A System for Finding Speech Formants and Modulations via Energy Separation

Helen M. Hanson, Petros Maragos, and Alexandros Potamianos

Abstract—This correspondence presents an experimental system that uses an energy-tracking operator and a related energy separation algorithm to automatically find speech formants and amplitude/frequency modulations in voiced speech segments. Initial estimates of formant center frequencies are provided by either LPC or morphological spectral peak picking. These estimates are then shown to be improved by a combination of bandpass filtering and iterative application of energy separation.

I. INTRODUCTION

The ability to automatically find and track resonant frequencies of the speech production system, called "formants," is an important part of speech processing, because formants play a major role in most speech applications [10]. Traditional methods for formant finding are peak picking of the cepstrally-smoothed or LPC spectrum [8], [11], or finding the roots of the LPC polynomial [1]. These methods assume that the formants are constant within an analysis frame. However, in a recently proposed modulation model [4]-[6], resonances are modeled as damped AM-FM signals

\[ x(t) = a(t) \cos(\phi(t)) \]

\[ = a(t) \cos(2\pi f_c t + f_m \int_0^t q(\tau) d\tau + \phi(0)) \]

(1)

with a time-varying instantaneous frequency (in hertz)

\[ f(t) = \frac{1}{2\pi} \frac{d}{dt} \phi(t) = f_c + f_m q(t) \]

(2)

and a (generally nonexponential) amplitude \( a(t) \). The formant frequency \( f_c \) may vary around its center value \( f_c \), according to a frequency modulating signal \( q(t) \in [-1, 1] \), with \( f_m \in (0, f_c) \) the maximum deviation. To estimate the amplitude and frequency signals, Maragos, Kaiser, and Quatieri [5] developed an energy separation algorithm (ESA) that uses a nonlinear energy operator to track the instantaneous energy of the source generating the AM-FM signal and separate it into its amplitude and frequency components. This energy operator, defined as \( \Psi_s[x(t)] = E[x(t)^2] - E[x(t^2)] \), was developed by Teager [13] and Kaiser [2], [3].

To apply the ESA to speech, one must isolate the resonances by bandpass filtering. We do this with a Gabor filter having impulse response \( g(t) = \exp(-\alpha t^2) \cos(\omega t) \), \( \omega = 2\pi f_c \), and frequency response

\[ G(\omega) = \frac{\sqrt{\pi}}{2\pi} \left( \exp\left(-\frac{(\omega - \omega_0)^2}{4\alpha^2}\right) + \exp\left(-\frac{(\omega + \omega_0)^2}{4\alpha^2}\right) \right) \]

(3)

II. BACKGROUND

Given an AM-FM signal \( x(t) = a(t) \cos(\phi(t)) \) as in (1), it has been shown in [4], [6] that

\[ \Psi_s[x(t)] \approx [a(t) \phi(t)]^2 \]

(4)

To separate amplitude from frequency in the above energy product, it has been shown in [5] that

\[ \frac{1}{2\pi} \sqrt{\Psi_s[x(t)]} \approx f(t) \]

(5)

\[ \frac{\Psi_s[x(t)]}{\sqrt{\Psi_s[x(t)]}} \approx |a(t)| \]

(6)

Thus, (5) and (6), referred to as the continuous-time energy separation algorithm (ESA), can estimate the amplitude envelope \( |a(t)| \) and instantaneous frequency \( f(t) \) of an AM-FM signal at each time instant. The approximations in (4)-(6) are valid under certain general, realistic constraints that restrict the bandwidths of \( a(t) \), \( f(t) \) and the maximum frequency deviation \( f_m \) of the FM part to be much smaller than the carrier frequency \( f_c \).

A similar set of equations has been derived for discrete-time AM-FM signals, defined by \( x[n] = a(n) \cos[\phi(n)] = a(n) \cos(2\pi T f_m n + \phi(n)) \), where \( |a(n)| \) is the discrete-time amplitude envelope, \( f(n) \) is the instantaneous frequency (in hertz), and a sampled version of (2) \( T \) is the sampling period. By applying the discrete-time Teager-Kaiser energy operator 2, \( \Psi_s[x(n)] = x'^2(n) - x(n-1)x(n+1) \), to the signal \( x(n) \) and its backward difference, \( y(n) = x(n) - x(n-1) \), it is shown in [4]-[6] that

\[ \Psi_s[x(n)] \approx a^2(n) \sin^2[2\pi T f(n)] \]

Manuscript received June 27, 1992; revised October 12, 1993. This work was supported by the National Science Foundation under Grant MIP-91-20624, by an NSF Presidential Young Investigator Award under Grant MIP-86-58150 with matching funds from Xerox, and by M.I.T. Lincoln Laboratory.

