Annales UMCS Informatica AI X, 2 (2010) 7-12
DOI: 10.2478/v10065-010-0047-3

Annales UMCS
Informatica
Lublin-Polonia
Sectio AI
http://www.annales.umcs.lublin.pl/

Formant paths tracking using Linear Prediction based methods

Ireneusz Codello*, Wiesława Kuniszyk-Jóźkowiak

Institute of Computer Science, Maria Curie-Skłodowska University,
Plac M.Curie-Skłodowskiej 1, Lublin, Poland.

Abstract – This paper focuses on formants as basic parameters for vowels recognition. There are used two different algorithms for formants finding based on the LP algorithm: spectral peak picking and root extraction algorithm - obtaining very good path estimations by each algorithm. Those methods are compared in a graphical form in our application 'WaveBlaster'.

1 Introduction

A sound is nothing but a complicated wave varying in time which is converted by our ears and brain causing a sound perception. Such a human speech sound wave can be divided into homogeneous chunks - phonemes, which are indivisible components of utterance (like letters in writing). Each phoneme has its own characteristic wave - that is why we can distinguish one phoneme from another. If we could implement a method which recognizes phonems, then we could, for instance , recognize speech - we only have to find a good method for wave decomposition.

The most popular method is to find all frequencies that the phonem consists of. To do so, we have to divide the wave into small frames (for instance 46ms) and compute a spectrum of each frame (we use FFT or LP for this purpose). Unfortunately, such a spectrum consists of too much data (frequencies), therefore we have to select only the most significant piece of information from it. One of such significant information are 'formants' - local extrema of frequencies in the spectrum.

In Fig. 1 the formants have frequencies: 910Hz, 1600Hz, 3160Hz, 4400Hz - they are marked with bold crosses. Usually only four first formants are taken into account because the others

*irek.codello@gmail.com

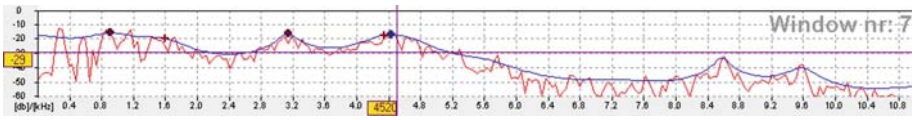

Fig. 1. Spectrum (in decibels) of vowel 'a' Smooth curve is obtained by Linear Prediction and the more frequency-varying one is obtained by the Fourier analysis. The sampling rate is 22kHz.

are irrelevant. It is very hard to calculate formants from the Fourier analysis - due to their variations.

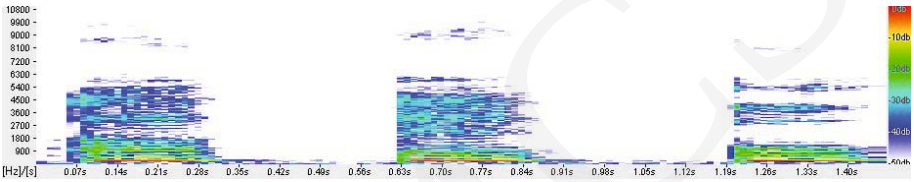

Fig. 2. Fourier spectrogram of the utterance "a e o" (sampling rate 22kHz).

It is far more easier to obtain them from the Linear Prediction analysis - because the spectrogram is smoother.

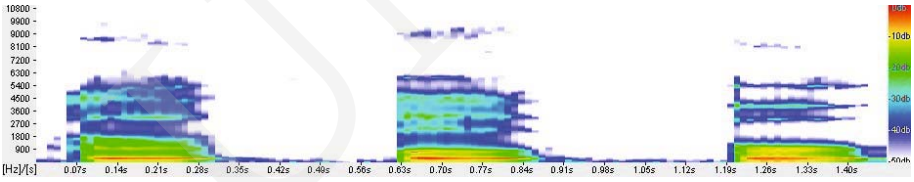

Fig. 3. Linear Prediction spectrogram of the utterance "a e o" (sampling rate 22kHz).

2 Computational procedure

In the LP analysis there is used the Levinson-Durbin algorithm which calculates the prediction parameters α_i and gains the coefficient *G* from the samples *x(t)* of each input signal window. Then, using the transfer function of the form (1):

$$H_z = \frac{S(z)}{U(z)} = \frac{G}{1 - \sum_{k=1}^{p} \alpha_k z^{-k}}, \tag{1}$$

where α - the prediction coefficients, G - the gain parameter. We obtain a spectrum (frequency characteristic) by using the arguments of the form $e^{\omega j}$, which gives:

$$\left|H(e^{j\omega}))\right| = \left|H(e^{j2\pi f/F})\right|, \tag{2}$$

where j – the imaginary unit, F – the sampling frequency, f – the interesting frequency.

A graph created from all spectra (from each window) - is called a spectrogram (Fig. 2 and 3).

