Hierarchical approach to formant detection and tracking through instantaneous frequency estimation

S. Ghaemmaghami, M. Deriche and B. Boashash

Indexing terms: Frequency estimation, Speech

Formant frequencies, represented by major peaks in the spectrum of speech signals, convey important information about speech. The authors propose a method for detecting the formants of voiced speech through 'instantaneous frequency' (IF) estimation using a recursive least square (RLS) algorithm. The accuracy of the technique is assessed by comparing it with conventional formant detection techniques. This method is also analysed from the viewpoint of phonetic conformity using 'temporal decomposition'.

Introduction: Formants are typically detected through searching for major peaks in spectral representations using short-time Fourier transform, filter-bank analysis, or linear prediction [1]. The main drawback of the first two (non-parametric approaches) is the tradeoff between frequency resolution (formant detection capability) and time resolution (formant tracking accuracy). In the linear prediction method (model based), two typical techniques are used: root finding, and peak picking of the reciprocal of the inverse filter in the LPC model. Both procedures rely on an all-pole model approximation of speech and the estimation accuracy is typically < 90% [1].

Instantaneous frequency (IF) estimation techniques are proven to be efficient in detecting and tracking frequency changes of mono-component signals but, in the case of multi-component signals, the result can be meaningless [2].

To detect and track individual components of multi-component signals through IF estimation, a number of methods have been proposed [2]. Most of these methods give high instantaneous resolution in both time and frequency domains but, on the other hand, tracking spurious peaks in speech spectrum can be a major drawback [2].

To alleviate such a problem, we use the recursive least square (RLS) algorithm which models the data as a linear prediction sequence, through the weighted covariance matrix [2]. The algorithm parameters are extracted in a recursive way, as [2]

$$\underline{a}_{n+1} = \underline{a}_n - e_{n+1}\mathbf{P}_n^{-1}\underline{z}_n^* \qquad (1)$$

$$e_{n+1} = z(n+1) + \underline{z}_n^T\underline{a}_n$$

$$\mathbf{P}_n = \alpha\mathbf{P}_{n-1} + \underline{z}_n^*\underline{z}_n^T$$

where n is the time index, P_n is an approximation to the covariance matrix, a_n is the vector of prediction filter coefficients at time n, e_{n+1} is the prediction error at time n+1, and α is the forgetting factor. z_n is the signal vector at time n given as

$$\underline{z}_n = [z(n)\ z(n-1)...z(n-L+1)]^T$$

where L is the length of the prediction filter and T represents transposition.

The RLS algorithm estimates only the frequency of the predominant component in speech, which is indeed the major formant. To find all desired formants, a hierarchical procedure is proposed in this Letter. We discuss the accuracy of the method and also present an evaluation by phonetic relevance analysis via temporal decomposition (TD) [4] of the matrix of estimated formants.

Method: First, we de-emphasise voiced speech using a linear phase 6dB/oct lowpass FIR filter to attenuate the high-frequency components. This filter reduces fluctuations in the IF which may appear during some instants due to the 'non-lowpass' spectral characteristics. Secondly, the RLS is applied to the de-emphasised speech to estimate the first formant as the IF of the predominant component.

To find other formants, we need to modify the spectral characteristics of the signal so that the predominant component be the desired formant detectable by the RLS algorithm. To do this, we apply the algorithm to the speech signal in consecutive steps, each associated with an appropriate pre-processing stage, to change the frequency predominance from one formant to the next. Accordingly, in each step, one formant is detected and hence tracked.

To remove the spectral components associated with the estimated formant and to continue the procedure, an adaptive time-varying filter is needed. This is performed using a sharp variable cutoff frequency highpass filter, adaptively. The cutoff frequency of the filter is set the basis of last estimated formant and the average formant bandwidths (50, 80, 120, 200, and 200Hz for the first five formants, respectively [1]).

Table 1: Overall error in formant detection and tracking [Hz]

Formant	F1	F2	F3	F4	F5
RLS-based	48	51	92	440	680
Convemtional	60	60	110		

Experimental results: The results obtained using the proposed formant detection method are shown in Table 1, using a number of voiced utterances spoken by different speakers.