P. M. Hansen is with the Research Lab of Electronics, M.I.T., Cambridge, MA 02139 USA.

P. Maragos and A. Potamianos were with the Division of Applied Sciences, Harvard University, Cambridge, MA 02138 USA. They are now with the School of Electrical Engineering, Georgia Institute of Technology, Atlanta, GA 30332 USA.

IEEE Log Number 9400746.

1063-6676/94$04.00 © 1994 IEEE
Fig. 1. (a) The Fourier magnitude spectrum of 25 ms of the vowel /e/. The Gabor filters (center frequencies at 3200, 3440, and 3700 Hz) used for isolating the formant that peaks at 3440 Hz are superimposed on the spectrum. (b) The instantaneous formant frequency, \( f(n) \), when the filter has center frequency \( f_c = 3200 \) Hz. (c) \( f(n) \) when \( f_c = 3440 \) Hz (the formant peak frequency). (d) \( f(n) \) when \( f_c = 3700 \) Hz. In (b)-(d) the center frequency of the filter and the average instantaneous frequency are indicated with dashed and dotted lines, respectively.

\[
\frac{1}{2\pi T} \arccos \left( 1 - \frac{\Psi_d[p(n)] + \Psi_d[p(n+1)]}{4\Psi_d[x(n)]} \right) \approx f(n) \tag{7}
\]

\[
\sqrt{1 - \left( \frac{\Psi_d[x(n)]}{\Psi_d[x(n)+\Psi_d[x(n+1)]} \right)^2} \approx |a(n)|. \tag{8}
\]

Equations (7) and (8) are referred to as the discrete-time ESA. At each sample it provides an estimate of the envelope and instantaneous frequency using only a 5-sample moving window, at a very small computational complexity. The approximations involved are valid as long as the amplitude envelope and instantaneous frequency do not change too much or too quickly in time compared with the carrier frequency. In implementing the ESA, we pre-smooth the energy signals \( \Psi_d[x(n)] \) and \( \Psi_d[p(n)] \) with a 7-point binomial smoothing filter, because this can reduce the approximation errors by about 50% [9].

III. AN ITERATIVE METHOD FOR DETERMINING FORMANT CENTER FREQUENCIES AND MODULATIONS

We next describe an iterative method that reduces the importance of having good initial estimates of \( f_c \) when applying the ESA, and allows us to follow formants within a voiced speech segment.

From the results of our early experiments with the ESA, we noticed that when the center frequency of the filter was off by even several hundred hertz, the average value of the instantaneous frequency was often close to the formant peak frequency. Fig. 1 shows an example of this.

Fig. 2. The iterative ESA was started at \( f_c^{(1)} = 1700 \) Hz for this speech segment. Only one iteration was necessary for it to converge to \( F_2 = 1563 \) Hz, a difference of about 140 Hz.

Based on this observation and a suggestion by Kaiser\(^1\), we reasoned that we might be able to use \( f(n) \) to iteratively estimate the center frequency of the formant, adjusting the center frequency of the filter on each iteration. Assuming in the AM-FM model for a speech resonance (1) that the frequency modulating signal \( q(t) \) has a zero mean within the short-time speech analysis frame, an estimate for the formant center frequency \( f_c \) can be the average of \( f(t) \). Thus, we have implemented the idea of iterative estimation by using the

\(^1\) J. F. Kaiser initially suggested the iteration in the ESA, 1991.
rule: \[ f_{c}^{(j+1)} = \sum_{n=0}^{N-1} f^{(j)}(n) \]. That is, the center frequency of the Gabor bandpass filter on the \((j+1)\)th iteration is set equal to the average value of \(f(n)\) on the \(j\)th iteration. We start the algorithm by setting \(f_{c}^{(1)}\) to be some initial estimate of the formant, and we consider the algorithm to have converged when the center frequency does not change by more than 5 Hz. This iterative application of the Gabor bandpass filter, the ESA, and the updating of the filter center frequency, while keeping \(\alpha\) fixed, is henceforth called the iterative ESA. Note that during the iteration one need not apply the amplitude estimation part of the ESA, except at the last iteration.

