Auditory Model Identification Using REVCOR Method

Lamia Bouafif
Signal and Image processing Laboratory, High Engineering Institute of Tunis, 1002 Tunisia
Email: b2lamia@yahoo.fr

Noureddine Ellouze
Signal and Image processing Laboratory, ENIT, 1002 Tunisia
Email: ellouz.noureddine@enit.rnu.tn

Abstract— Auditory models are very useful in many applications such as speech coding and compression, cochlea prosthesis, and audio watermarking. In this paper we will develop a new auditory model based on the REVCOR method. This technique is based on the estimation of the impulse response of a suitable filter characterizing the auditory neuron and the cochlea. The first step of our study is focused on the development of a mathematical model based on the gammachirp system. This model is then programmed, implemented and simulated under Matlab. The obtained results are compared with the experimental values (REVCOR experiments) for the validation and a better optimization of the model parameters. Two objective criteria are used in order to optimize the audio model estimation which are the SNR (signal to noise ratio) and the MQE (mean quadratic error). The simulation results demonstrated that for the auditory model, only a reduced number of channels are excited (from 3 to 6). This result is very interesting for auditory implants because only significant channels will be stimulated. Besides, this simplifies the electronic implementation and medical intervention.

Index Terms— Auditory Filter, Gammashirp, SNR, MQE, Revcor, Channel Stimulation

I. INTRODUCTION

The study of the auditory models was developed by psychoacoustics and biomedicine specialists especially in the field of the cochlea and auditory implants.

Several auditory models have been developed and implemented on ship such as the DSAM of the HUTear, the Earlab data viewer application which was developed by the hearing research Lab at Boston University and the Auditory Toolbox [1].The DSAM library is programmed in C. It simulates many models such as the Gammatone auditory filter, Meddis [2], and the auditory imagery Model (AIM) of Patterson. In addition of the Gammatone filter, the library contains the auditory model of Lyon [3] and Seneff[4]. Besides, we find in the several references many others models of auditory filters such as those of Revcor[5], Boer[6], Houtgast[7], Patterson an Roex [8].

In this study, we will develop a new auditory model based on the gammachirp filterbank analysis. At the first step, two auditory models will be presented essentially the analytic model and the REVCOR experimental model. In order to validate this model and to optimize its parameters, a comparison between the experimental and the simulated responses will be conducted. Finally, we will apply our strategy on several database speeches in order to determine the hearing bands containing the maximum information and to deduce the appropriate stimulation channels.

II. THE REVCOR PRINCIPLE

The technique of REVCOR or "Reversed Correlation" is a method used by neurophysiologists to estimate the response of an auditory neuron [7]. This technique, initiated by Boer and Jongh, is based on the fact that when the filter input is a white noise, the correlation function between the filter input and output is its impulse response [9], [10].

The following expression presents the impulse response filter under the form of a pulse trains noted S(t) :

$$S(t) = \sum_{i=1}^{N} \delta(t - t_i) \tag{1}$$

ti: represents the releasing times of the electric impulses.
The cross-correlation or REVCOR of the signal X(t) with the output filter is noted $\Phi_{XS}(t)$:

$$\phi_{XS}(\tau) = \int_{-\infty}^{+\infty} X(t)S(t+\tau)dt \tag{2}$$

$$\phi_{XS}(\tau) = \int_{-\infty}^{+\infty} X(t)\big[(X*h)(t+\tau)\big]dt$$

$$= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} X(t)X(t+\tau-t')h(t')dt'dt \tag{3}$$

with:
$h(t)$ represents the filter impulse response.
Changing the order of integration, we make appear the cross-correlation function:

$$\phi_{XS}(\tau) = \int_{-\infty}^{+\infty} \phi_{XX}(\tau-t')h(t')dt' \tag{4}$$

If the input signal is a white noise, then we obtain the following autocorrelation function:

$$\phi_{XX}(\tau) = N_0\delta(\tau) \; \phi_{XX}(\tau) = N_0\delta(\tau) \tag{5}$$

N_0 : is the spectral density of the noise.