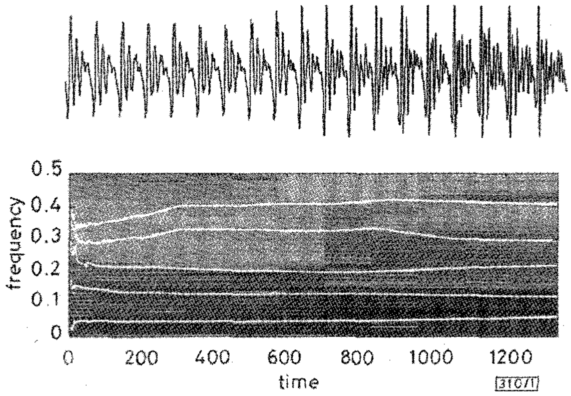

Fig. 1 *Speech waveform, spectrogram, and formant trajectories (white lines on spectrogram)*

The results represented in Fig. 1 are obtained by processing a composite voiced speech signal composed of /o/, /r/, and /a/. It has been lowpass filtered at ~4kHz and sampled with 8kHz. The total duration is 200ms, which can approximately be divided into three parts of 87, 48, and 65ms duration for the three aforementioned sounds, respectively.

Table 1 shows the overall error in formant estimation using the proposed method. In comparison with conventional formant trackers, in which the overall error is 60Hz for the first two formants, and 110Hz for the third one [1], our method yields higher accuracy. The relatively larger error in the fourth and fifth formants, arises from the non-stationary behaviour of voiced speech at higher frequencies.

As seen in Fig. 1, close formants are well resolved by the proposed method. Indeed, the RLS algorithm uses the preceding information to predict the new parameters of the signal, in the least square sense. This is controlled by a forgetting factor which is typically between 0 and 1. The larger the forgetting factor, the smoother the formant trajectories. We found that a value between 0.9 and 0.95 is suitable to obtain good speech quality.

To perform an evaluation from the viewpoint of phonetic conformity, we used the matrix of formants as the matrix of spectral parameters. Then, we decomposed the matrix using temporal decomposition [4] and extracted speech events. The experiments, conducted using a number of different utterances, resulted in good relevance of the detected events to the phonemic structure of speech (considering the events location and duration). We also repeated the experiments using the STFT technique in IF estimation for a comparison from the viewpoint of phonetic relevance. For most cases, the latter technique produced spurious or unmatched events.

An illustrative example, using the same short composite utterance (depicted in Fig. 1), is shown in Fig. 2. As can be seen, events extracted using the proposed method (solid) adapt well with the phonemic structure of the utterance, while nearly irrelevant events (dashed) are obtained using the STFT technique.

Conclusion: We have proposed a formant detection method using an RLS algorithm to find and track the IF of voiced speech signals. The method relies on the predictability of formant evolution and the structure of voiced speech spectrum using the weighted covariance matrix associated with short segments of the signal.

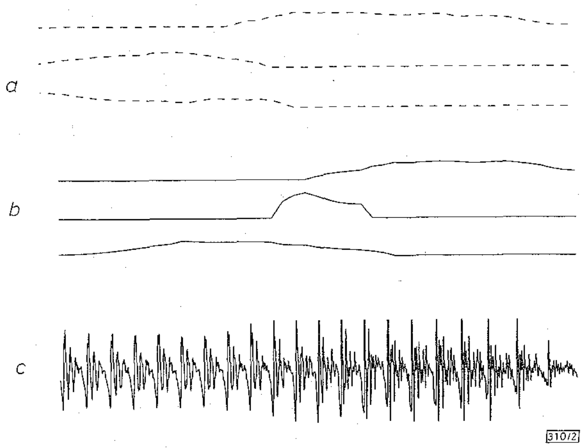

Fig. 2 *Temporal decomposition using formant-matrix as spectral parameters*

a Events using STFT
b Events using proposed method
c Speech waveform

The method has been evaluated through comparison of the achieved accuracy in formant estimation with that of conventional formant detection techniques. The results obtained using the proposed technique clearly indicate better detection and tracking ability (see Table 1 and Fig. 1).

The proposed technique has also been analysed and compared to the STFT-based methods from the viewpoint of phonetic conformity using temporal decomposition. Our experiments show that while STFT-based methods generally fail to follow the speech phonetic evolution, the proposed technique mostly conforms with the phonemic structure (see Fig. 2). This result, axiomatically, shows the superior performance of the proposed formant estimator over STFT-based methods in extracting the main spectral components of speech which are related to phonetic events. These components are particularly useful in systems where only a few formants are considered for representation.