We have been using the iterative ESA for some time now and, overall, the results are good. For well-defined spectral peaks and fairly good initial estimates, the algorithm converges quickly. Fig. 2 shows an example where the initial estimate was off by 140 Hz and only one iteration was required for convergence. A poorer initial estimate or a poorly defined peak requires more iterations.

In our experience, the algorithm converges to the spectral peak closest to the initial estimate as long as the Gabor bandpass filter is narrow enough to exclude neighboring formants. Problems can arise for the first formant, when larger ESA errors due to a low \(F1\) center frequency may be compounded by undesired effects of a nearby \(F2\). Our solution has been to use narrower filters for formant estimates below 1000 Hz, as described in Section IV. Another problem may occur with back vowels, where the first two formant peaks may not be well-separated and appear as one peak rather than two. As with many standard formant finders, the iterative ESA could fail to find two separate peaks.

In Fig. 3 we superimpose the results of the iterative ESA onto the LPC spectrum for two vowels. In Fig. 3(a), the results of iteration agree well with the peaks in both the speech and LPC spectrum. However, in Fig. 3(b), the LPC spectrum has some peaks that are difficult to distinguish, while the iteration results correspond well with peaks in the speech spectrum. In addition, the iterative ESA has the advantage over LPC that it also finds the modulations, i.e., the signals \(x(n)\) and \(f(n)\).

Fig. 4 shows an example of how better positioning of the Gabor filter due to iteration improves the output of \(\Psi_{t}\), and thus improves the estimates of \(|\omega(n)|\) and \(f(n)\). In Fig. 4(a), we show the speech spectrum with two Gabor filters superimposed. One filter is centered on \(F2\) and the other filter is off by several hundred hertz. Fig. 4(b) shows the waveform and Fig. 4(c)-(d) demonstrates a clear difference between \(\sqrt{\Psi_{t}}\) for the two filters. Here, better positioning of the filter results in additional modulations being revealed, but there may be cases where it results in fewer modulations.

Since the iterative ESA seems to gravitate to peaks, one might assume that it would be feasible to use the algorithm for finding formants, where by "finding formants" we mean that there is no prior information about where the formants might be located. That is, we blindly start the algorithm with arbitrary values and it converges to a formant. Note that this is not the same as what we described above, which was to start the algorithm by giving it a push in the direction where we thought a formant was located, and then letting it track or refine the formant value. Extensive experience with the iterative ESA has led us to conclude that it cannot be useful as a formant finder when it has no prior information about the formant center frequencies. In many cases where \(f_{c}^{(1)}\) was completely arbitrary, the algorithm converged to a spectral peak if there was no strong peak within about 500 Hz of the initial estimate. Instead, the algorithm seems to be most useful when used in conjunction with a more standard formant finder. We will describe this in Section IV.

We now briefly turn to the issue of how the iterative ESA may be converging. We experimentally observed that the resulting average value of the instantaneous frequency \(f(n)\) seemed to be drawn close to peaks or local maxima in the power spectrum. Since the output of \(\Psi_{d}[x(n)]\) is proportional to the energy required to produce \(x(n)\), we reasoned that the algorithm could be maximizing the mean (modulation) energy of the bandpass filtered speech. We defined the average energy for an iteration as \(E(j) = \frac{1}{N} \sum_{n=0}^{N-1} \Psi_{d}[x^{(j)}(n)]\), where \(x^{(j)}(n)\) is the bandpass filtered speech on the \(j\)th iteration.

This quantity was then computed on each iteration while finding the formants of speech segments. We have experimentally found that for the majority of formants \(E(j)\) increased as the algorithm converged. However, there were a few exceptions where it would rise and then either oscillate a little, or fall slightly. We have also found that \(E(j)\) peaks in the vicinity of formant peaks. Perhaps locally searching for the peaks of the average energy \(E\) (as a function of \(f_{c}\)) may be useful as a convergence criterion. We are continuing to investigate this issue.

