude of the Fourier transform of the female spoken word "dark."

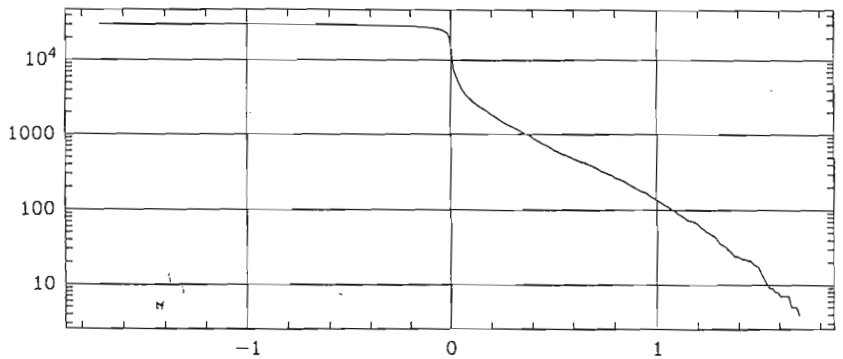

FIG. 42. λ distribution of the female spoken word "dark."

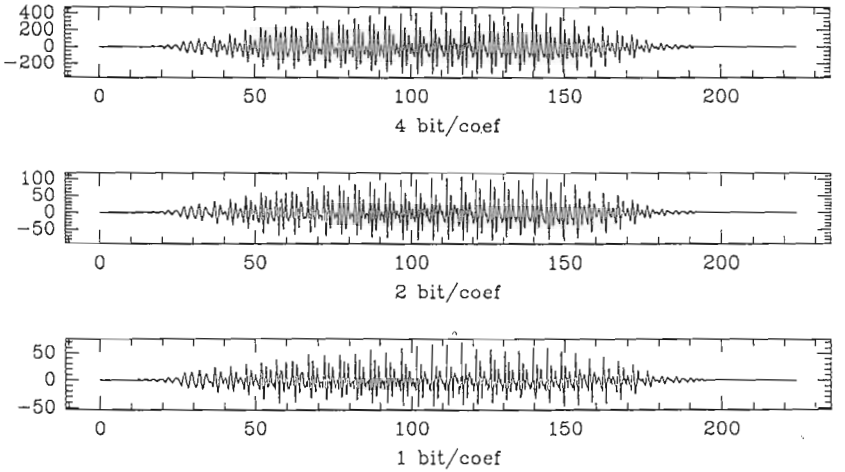

FIG. 43. Time domain reconstructions of the female spoken word "dark" as the bit allocation per coefficient b_c varies (1, 2, and 4 bits per coefficient).

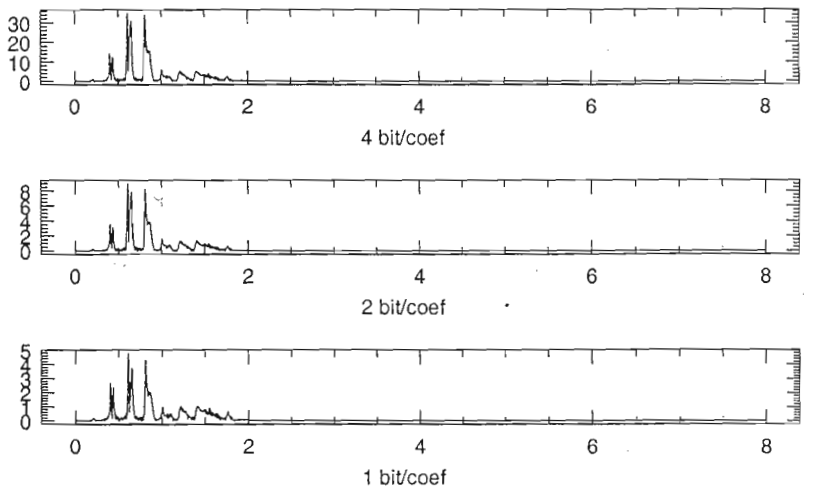

FIG. 44. Magnitude of the Fourier transform of reconstructions of the female spoken word "dark" as the bit allocation per coefficient b_c varies (1, 2, and 4 bits per coefficient).

