Formant Estimation using Gammachirp Filterbank

Kaïs OUNI, Zied LACHIRI and Noureddine ELLOUZE

Laboratoire des Systemes et Traitement du Signal ( LSTS )
ENIT, BP 37, Le Belvédère 1002, Tunis, Tunisie
E-mails : kais.ouni@enit.rnu.tn ; zied.lachiri@enit.rnu.tn & N.ellouze@enit.rnu.tn

Abstract

This paper proposes a new method for formants estimation using a decomposition of speech signal in Gammachirp functions base. It is a spectral analysis method performed by a gammachirp filterbank. A similar approach to the modified spectrum estimation, which allows a smooth and an average spectrum is adopted. In fact, instead of using an uniform window commonly used in short-time Fourier analysis, a bank of gammachirp filters is applied on the signal. A temporal average of the estimated spectra is then applied to obtain one spectrum highlighting the formants structure. This approach is validated by its application on synthesized vowels. The formants are detected with good estimation in comparison with the values given in synthesis. In the same way, this analysis is applied on natural vowels. All the results are compared to three traditional methods, LPC, cepstral and spectral one’s and also to a same analysis given by a gammatone filterbank. The tracking of formants shows that this method, which based on gammachirp filters, gives a correct estimation of the formants compared to traditional methods.

1. Introduction

Over the past forty years and especially in the last decade computational models of the peripheral auditory system have gained popularity in speech processing. These models have shown that in complex speech processing applications, classical spectral analysis can be modified to one’s advantage by adding properties of the auditory system [3][10]. The basic promise of this approach is that understanding cochlear function and central auditory processing such as the remarkable abilities of the human auditory system to detect, separate and recognize speech, will provide new insights into speech processing and will motivate novel approaches to the problems of analysis and robust recognition of acoustic patterns [11][15]. The adoption of such auditory processes has usually led to significant improvements in performance over systems using traditional approaches such as LPC, spectral and cepstral methods [1][12].

The auditory path of a sound is presented as follow. When sound pressure waves impinge upon the eardrum, in the outer ear, they cause vibrations that are transmitted via the stapes, at the oval window to the fluids of the cochlea in the inner ear [2][3]. These pressure waves in turn produce mechanical displacements in the basilar membrane. The amplitude and time course of these vibrations reflect directly the amplitude and frequency content of the sound stimulus [13][15]. These mechanical displacements at any given place of the basilar membrane, can be viewed as the output signal of a band-pass filter whose frequency response has a resonance peak at frequency that is characteristic of the place. Filters with a so-called gammatone impulse response are more used for modelling of the cochlear filterbank [8][16]. The gammatone function has also been used to characterize the spectral analysis of the cochlear filters at moderate levels [14]. It is distinguished by a spectral bandwidth that depends on the central frequency of its corresponding cochlear filter, which is measured in Equivalent Rectangular Bandwidth (ERB) [2][13]. The ERB is related to the psychophysical critical band assignment. The critical bandwidth is about 50-100 Hz at low frequencies and changes gradually to around 20 percent of the frequency at high frequencies. Recently, IRINI [8] has proposed an excellent candidate for asymmetric, level-dependent cochlear filter : the Gammachirp filter. It is an extension of the gammatone filter with an additional chirp term to produce an asymmetric amplitude spectrum. In this work, we tried to use the gammachirp filter in order to estimate formants with a similar approach to the modified spectrum estimation.

In the following sections, we present a time-frequency study of the gammachirp function. Then, we introduce the approach of the gammachirp spectrum estimation and its validation for formants estimation.