IV. AUTOMATED SYSTEM

We now describe an automated system that we have been developing to automatically find the formant center frequencies and modulations of a speech segment. In this system, the iterative ESA described in the previous section is employed to determine the center
frequencies of the resonances. Bandpass filtering is implemented using a truncated, discretized Gabor filter with impulse response

$$g(n) = \begin{cases} \exp(-\alpha n^2) \cdot \cos(2\pi f_c T_n), & |n| \leq N, \\ 0, & |n| > N \end{cases}$$

(9)

$N$ is chosen such that the envelope of $g(n)$ is nearly zero at $n = N$. We have found that a good choice is $N$ such that $\exp(-\alpha N^2) \approx 10^{-6}$. Through extensive experience, we have found that it is reasonable to use fixed bandwidths of $\alpha = 800$ Hz when $f_c < 1000$ Hz, and $\alpha = 1100$ Hz for all other resonances. We discuss the possibility of varying bandwidth values in the Conclusion.

An important issue for the system is getting good initial estimates of the formant center frequencies. This is to ensure that the iterative ESA does not converge to false formants. We now briefly discuss the two methods that we have implemented: a standard method, LPC, and a new method that we call morphological peak picking.

For LPC, the advantages are that is easy to implement and often does a good job of estimating the spectral peaks. However, it sometimes performs poorly, especially for female speakers or children. Formant center frequencies can be estimated by finding the roots of the LPC polynomial [1], or by peak picking the spectral envelope [8]. The former method is computationally expensive, while the latter may be problematical when the peaks are not strong. We have implemented the latter method.

Our other formant finding method is to perform a morphological closing of the speech spectrum. A closing of a signal by a set $B$ is a nonlinear filter that is a cascade of a dilation (local maximum within the moving window $B$) followed by an erosion (local moving minimum) [7], [12]. Formally, let $S(k)$ be the speech magnitude spectrum as a function of discrete frequency index $k$ and let $B = \{-\frac{W-1}{2}, \ldots, \frac{W-1}{2}\}$ be the set of indices, with $W$ being the window width. The closing of $S(k)$ by $B$ is defined as the transformed spectrum

$$S \ast B(k) = \min_{j \in B} \max_{i \in B} S(k-i+j).$$

(10)

Figs. 5(b)-(d) show examples of dilation, erosion, and closing of the speech spectrum in Fig. 5(a). In the closing, the narrow valleys of the spectrum get filled up, so to speak, and the peaks of the closing correspond to peaks of the spectral envelope. Then it only remains to pick the peaks and we have our initial formant estimates.

Extraneous peaks can be produced by this method, but if the iterative ESA is started at one of these peaks, it almost always converges onto an actual formant. The exceptions tend to be tiny peaks occurring on the spectral plains that are often found at frequencies above 1000 Hz.

A requirement of this method is to carefully choose the width of the filter, $W$. If the filter is too wide, a formant that is close to another stronger formant might be missed. At the same time, we must keep it from being too narrow to avoid treating individual harmonics as peaks. Due to the latter restriction, the minimum width of the filter is essentially a function of the fundamental frequency, $F_0$. We estimate $F_0$ by peak picking the spectrum over the first 1000 Hz, which, ideally, gives the location of the first five to ten harmonics. The distance between these points is averaged for an estimate of $F_0$. To avoid getting estimates that are too low due to spikes that might occur between harmonics, we set a lower limit of 250 Hz on the width of the filter. This lower limit is based on the fact that the lowest fundamental frequency that we expect to encounter is, on average, 100 Hz, and we would like the filter to overlap about three harmonics.

Advantages of morphological filtering are that it is very cheap to implement and can be used rigorously to extract peak or valley features on arbitrary signals. In addition, it is nonparametric, i.e., it does not presuppose anything about the speech spectrum, while
LPC assumes that the vocal tract transfer function can be modeled by an all-pole model. Finally, it formalizes what we most likely do when we visually identify formants from a speech spectrum using geometrical features.

Back vowels present a problem for many formant finding techniques because if $F1$ and $F2$ are not well-separated they could be treated as one spectral peak. For morphological peak picking, the problem could be compounded by high-pitched speech, when $F2 - F1$ may be less than the filter width, $W$. In such a case, the closing will have a peak at either $F1$ or $F2$, but not at both. This will clearly not do. A possible solution is to perform the closing on the spectrum with a logarithmic frequency axis and a narrower filter. Not only might $F1$ and $F2$ be better separated, but spurious peaks at high frequencies could be avoided. Alternatively, the width of the window could be increased exponentially as a function of frequency.