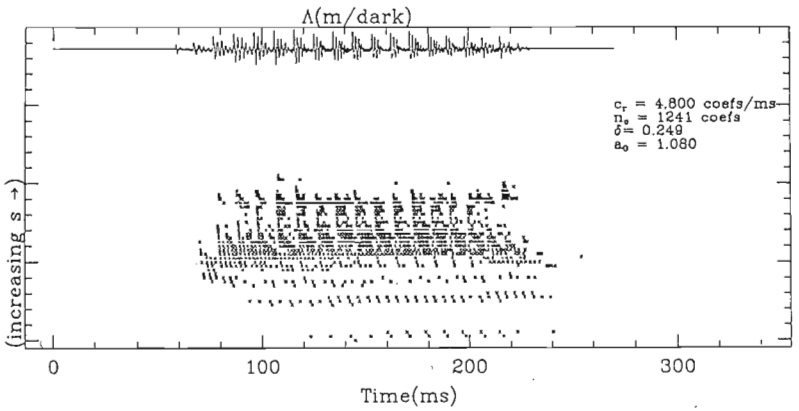

FIG. 45. Male spoken word "dark" and its thresholded WAM representation.

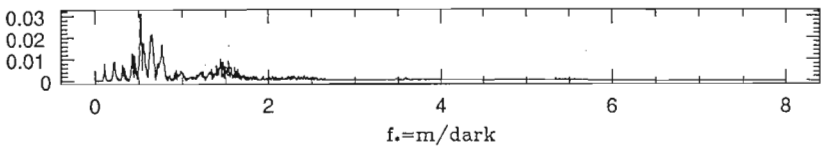

FIG. 46. Magnitude of the Fourier transform of the male spoken word "dark."

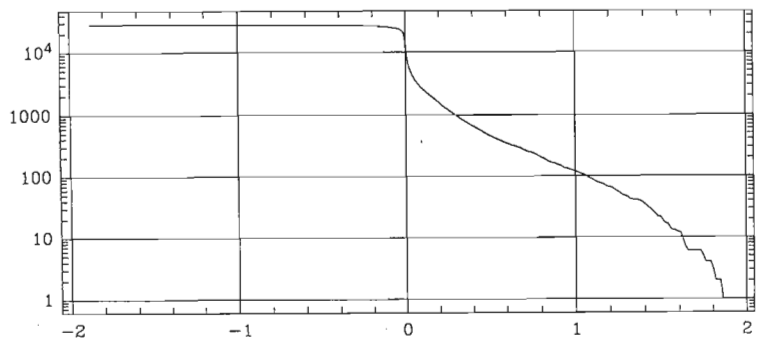

FIG. 47. λ distribution of the male spoken word "dark."

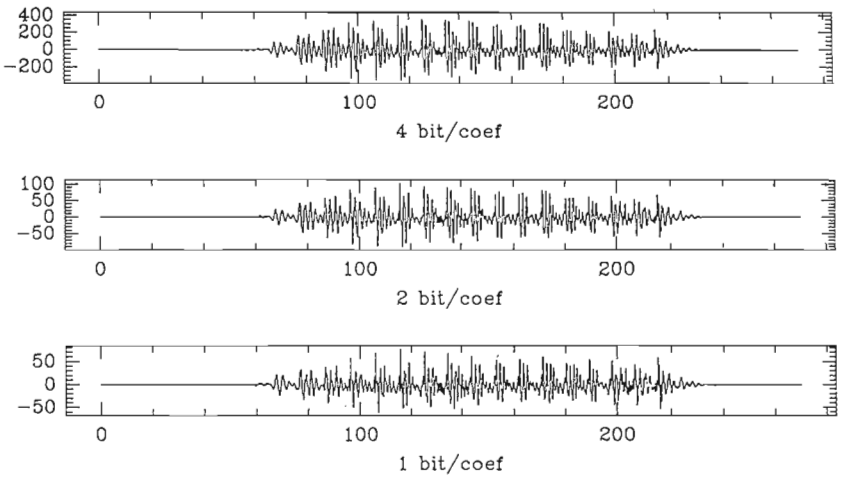

FIG. 48. Time domain reconstructions of the male spoken word "dark" as the bit allocation per coefficient b_c varies (1, 2, and 4 bits per coefficient).