3 Spectral peak picking

The most obvious and simple method for formant finding is the spectral peak picking algorithm. Every local extremum in the spectrogram is treated as a formant. Therefore accuracy of spectrogram computation is very important. We have several parameters which effect the algorithm run: window width and window overlap, time window, pre-emphasis and the most important - prediction order. The window width was set to 46ms with 25% overlap - we decided that the frequency precision of such a window is sufficient. We also used a pre-emphasis filter:

$$x(t) = s(t) - \alpha \cdot s(t-1), \tag{3}$$

where s - the input signal, x - the output signal, α - the pre-emphasis coefficient. With the factor $\alpha = \frac{15}{16}$ to eliminate F0 formant (it corresponds to the fundamental frequency which is not desired in our case). Choosing a proper prediction order is quite difficult - the bigger it is, the more detailed spectrum we get. If it is too small - we obtain too small formants, because a spectrum is rather flat with not many extrema, and if it is too large - the spectrum is too detailed and we find formants in the areas where none are supposed to be. Therefore the prediction order should be proportional to the sampling frequency of input signal 1 - we used the formula:

$$p = \frac{F_s}{1000} + C, \tag{4}$$

where p - the prediction order, FS - the sampling rate, C - the constant value from

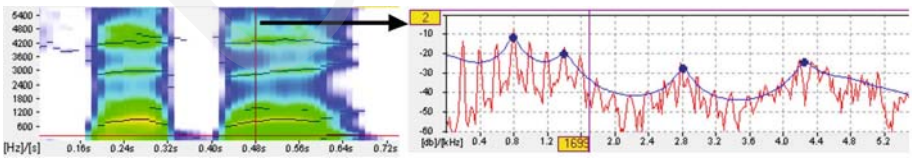

Fig. 4. Spectrogram (on the left) the utterance "papa" (sampling rate 22kHz) and the spectrum (spectrogram intersection) of the selection in the spectrogram computed with FFT and LP. The linear prediction order p=22.

In Fig. 4 (on the right) we can see four formants on the selected spectrogram of vowel 'a'. In the spectrogram (Fig. 4, on the left) the second formant F2 disappears from time to time - the reason is too small prediction order. If we increase the prediction order we will get more and more detailed spectrograms - sometimes too detailed.

4 Root extraction algorithm

Another algorithm taken into consideration was based on the poles extraction of transfer function H(z). To calculate poles (i.e. roots of denominator of H(z) - equation (1)) we have to find complex roots c_k of a complex polynomial:

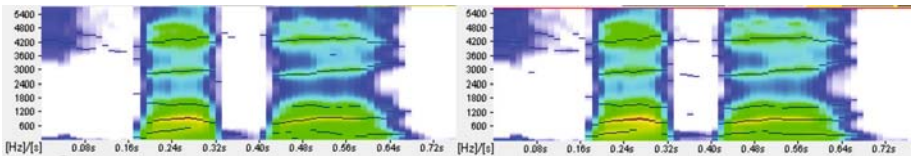

Fig. 5. Spectrogram of the same utterance "papa" (sampling rate 22kHz) but the prediction order is equal to 26 (on the left) and 30 (on the right). More detailed formant trackscan be seen.

$$U(z) = 1 - \sum_{k=1}^{p} \alpha_k z^{-k} = \prod_{k=1}^{p} (1 - c_k z^{-1})(1 - c_k^* z^{-1}). \tag{5}$$

Therefore we solve the equation:

$$U(z) = 1 - \sum_{k=1}^{p} \alpha_k z^{-k} = 0 \Rightarrow A(z) = z^p - \sum_{k=1}^{p} \alpha_k z^{p-k} = 0. \tag{6}$$

Then we have to choose a numerical algorithm to solve equation (5) using the Laguerre's, Muller's or Eigenvalue method - we have chosen the Laguerre's one. All these methods find the first root c_1 and then divide the polynomial $A(z)$ by monomial $(1 - c_1)$ obtaining polynomial $A^1(z)$ with a lower degree. Then we repeat the procedure on $A^1(z)$ until the degree is not zero. Every polynomial division and root finding result in round-off errors which accumulate due to recursive nature of the procedure accumulate. As a result, after finding all roots, we have to 'polish' them - for this purpose we use the same Laguerre method (where unpolished roots are starting points of computations). As formant candidates we take only those roots (poles) which meet the condition: $imag(c_k) > 0$ - we take only complex roots, one from each conjugate root pair $(1 - c_k z^{-1})(1 - c_k^* z^{-1})$. After computing magnitude and phase of each root (by converting $c_k = (a + ib)$ into $c_k = re^{j\varphi}$) we obtain formant's frequency F and 3db formant bandwidth B (in Hz units) using the following equations:

$$\begin{aligned} F &= \frac{f_s}{2\Pi}\varphi, \\ B &= -\frac{f_s}{2\Pi} ln(r), \end{aligned} \tag{7}$$

where f_s - the input signal sampling frequency, φ - the root phase, r - the root magnitude.

5 The application

We created a module in our application 'WaveBlaster', for the graphical representation of formant paths. In the application we can choose which method we want to use to track formants. We can also pick both of them - then on the graph we will see a path in three

colours: red for root extracting, blue for peach picking and green in places where both tracks overlap.