For the latter alternative, the window $B$ in (10) is replaced with a frequency-varying window $B_k$, where $B_k = 10, 1.2, \ldots, W(k)$, $W(k) = [W(0) \exp(kC)]$, and $C$ is a constant that controls the rate of exponential growth. We have tested the latter method and the initial results are promising, as seen in Fig. 6. Fig. 6(a)–(b) show spectra for
Fig. 7. Morphological filtering of a spectrum of the back vowel /a/. Dotted vertical lines indicate formant center frequencies found by the IESA. (a) For normal style speech, a constant window of length $W = 266$ Hz separates the first two formants. (b) For angry style speech, $F_0$ is high and $W = 468$ Hz is used for the closing. However, $F_2 - F_1 < W$, so they are treated as a single peak. (c) Using a narrower window of $W = 281$ Hz separates the first two formants of the angry style speech, but results in extraneous peaks at high frequencies. (d) $W$ is varied exponentially from 250 Hz to 750 Hz. $F_1$ and $F_2$ are separated, while spurious peaks at high frequencies are smoothed.

Fig. 8. Results of the automated system. (a) The vowel /i/, from a male speaker. (b) The spectrum. Dotted vertical lines indicate formant center frequencies following iteration.

the vowel /a/, spoken by a male speaker in normal and angry speech styles. Closings computed using constant structuring elements are superimposed on both spectra. For the normal speech style, $F_1$ and $F_2$ are separated in the closing, but for the angry style, $F_0$ is high.
so that $W$ is quite wide, and $F1$ and $F2$ are not separated. Fig. 6(c) shows the closing of the angry style spectrum, using a narrower, but still constant window length. $F1$ and $F2$ are separated, but many extraneous peaks are produced, particularly around 6000 Hz, where there does not seem to be a spectral peak. Finally, Fig. 6(d) shows the closing of the angry style spectrum using structuring elements that increase exponentially with frequency, with $W(0)$ and $C$ chosen so that the initial width of the window is about 250 Hz and the width at the high end of the spectrum is 750 Hz. Not only are the first two formants separated, but use of the wider window at higher frequencies has removed the extraneous peaks. These results are promising and suggest that further testing of this method is warranted for possible implementation in our system.

For the time being, we have implemented morphological peak picking with constant size window, and we give the user of our system the choice of that or LPC. Fig. 7 shows an example of the initial values that result for both methods. The results are similar, except that for the peaks near 3800 Hz, LPC finds one formant while morphological filtering finds two. Overall, we have found that LPC and morphological peak picking give similar results after iteration, so it may be that morphological filtering is the best choice, since LPC is more expensive. However, more rigorous testing must be done to determine how their performance compares before iteration. If morphological filtering requires many more iterations due to extraneous peaks, then it may be no less expensive than LPC.

Finally, an example of the output of the automated system is shown in Figs. 8–10. For this speech segment, the initial estimates of formant center frequency were found using morphological peak picking. Fig. 8(a) shows the speech waveform, while Fig. 8(b) shows the speech spectrum, with the formant center frequencies found by the iterative ESA indicated by vertical dashed lines. Figs. 9–10 show $|a(n)|$ and $f(n)$ for the first four formants. For this example, there are more modulations present in the higher formants than in the lower formants. Median filtering has been applied to the extracted instantaneous frequency signals to suppress the narrow spikes that are due to pitch period effects or isolated numerical instabilities of the ESA [5].

V. CONCLUSION

An automated system that finds formant center frequencies and speech modulations using energy separation has been described. Initial values for the formant center frequencies are found using LPC or morphological peak picking. These values are then refined using an iterative energy separation algorithm. The system has been shown to be effective.

We are continuing to improve the system by investigating the following refinements. First, we would like to implement the automatic selection of the filter bandwidths, based on the distance between neighboring formants. We have done some preliminary work on this, using synthetic AM-FM signals, and the results suggest that the optimum choice of bandwidth could be a linear function of the
distance between formants, $\Delta f$. Our next step is to try to apply
this result to speech. Complications may occur, however, when we
try to incorporate this with the iterative ESA, because $\Delta f$ will
change during iteration. The solution that we propose is to apply
the iterative ESA to all formants in parallel. Then we can vary
the filter bandwidths on each iteration according to the values of
$f_i$.