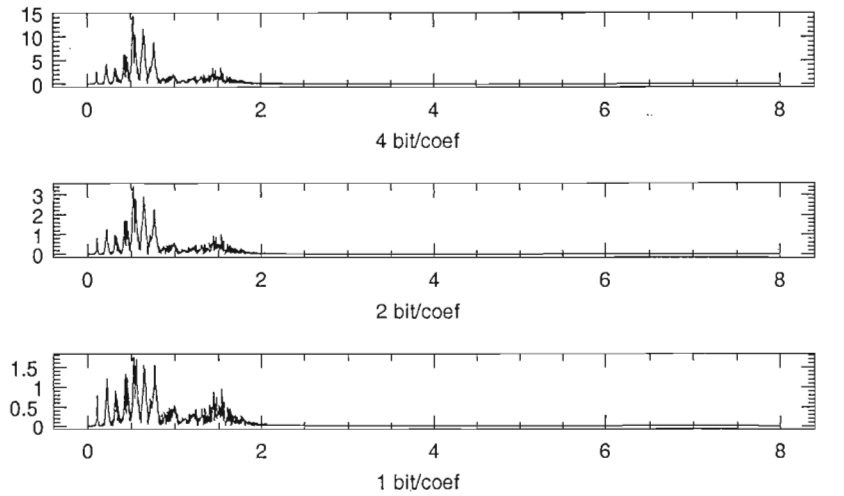

FIG. 49. Magnitude of the Fourier transform of reconstructions of the male spoken word "dark" as the bit allocation per coefficient b_c varies (1, 2, and 4 bits per coefficient).

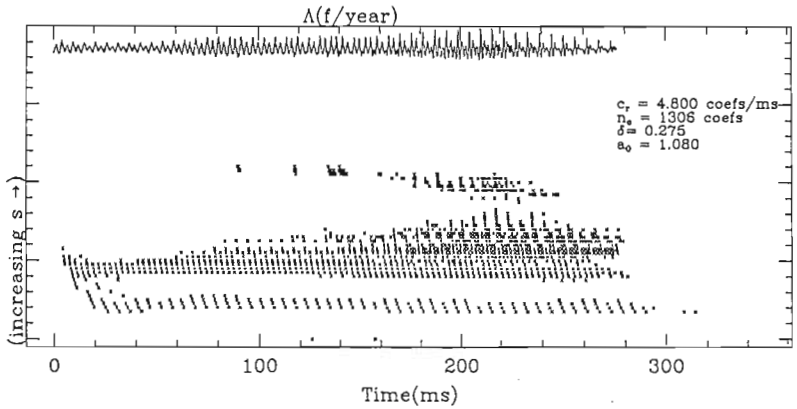

FIG. 50. Female spoken word "year" and its thresholded WAM representation.

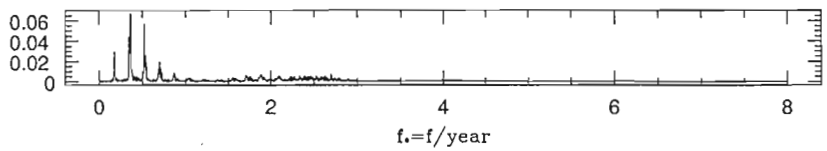

FIG. 51. Magnitude of the Fourier transform of the female spoken word "year."