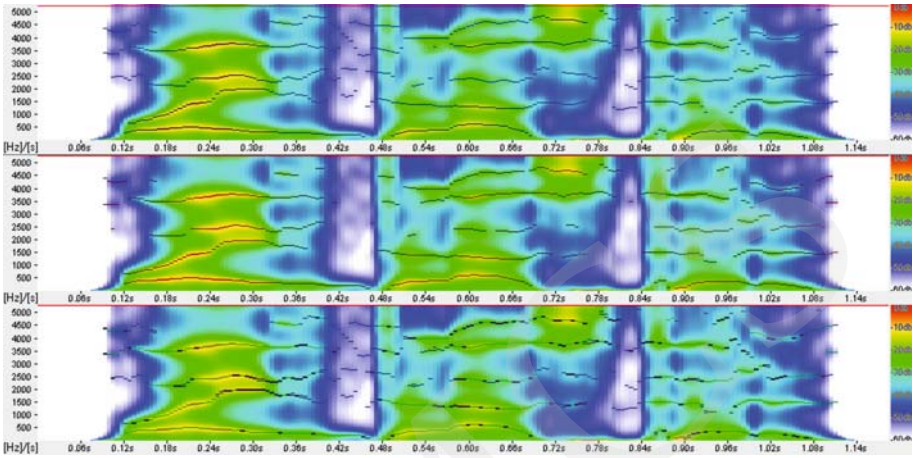

Fig. 6. Formant paths computed with LP p=24 for the utterance 'wave blaster'. On the top - the root extracting formant paths, in the middle - the peach picking formant paths and at the bottom - both of them (overlapping).

6 Comparison

In our opinion, there are no major differences between the described algorithms. Partly it is due to the fact that both methods are based on the same LP transfer function model. The root extracting is slightly more accurate than peak picking because of more theoretical approach and it is a little bit less sensitive to the prediction order. Moreover, the prediction order can be a little smaller than in peak picking to obtain comparable results.

The figures below and above (Fig. 6, 7) show the exemplary comparison of these two methods.

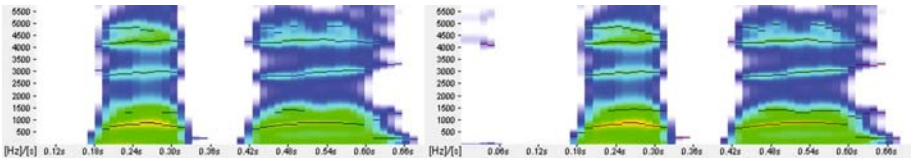

Fig. 7. Spectrograms of the "papa" p=22. On the left the peak picking formant tracks, on the right the root extraction formant tracks.

7 Summary

As can be seen (Fig. 6), both methods give satisfactory results. To make complete use of their advantages, in further analysis, we have to combine the formants into the paths. Then, by

filtering those paths by a low-pass filter, we obtain compact and smooth lines, which are more useful in computations. Then, it is only one step to recognize the voiced phonems like vowels. As one vowel determines one syllable, we can count a number of syllables in an utterance.

Using this approach, as the next step, we will attempt writing a program for determining the utterance rate, whose unit is a syllable per minute. Such a measure is very helpful, for instance, in estimating frequency of disorder occurrence in the utterance, as the utterance length is as important as speech rate in this issue.

References

[1] Codello I., Kuniszyk-Jóźkowiak W., Digital signals analysis with LPC method, Annales UMCS Informatica AI 5 (2006): 315–321.

[2] Rabiner L. R., Schafer R. W., Digital processing of speech signals (Prentice-Hall Inc., New Jarsey, 1978): 354–360.

[3] Zieliński T. P., Od teorii do cyfrowego przetwarzania sygnałów (Wydział EAIiE AGH, Kraków, 2002): 220–235.

[4] Wakita H., Direct estimation of the vocal tract shape by inverse filtering of acoustic speech waveforms, IEEE Transactions on Audio and Electroacoustics, AU-21(5) (1973): 842–866.

[5] Ananthakrishnan K. S., Computer aided pronunciation system (CAPS) (University of South Australia, 2003), http://www.itr.unisa.edu.au/rd/pubs/thesis/ksa.pdf.

[6] Komaee A., Sepehri A., Linear prediction and synthesis of speech signals (Department of Electrical and Computer Engineering, University of Maryland), http://www.enee.umd.edu/ afshin/adsp2/proj2.pdf.

[7] Chanwoo Kim, Kwang-deok Seo, Wonyong Sung, A robust formant extraction algorithm combining spectral peak picking and root polishing, EURASIP Journal on Applied Signal Processing (01.01.2006): 1–16.

[8] Deng L., Attias H., Acero A., Adaptive Kalman Filtering and smoothing for tracking vocal tract resonances using a continuous-valued hidden dynamic model, IEEE Transactions On Audio, Speech and Language Processing 15(1) (2007): 13–23.