The other refinement is to possibly reduce discretization effects as
follows. Instead of convolving the speech signal $s$ with a discrete-time
Gabor bandpass filter $g$ and then applying the discrete-time energy
operator, we apply the following combination of the continuous-time
energy operator and bandpass filtering, which introduces discretiza-
tion only at the very last step, i.e., at sampling time $t = nT$: $\Psi[s(t) * g(t)]_{n \rightarrow \infty} \approx \left[(s(t) * g(t))^2 - \left(s(t) * g(t) \right) \left(s(t) * \hat{g}(t)\right)\right]_{t=nT}$, where $\hat{g}(t)$ is $g(t)$'s derivative. The derivatives of the Gabor bandpass filter,
are functions with simple known formulas. In this way, we avoid
the approximation of the signal derivatives with first differences
that maps $\Psi$ to $\hat{\Psi}_f$ [4], [6], and this may improve the results of applying
the energy operator and the ESA to sampled speech signals.

ACKNOWLEDGMENT

The authors wish to thank the reviewers for their very useful
suggestions. Thanks are also due to J. Kaiser of Rutgers University
and T. Quatieri of M.I.T. Lincoln Laboratory for many helpful
discussions.

REFERENCES

linear prediction of the speech wave,” J. Acoust. Soc. Am., vol. 50,

381–384.

to continuous signals,” in Proc. 4th IEEE Digital Signal Processing

frequency demodulation using energy operators,” IEEE Trans. Signal

signal modulations with application to speech analysis,” IEEE Trans.

modulations, and energy operators,” in Proc. IEEE ICASSP (Toronto,

set-theoretic analysis and relations to linear shift-invariant filters,”


and Hilbert transform approaches to signal and speech demodula-
tion,” to be published.


mechanisms in the vocal tract,” in Proc. NATO Advanced Study Institute
on Speech Production and Speech Modeling (Donaus, France), July 1989.

II. LATTICE ANALYSIS USING QUANTIZED PARCOR COEFFICIENTS

A lattice analysis filter for linear prediction of speech is depicted
in Fig. 1. During the analysis, the speech signal is fed to the input
of the lattice, and the prediction error is minimized with respect to each
PARCOR coefficient [7]. Let $f_m(n)$ and $g_m(n)$ be, respectively,
the forward and backward prediction errors of the $m$th stage lattice
filter at time $n$. Then, from the lattice structure shown in Fig. 1, two
order recursive equations are derived as

$$f_{m+1}(n) = f_m(n) - h_m g(n - 1)$$

and

$$g_{m+1}(n) = g_m(n) - h_{m+1} f_m(n).$$

Split-Dimension Vector Quantization of Parcor
Coefficients for Low Bit Rate Speech Coding

Kwok-Wah Law and Cheung-Fat Chan

Abstract—A novel split vector quantization (SVQ) scheme for low
bit rate coding of speech signals is proposed. In this scheme, the LPC
parameter vector, which is represented by PARCOR coefficients, is
split into small-dimension subvectors, and each subvector is sequentially
quantized according to a multistage structure that resembles a segmented
lattice filter. The forward and backward prediction residuals in the
segmented filter are coupled across VQ stages. The quantizer in each
stage operates on the principle of minimizing the forward and backward
prediction error energies similar to linear predictive analysis. Simulation
results show that the new split VQ scheme can achieve transparent
quantization of LPC parameters at 25 b/frame.

I. INTRODUCTION

Linear predictive coding (LPC) is a well-established technique for
speech compression at low rates. In order to achieve transparent
quantization of LPC parameters, typically 30 to 40 b are required
in scalar quantization [8]. Vector quantization can reduce the bit rate
to 10 b/frame [4], but vector coding of LPC parameters at such a
bit rate introduces large spectral distortion that is unacceptable for
high-quality speech communications. In the past decade, structurally
constrained VQ's such as multistage (residual) VQ and partitioned
(split) VQ [1], [6] have been proposed to fill the gap in bit rates
between scalar and vector quantization. In multistage schemes, VQ
stages are connected in cascade such that each of them operates on
the residual of the previous stage [3]. In split vector schemes, the
input vector is split into two or more subvectors, and each subvector
is quantized independently. Recently, Palival and Atal proposed a
split vector scheme to achieve transparent quantization of line
spectrum frequency (LSF) parameters using only 24 b/frame [8].
In this correspondence, we propose a novel scheme to decompose LPC
parameters represented by PARCOR coefficients into subvectors for
split vector quantization. The structure of the proposed scheme re-
sembles a segmented lattice filter where PARCOR coefficients of each
segment are grouped as vectors for quantization. All lattice segments
are connected in a cascaded structure similar to the multistage VQ.