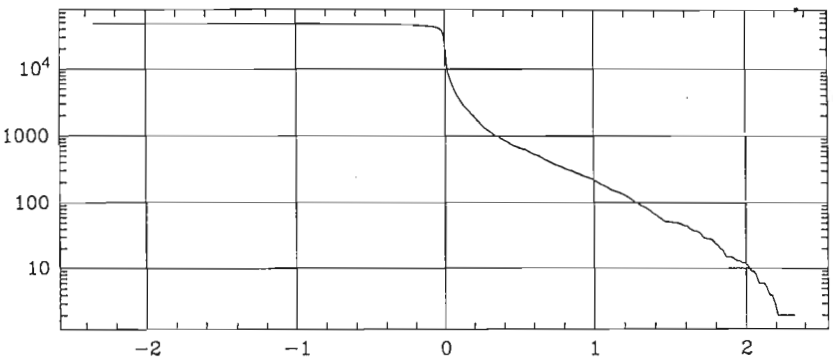

FIG. 52. λ distribution of the female spoken word "year."

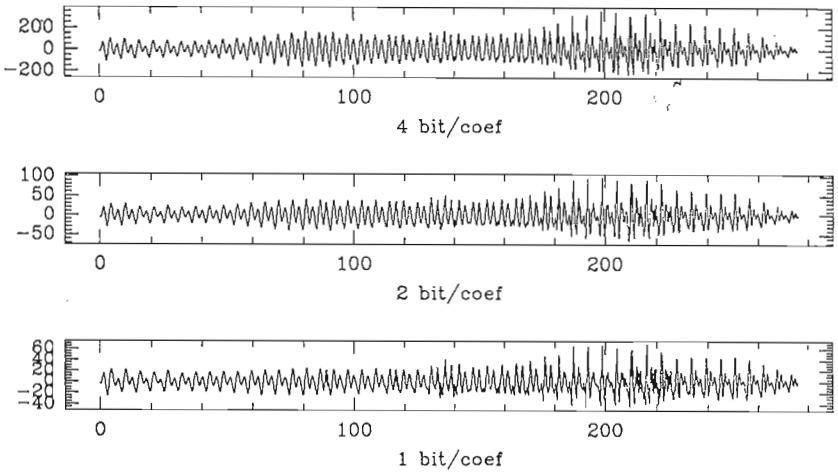

FIG. 53. Time domain reconstructions of the female spoken word "year" as the bit allocation per coefficient b_c varies (1, 2, and 4 bits per coefficient).

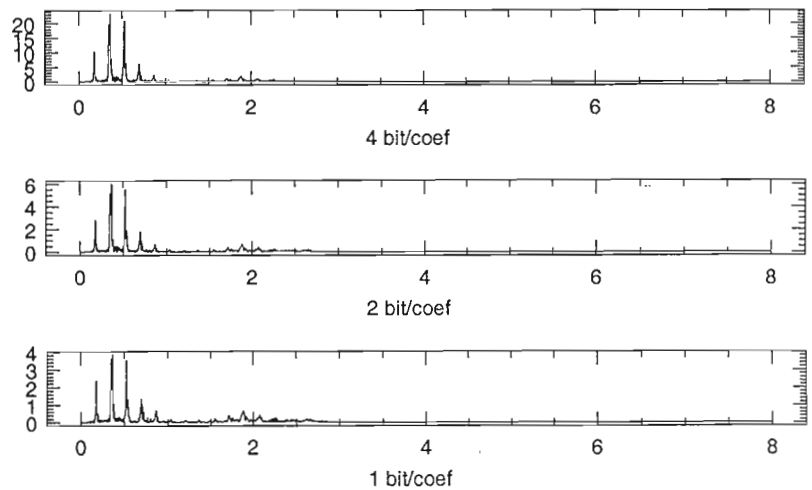

FIG. 54. Magnitude of the Fourier transform of reconstructions of the female spoken word "year" as the bit allocation per coefficient b_c varies (1, 2, and 4 bits per coefficient).