2. The Gammachirp Filter

The gammachirp filter is a good approximation to the frequency selective behaviour of the cochlea. It is an auditory filter which introduce an asymmetry and level – dependent characteristics of the cochlear filters and it can be considered as a generalization and improvement of the gammatone filter. The gammachirp filter is defined in temporal domain by the real part of the complex function \(g_r(t)\) [8][9].

\[
g_r(t) = A_{g,B} e^{j\varphi} \exp(-2\pi B t) \exp(j2\pi f_0 t + j c \ln(t) + j \varphi)
\]

with \(B = b \cdot \text{ERB}(f_0)\) and \(\text{ERB}(f_0) = 24.7 + 0.108 f_0\), in Hz

(2)

which is the equivalent rectangular bandwidth at frequency \(f_0\), \(n\) is the filter order, \(f_0\) is the frequency modulation, \(A_{g,B}\) is some normalization constant, \(b\) is a parameter defining the envelope of the gamma distribution and \(c\) a parameter for the chirp rate.

2.1. Energy

The energy of the impulse response \(g_r(t)\) is obtained with the following expression :

\[
E_{g,r} = \|g_r\|^2 = \langle g_r, g_r \rangle = \int_{-\infty}^{\infty} |g_r(t)|^2 dt
\]

(3)

\[
E_{g,r} = A_{g,B} \Gamma(2n-1) \left(4 \pi B \right)^{-1}
\]

(4)

with \(\Gamma(n)\) is the n-th order gamma distribution function. Thus, for energy normalization the normalisation constant must be :

\[
A_{g,r} = \frac{\Gamma(2n-1)}{\Gamma(2n-1)}
\]

(5)
The variances around these average values are respectively:

\[ \sigma_t = \frac{1}{E_{s,s}} \int \left| g_s(t) \right|^2 dt \]

and:

\[ \sigma_f = \frac{1}{E_{s,s}} \int \left| G_s(f) \right|^2 df \]

In the case of the energy-normalized gammachirp function we obtain:

\[ \mu = \frac{2n-1}{4\pi B} \]

\[ \sigma_t = \frac{2n-1}{(4\pi B)} \]

\[ \xi = f_0 \]

\[ \sigma_f = \frac{B}{2n-3} \]

We remark that these values are the same in the case of gammatone function.

The temporal and frequency variances of \( g_s \) satisfy the Heisenberg uncertainty principle for \( n > 3/2 \):

\[ \sigma_t \sigma_f > \frac{1}{4\pi} \]

### 2.4. The Equivalent Rectangular Bandwidth (ERB)

The ERB of a band-pass filter is defined as the width of a rectangular filter with the same peak gain and impulse response energy:

\[ \text{ERB}(G) = \left( \frac{2\pi B}{\Gamma(2n-1)} \right) \]

from (4) and (7) it follows:

\[ \text{ERB}(G) = \left( \frac{2\pi B}{\Gamma(2n-1)} \right) \]

and:

\[ \text{ERB}(G) = \left( \frac{2\pi B}{\Gamma(2n-1)} \right) \]

then, for normalized-energy gammachirp:

\[ \text{ERB}(G) = \left( \frac{2\pi B}{\Gamma(2n-1)} \right) \]

comparing the ERB to the root mean square of the frequency variance \( \sigma_f \) by inserting its expression from (16) it follows:

\[ \text{ERB}(G) = \left( \frac{2\pi B}{\Gamma(2n-1)} \right) \]

It is seen that for \( n = 3 \), the \( \text{ERB} \approx 2\sigma_f \). For this reason, we take this value of \( n \) in (1).

### 3. Gammachirp Spectrum Estimation

In this work, a gammachirp spectrum estimation is performed to estimate formants in a multiresolution and auditory context. A similar approach as the modified spectrum estimation, which allows a smooth and an average spectrum is then adopted. In fact, instead of using an uniform window like Hamming or Blackman which is commonly used in short-time Fourier analysis and which has a constant time-frequency resolution, the gammachirp function, which has not a constant time-frequency resolution but a constant quotient frequency assessment is then used as speech-analysis window centred in a linear-frequencies assessment. The modified estimator has the following expression [12]:

\[ R_s(f) = \frac{1}{K} \sum_{k=1}^{K} R_s(f) \]

