THE USE OF LPC AND FFT IN PHONETIC ANALYSIS

J. Rosenhouse and G. Rosenhouse
Department of General Studies
Faculty of Civil Engineering

The Technion, I.I.T. Haifa, Israel

ABSTRACT
The application of FFT and LPC methods in speech analysis is discussed here. When used side by side,these methods are complementary, which helps clarify various points that may remain if only one method is used.This approach is exemplified here for some Hebrew speech sounds (vowels, consonants) and some general speech features.

1. INTRODUCTION: FFT AND LPC

The Fast Fourier Transform (FFT) yields frequency spectra for given signals of a certain duration.This method is used in speech analysis to represent the speech output in the frequency domain for the given duration, and is a result of both the input signal and the filter (the vocal tract). On the other hand, the Linear Predictive Coding(LPC) method yields the response function of the vocal tract, eliminating as much as possible the effect of the input signal. This method is based on the approximation of the speech signal by a linear combination of past speech samples. Minimizing the sum of square differences over a finite interval, between the actual speech samples and the linearly predicted samples, a unique set of prediction coefficients is defined.

The advantage of the LPC is its accuracy and reliability in defining the basic speech parameters, mainly formants and spectra. It is analytically tractable, easy to implement and suitable for time varying speech signal analysis.

In ordr to define precisely the spectrum of vowels and consonants by the FFT method first a pitch detector is to be used. This enables the selection of signal duration that corresponds to integer numbers of pitch periods. Using inaccurate durations of input signals may yield errors. Another deficiency of the FFT compared to LPC is the large number of terms to be calculated by it, while only a small number of poles is required for the LPC values. Yet the FFT method can help in defining the number of poles to be used in an LPC program.

The LPC method is generally considered a more efficient method.Yet much experimental evidence from our work(using samples of 35ms duration)has shown a good correspondence between FFT and LPC in formant frequency definition (see Figures 1,2), though large differences have been found for amplitudes due to the input effect.In addition the FFT method yields the Fo and certain effects that are difficult to identify by LPC only. Hence, the use of both LPC and FFT in speech analysis to complement each other seems to be favorable.

In the literature comparison of FFT, LPC or other methods is to be found (see, e.g., |3|,|6|). Woods (|6|), for example, compares the spectrograph output with the LPC method. It should be noted, that FFT can be used so that the results would be easier to read. (This can be found in e.g. improved spectrographs, and the MATLAB(c) package which applies the FFT function may be used by users to write programs according to needs).

In the sequel, we note some examples based on our experiments performed at the Lab of Medical Electronics, the Faculty of Electrical Engineering at the Technion. Some experiments were also done in the framework of a Technion D.Sc. thesis (|1|). Our recorded natural speech material was analyzed by programs written at the Technion, by both LPC and FFT methods.

2. EXAMPLES

2.1. Pitch Detection

The LPC method is normally not intended for pitch analysis(Fo).As the FFT program gives "raw" harmonics (without "smoothing") of both source and filter, it is easy to find the Fo and the other formant values and distinguish between them visually. Applied natural speech analysis(for linguistic or even medical purposes of voice quality measurements) often needs to define or find the speaker's pitch (i.e., Fo) as well as the formants, and in such a case combining both these methods for the analysis seems important. (See Figure 1.)

2.2. Formant Frequencies, Band-Width Variations

It is well-known that speech is not stationary.Therefore, no speech segment is the same as any other segment, even if they are adjacent. There is thus always also a "movement" of formant bandwidths along the frequency axis of the spectrum. It is hard to decide just by an FFT program output what the really important formant areas are (besides formant peaks), as the final output of a speech signal analyzed by an FFT analysis program is a series of harmonics along the spectrum which are effected by the input signal In this case, then, the LPC program may be more suitable because it presents formants including the full bandwidth covered by each formant. Thus, even if there are some local peaks within this formant band area,it can become clear that they are not individual formants but part of a specific frequency domain. This presentation is advantageous over spectrographic outputs, where formant limits are not clear and formant centers (their peaks) are not accurate. An LPC program may also provide the precisely calculated point of a formant peak frequency.

2.3. Amplitude Features

As mentioned,the FFT program calculates signals including both their source and filter while the LPC program calculates only filter features, namely the formants. Thus, formant amplitudes are more accurate in the LPC program, although formant amplitudes of an FFT program output seem to be more conspicuous, due to some energy gain values of the voice source. As a matter of fact, for the hearing system the whole formant range is important rather than a single peak-frequency, which even more justifies the use of LPC for speech sound analysis.

2.4. Sex-Dependent Phonetic Features of Native Speakers of Hebrew

It is likewise well-known that for the same phonemes there are different formant values, which depend on the speaker's sex(and the physical structure of the vocal tract). Such differences may occur also in F2 and F3 (which in usual spectrograms are hard to see) and in relative amplitudes of each formant. Sex-related differences were found, for instance, in the pronunciation of the vowels /o,u,a/ by some speakers (|2|) and /ħ,x/ (|4|).

2.5. Fricative Features

In Hebrew as in many languages, there are fricative phonemes.Some are more common than others, as in other languages (e.g., /f, s, sh/), and some are less common (the laryngeal and velar /h,ḥ,x/). These sounds are hard to analyze accurately because of the large amount of noise involved in their articluation and the lack of voicing, and traditional sound spectrograms yield rather blurred printouts of such phonemes. In this case, then, the FFT analysis seems again to be of less value than the LPC program which yields well defined formant domains (|4|, |5|).