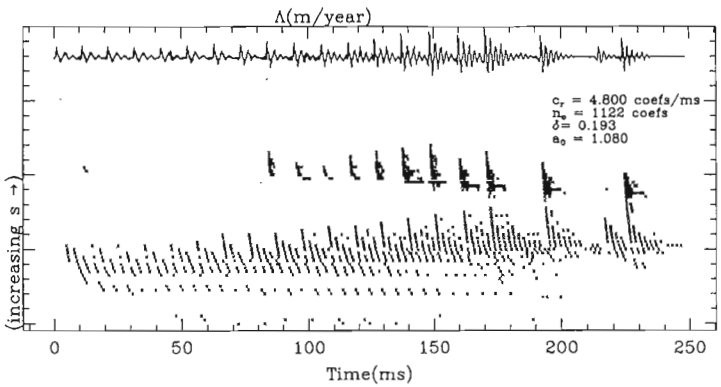

FIG. 55. Male spoken word "year" and its thresholded WAM representation.

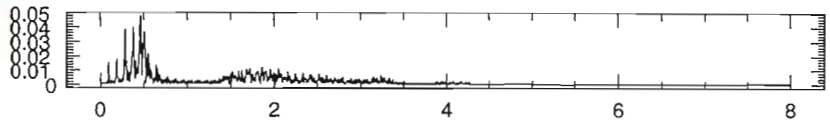

FIG. 56. Magnitude for the Fourier transform of the male spoken word "year."

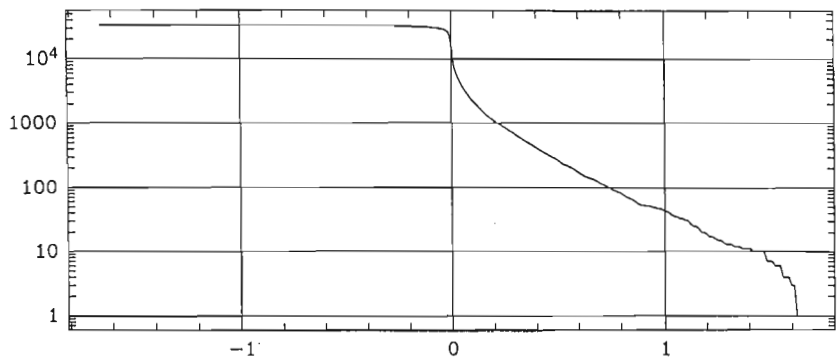

FIG. 57. λ distribution of the male spoken word "year."

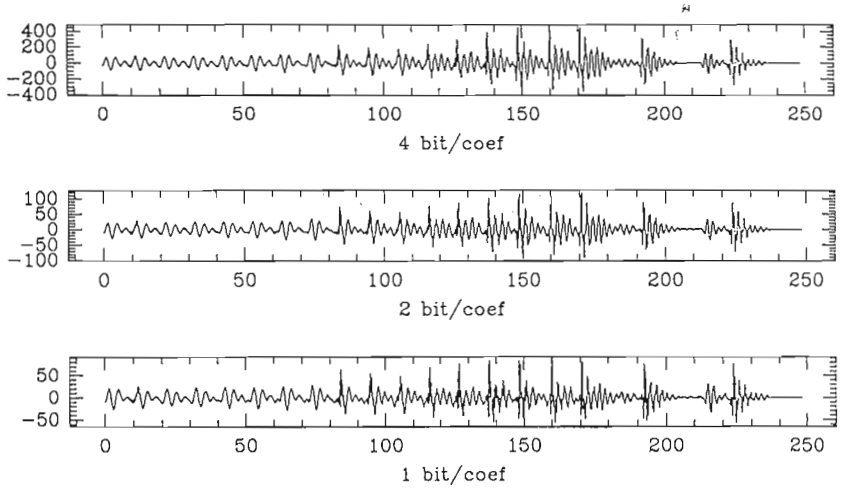

FIG. 58. Time domain reconstructions of the male spoken word "year" as the bit allocation per coefficient b_c varies (1, 2, and 4 bits per coefficient).