where \( K = LM \) is the number of sections, \( L \) is the length of the signal and \( M \) is the length of each section:

\[ R_s(f) = \frac{1}{MP} \sum_{k=1}^{K} s_k(t) w(t) \exp(-j2\pi f t) \]

is the equivalent expression of the simple spectrum estimator, where:

\[ p = \frac{1}{M} \sum_{k=1}^{K} w(t) \]

is the normalization factor introduced in the expression to make the estimator asymptotically non biased. Thus, the expression of the window \( w(t) \) in (23) is replaced by the expression of the energy-normalized gammachirp \( g_s(t) \) centred at \( f_0 \).

The form of the modified estimator becomes:

\[ R_s(f_0) = \left( \frac{2\pi B}{\Gamma(2n-1)} \right) \]

512 channels in a linear-frequencies assessment is then performed using the expression (25) and which each channel is corrected in its peak frequency by \( (c B/n) \) (with \( c=1.5 \) and \( b=1 \)) to obtain one spectrum highlighting the formants structure.
4. Validation

This approach, described above, is validated by its application on synthesized vowels /a/, /i/ and /u/ in noise free with a 11025 Hz sampling rate and 100 Hz as pitch frequency, for formants estimation. The obtained temporal average of the estimated spectra are compared to the spectra given by three classical methods commonly used for formant detection. The first one is the LPC method based on a model of the vocal tract with 17 linear predictive coefficients. The second one is the smooth cepstral method which operate a smoothing on the cepstral coefficients to make smooth the estimated spectra. The third one is the more classical one, the short-time Fourier spectra. Table 1. shows the values given in synthesis and Figures 4, 5 and 6 show the Gammatone- temporal - average spectra compared to LPC, smooth-cepstre and short-time Fourier spectra.

Table 1: The different parameters given in synthesis.

<table>
<thead>
<tr>
<th>Vowels</th>
<th>Pitch</th>
<th>F 1</th>
<th>F 2</th>
<th>F 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>/a/</td>
<td>100 Hz</td>
<td>730 Hz</td>
<td>1090 Hz</td>
<td>2440 Hz</td>
</tr>
<tr>
<td>/i/</td>
<td>100 Hz</td>
<td>270 Hz</td>
<td>2290 Hz</td>
<td>3010 Hz</td>
</tr>
<tr>
<td>/u/</td>
<td>100 Hz</td>
<td>300 Hz</td>
<td>870 Hz</td>
<td>2240 Hz</td>
</tr>
</tbody>
</table>

In the same way, this analysis is applied on natural vowels /a/, /i/ and /u/, spoken by a male in environment noise and recorded at 11025 Hz. All the results are compared to the formants given by the same three traditional methods. The tracking of formants shows that this method gives a correct estimation of the formants compared to traditional methods. It also detects the beginning of voiced-speech sound by the first harmonics of pitch which can be used as a method to voiced speech detection. Moreover, it presents a different formants-bandwidth. This means that this auditory approach gives, in addition to the positions of the formants, their bandwidths that are perceived by the auditory system. It also presents a different ratio between the relative amplitude formants (F1/F2 and F2/F3), which are different from the others ratios given by the classical three methods.

The Figures 7, 8, 9 illustrate the comparison between Gammatone-energy spectrum obtained for /a/,/i/ and /u/ natural vowels and the same ones obtained by LPC, smooth-cepstre and short-time Fourier methods.
5. Conclusion

In this work, we present a spectral analysis for formants estimation performed by a gammachirp filterbank. This spectral analysis is based on an auditory-modified spectral estimation. This method is validated by its application, for formants detection, on both synthesized and natural vowels. The obtained spectra are compared to the LPC, smooth - cepstr and short-time Fourier spectra which show that this method gives a correct formant estimation. It also detects, the voiced speech sound and presents some different spectral characteristics from classical methods. Finally, it seems that this method gives an estimation of formants with less variance.

To test the validity of this assumption, we have to apply this method to a large data base in future work.

6. References