© IEE 1997 *5 November 1996*
Electronics Letters Online No: 19970063

S. Ghaemmaghami, M. Deriche and B. Boashash (*Signal Processing Research Centre, School of Electrical and Electronic Systems Engineering, Queensland University of Technology, GPO Box 2434, Brisbane, Q 4001, Australia*)

E-mail: m.deriche@qut.edu.au

References

1 DELLER, J.R. Jr., PROAKIS, J.G., and HANSEN, H.L.: 'Discrete-time processing of speech signals' (MacMillan Publishing Co., 1993)

2 BOASHASH, B.: 'Estimation and interpreting the instantaneous frequency of a signal – Parts 1 & 2', *Proc. IEEE*, 1992, **80**, (4), pp. 520–568

3 ASSALEH, K.T., and MAMMONE, R.J.: 'Spectral-temporal decomposition of multicomponent signals'. Proc. ICASSP 93, 1993, pp. 206–209

4 ATAL, B.S.: 'Efficient coding of LPC parameters by temporal decomposition'. Proc. ICASSP 83, 1983, **1**, pp. 81–84

Low cost wideband I-Q vector modulator

J.M. Blas and J.I. Alonso

Indexing terms: Quadrature phase shift keying, Modulation

One of the applications of QPSK modulators is to control the amplitude and phase of an RF signal, working as a vector modulator. Recently there are several monolithic commercial circuits that can be easily used to frequencies of > 4GHz. First, the input signal must be decomposed into two quadrature components, which is usually done in a 3dB 90° hybrid coupler. Using microstrip techniques, this I-Q vector modulator exhibits narrowband characteristics, due to the degradation of the hybrid performances with frequency. The authors show how the working bandwidth can be extended up to an octave by means of a versatile control system as described here, as an alternative to a new design of the RF stage.

Introduction: A vector modulator has recently been developed as the main element of an electronically steered array antenna for mobile communications applications in the L band [1]. A classical configuration has been used [2], with a 3dB 90° hybrid coupler to obtain two quadrature outputs, and a commercial MMIC QPSK modulator (HPMX2001) to make the control vector function. By using microstrip techniques, small size circuits with high integration facilities are easily obtained at a very low cost. The vector modulator has been initially designed at a frequency of 1.575GHz, used by the GPS navigation system, although the working bandwidth will be mainly limited by hybrid characteristics. It can be easily extended up to the whole L band (1–2GHz), by introducing the appropriate compensation through the control subsystem without any modification in the RF stage, as is presented here. To provide complete control by software, PC based boards have also been developed, so the device could also be used in any of the navigation and communications systems working in the L band.

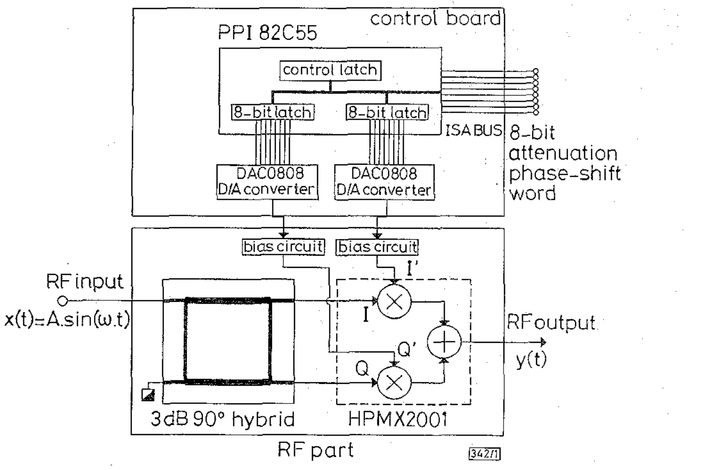

Fig. 1 *Block diagram of I-Q vector modulator*

Design: The block diagram of the vector modulator is shown in Fig. 1. It is composed of a 3dB 90° hybrid, into which the RF signal is applied to obtain the two equal amplitude quadrature outputs needed, I and Q. These signals are introduced in the QPSK modulator, in which the control vector function is calculated by multiplying the amplitude of each component by two DC control signals (I' and Q'), acting as a scaling factor between –1 and +1. The output is now obtained as a vectorial combination of them, resulting in a signal with equal frequency and different amplitude and phase characteristics. The working frequency range of this circuit in the RF inputs is from DC to 2GHz. These DC control signals are obtained through 8 bit D/A converters, with 8 bit latches and an ISA interface in a PC board, constituting a system completely controlled by software, which is one of the key elements of this design. To calculate these control signals properly the following effects must be taken into account. At the design frequency of the hybrid (1.575GHz), the imbalance in the amplitude and phase will not be null due to small imperfections in the circuit, although they will have the minimum values. This fact has little influence