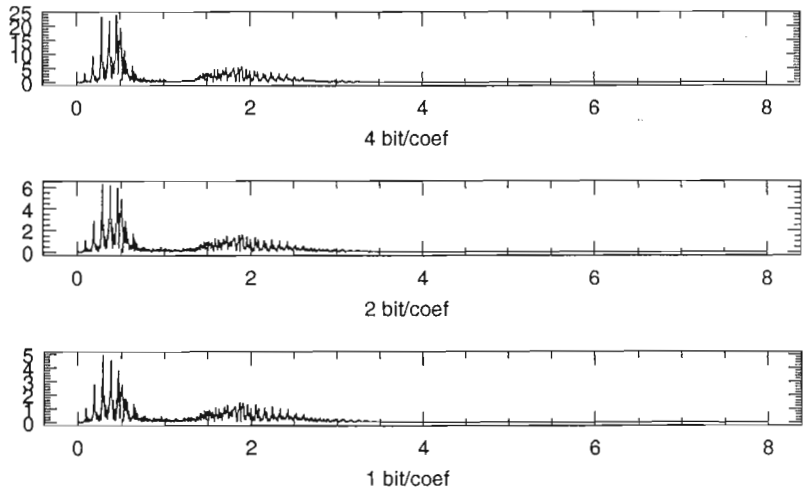

FIG. 59. Magnitude of the Fourier transform of reconstructions of the male spoken word "year" as the bit allocation per coefficient b_c varies (1, 2, and 4 bits per coefficient).

ACKNOWLEDGMENT

It is our pleasure to thank A. Ephremides, W. Evans, and S. Shamma for sharing their expertise and insights with us on several critical issues in this work.