3. CONCLUSIONS

Many speech analysis techniques exist now, relying on various theoretical approaches and algorithms. It seems useful to find the merit of each method in order to extract the best results of all of them in order to fully understand speech structure. Combining various analysis methods, more insight may be gained as to many problems that still exist in this field. The few examples shown here represent clearly this viewpoint concerning language-specific and general(universal) phonetic issues.

4. REFERENCES

(1)Aronson,L.(1990) Electrical Parameters of Cochlear Prosthesis Adapted to Deaf Hebrew Speakers,D.Sc.thesis, School of Medicine, Technion (supervisers: Prof. G.Rosenhouse and Prof. L. Podoshin)

(2)Aronson,L., J.Rosenhouse, G.Rosenhouse and L.Podoshin, "An analysis of Modern Hebrew Vowels and Voiced consonants" (submitted for publication)

(3)Iivonnen,A. (1987) "A Set of German Stressed Monophthongs Analyzed by RTA, FFT and LPC," in: Channon,R. and L. Shockey (eds.) In Honour of Ilse Lehiste, Dordrecht, Holland/Providence, U.S.A., pp. 125-138.

(4) Rosenhouse, J. (1989) "Issues in Hebrew Phonetics: the Articulation of Ḥ/X and Tsere",Proceedings of the 10-th World Congress of Jewish Studies, Division S, vol. I, pp. 125-132 (in Hebrew)

(5) Rosenhouse, J. "Two Unstable Phonemes in Israeli Hebrew and Colloquial Arabic :'Aleph and 'Ayin"(to appear in the volume In Honour of Prof. W. Leslau, 1991).

(6) Woods, S. (1987) "The Precision of Formant Frequency Measurement from Spectrograms and by Linear Prediction", Abstracts of Papers, The 3rd Swedish Phonetics Symposium,Fonetik-89, in: Speech Transmission Lab. Quarterly Progress & Status Report,April 1989,Stockholm, R.I.T. pp. 91-93.

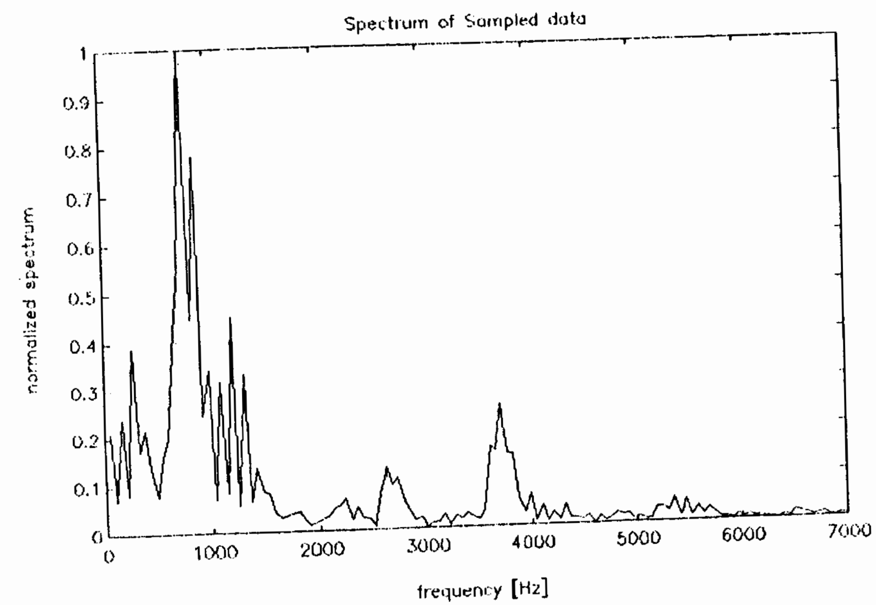

Figure 1. FFT analysis of /ʿ/ as uttered by a male native speaker of Hebrew (cf. |5|)

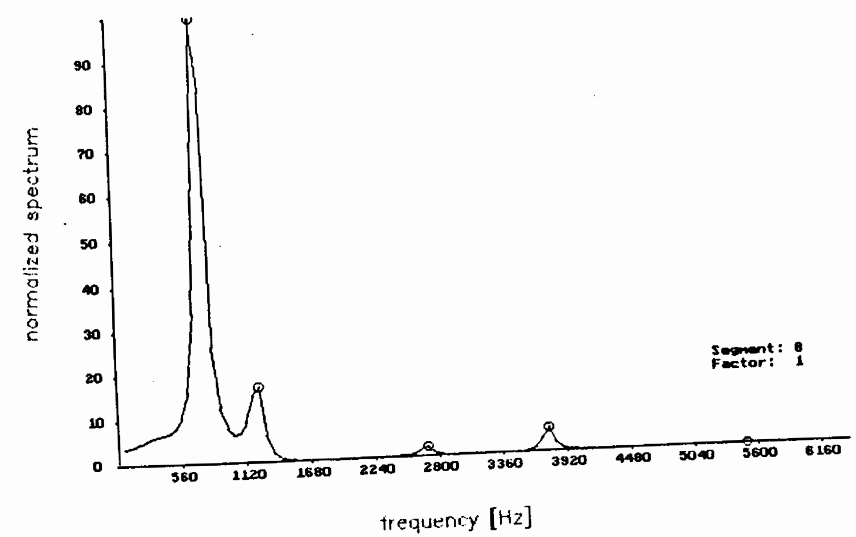

Figure 2: LPC analysis of /ʿ/ (same segment as Figure 1) as uttered by a male native speaker of Hebrew (cf. |5|)