REFERENCES

1. J. Allen, Cochlear modeling, *IEEE Acoust. Speech Signal Process Mag.* **2** (1985), 3–29.
2. B. Atal and M. Schroeder, Predictive coding of speech signals, *in* "Proceedings, 1967 Conference on Communications and Processing," pp. 360–361.
3. B. Atal and M. Schroeder, Adaptive predictive coding of speech signals, *Bell System Tech. J.* **49** (1970), 1973–1986.
4. G. Battle, Phase space localization theorem for ondelettes, *J. Math. Phys.* **30** (1989), 2195–2196.
5. J. Benedetto, "Real Variables and Integration," Teubner, Stuttgart, 1976.
6. J. Benedetto, "Wavelet Auditory Models and Irregular Sampling," Technical Report, Prometheus Inc., Newport, RI 02840, 1990.
7. J. Benedetto, Irregular sampling and frames, *in* "Wavelets: A Tutorial in Theory and Applications" (C. Chui, Ed.), pp. 445–507, Academic Press, Boston, 1992.
8. J. Benedetto, Frame decompositions, sampling and uncertainty principle inequalities, *in* "Wavelets: Mathematics and Applications" (J. Benedetto and M. Frazier, Eds.), CRC Press, Boca Raton, FL, 1993.
9. J. Benedetto and W. Heller, Irregular sampling and the theory of frames, *Mat. Note* **10,** Suppl. 1, (1990), 103–125.
10. J. Benedetto and A. Teolis, An auditory motivated time-scale signal representation, *in* "IEEE-SP International Symposium on Time-Frequency and Time-Scale Analysis," October 1992.
11. J. Cohen, Application of an auditory model to speech recognition, *J. Acoust. Soc. Amer.* **85** (1989), 2623–2629.
12. I. Daubechies, The wavelet transform, time-frequency localization and signal analysis, *IEEE Trans. Inform. Theory* **36,** No. 5 (1990), 961–1005.
13. I. Daubechies, "Ten Lectures on Wavelets," CBMS-NSF Regional Conference Series in Applied Mathematics, SIAM, Philadelphia, 1992.
14. L. Deng, C. Geisler, and S. Greenberg, A composite of the auditory periphery for the processing of speech, *J. Phonet.* **16** (1988), 93–108.
15. I. Daubechies, A. Grossman, and Y. Meyer, Painless nonorthogonal expansions, *J. Math. Phys.* **27,** No. 5 (1986), 1271–1283.
16. R. Duffin and A. Schaeffer, A class of nonharmonic Fourier series, *Trans. Amer. Math. Soc.* **72** (1952), 341–366.
17. C. G. Fant, "Speech Sounds and Features, Technical Report," MIT, Cambridge, MA, 1973.
18. G. Geisler, Representation of speech sounds on the auditory nerve, *J. Phonet.* **16** (1988), 19–35.
19. O. Ghitza, Temporal non-place information in the auditory nerve firing patterns as a front-end for speech recognition in a noisy environment, *J. Phonet.* **16** (1988), 109–124.
20. S. Greenberg, Acoustic transduction in the auditory periphery, *J. Phonet.* **16** (1988), 628–666.
21. K. Hoffman, "Banach Spaces of Analytic Functions," New York, 1962.
22. C. Heil and D. Walnut, Continuous and discrete wavelet transforms, *SIAM Rev.* **31** (1989), 628–666.
23. F. Itakura and S. Saito, Paper c-5-5, *in* "Proceedings, International Congress on Acoustics," Aug. 1968.
24. S. Jaffard, A density criterion for frames of complex exponentials, *Michigan Math J.* **38** (1991), 339–348.
25. D. Klatt, The representation of speech in the peripheral auditory system, *in* "Speech Processing Strategies Based on Auditory Models" (R. Carlson and B. Granström, Eds.), pp. 181–196, Elsevier Biomedical Press, Amsterdam, 1982.
26. W. Liu, A. Andreou, and M. Goldstein, Voiced speech representation by an analog silicon model of the auditory periphery, *IEEE Trans. Neural Networks* **3** (1992), 477–487.
27. Y. Meyer, "Ondelettes et Opérateurs," Vols. I, II, and III, Hermann, Paris, 1990.
28. I. Morishita and A. Yajima, Analysis and simulation of networks of mutually inhibiting neurons, *Kybernetik* **11** (1972), 154–165.
29. R. R. Pfeiffer and D. O. Kim, Cochlear nerve fiber responses: Distribution along the cochlear partition, *J. Acoust. Soc. Amer.* **58** (1975), 867–869.
30. P. Porcelli, "Linear Spaces of Analytic Functions," Rand McNally, Chicago, 1966.
31. R. Paley and N. Wiener, "Fourier Transforms in the Complex Domain," Vol. 19, Amer. Math. Soc., Colloquium Publications, Providence, RI, 1934.
32. R. J. Ritsma, Frequencies dominant in the perception of pitch of complex sounds. *J. Acoust. Soc. Amer.* **42** (1967), 191–198.
33. S. Shamma, Speech processing in the auditory system. I. The representation of speech sounds in the responses of the auditory nerve, *J. Acoust. Soc. Amer.* **78** (1985), 1612–1621.
34. S. Shamma, Speech processing in the auditory system, II. Lateral inhibition and the central processing of speech evoked activity in the auditory nerve, *J. Acoust. Soc. Amer.* **78** (1985), 1622–1632.
35. E. Stein and G. Weiss, "An Introduction to Fourier Analysis on Euclidean Spaces," Princeton Univ. Press, Princeton, NJ, 1971.
36. A. Teolis and J. Benedetto, "The Theory of Local Frames," Technical Report, Institute for Systems Research, University of Maryland, College Park, MD, 1992.
37. A. Teolis, "Discrete Signal Representation," Ph.D. Thesis, University of Maryland, College Park, MD, 1993.
38. R. M. Young, "An Introduction to Nonharmonic Fourier Series," Academic Press, New York, 1980.
39. X. Yang, K. Wang, and S. Shamma, Auditory representations of acoustic signals, *IEEE Trans. Inform. Theory,* **38,** No. 2 (1992), 824–839.
40. G. Zweig, Finding the impedance of the organ of Corti, *J. Acoust. Soc. Amer.* **89,** No. 3 (1991), 1229–1254.