Like this, we obtain the following function:

$$\phi_{XS}(\tau) = N_0 h(\tau) \tag{6}$$

III. EXPERIMENTATION

The validation of this method was performed by the experiment of Fig.1 applied on a cat: the result is the impulse response REVCOR.

The experiments values are extracted from Carney Database developed by the EARLAB laboratory of Wisconsin University [9].
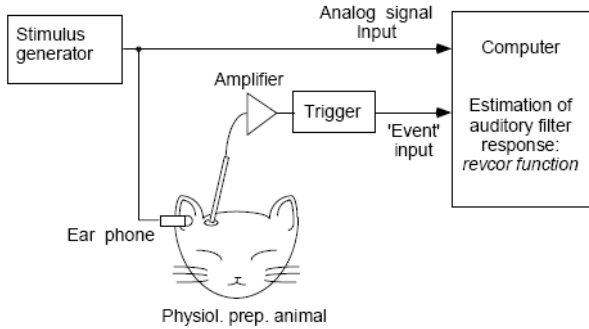

Fig. 1. REVCOR experimentation

An example of the impulse response recorded by the technical Revcor is given by Fig.2.
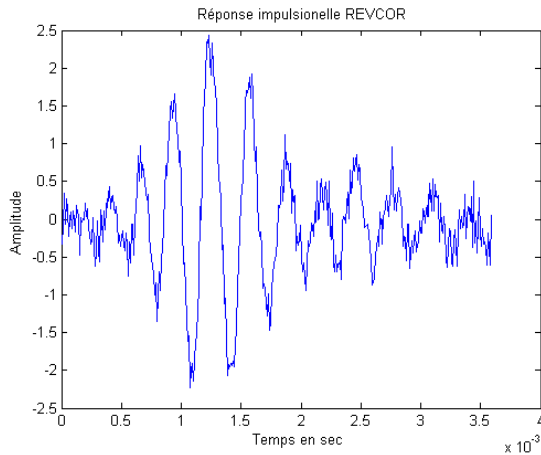

Fig. 2. Revcor impulse response measured on a nerve of an auditory fiber of a cat with 3000 Hz

IV. GAMMACHIRP AND GAMMATONE MODEL

A. Theorical auditory model

The gammatone model has been proposed to the first time by Johannesma. This temporal model was deduced

from the impulse responses measured from the electric impulses of the nervous fibres of the internal ear [4]. Irino and Patterson proposed a new model of the auditory filter called gammachirp, taking into account rectangular auditory bands [11]. The impulse response of the gammachirp filter is given by the following expression [12]:

$$g_c(t) = at^{n-1}\exp(-2\pi bERB(f_r)t)$$
$$\exp(j2\pi f_r t + jc\ln t + j\phi) \tag{7}$$

with:

n : filter order,

f_r : is the modulation frequency of the gamma function,

a: is the carrier normalization parameter,

c : is the asymmetry coefficient of the filter,

ϕ : is the initial phase

bERB: is filter envelope,

ERB : represents the equivalent rectangular band given by [7], [2]:

$$ERB(fr) = 24,7 + 0,108 . fr \tag{8}$$

The ERB of each gammashirp filter is calculated in function of the central frequency (fr) according to Fletcher [3]. If we use the formula of Glasberg and Moore [2] and if we suppose that the signal band is between f_H and f_L with a filter recovery ratio (v) hence, the number of filters (N) is selected like this [13]:

$$N = \frac{9.26}{v}\ln\frac{f_H + 228.7}{f_L + 228.7} \tag{9}$$

However, the central frequencies (fr) can be deduced by the expression [14]:

$$f_r = -228.7 + (f_H + 228.7)e^{-\frac{vn}{9.26}} \tag{10}$$

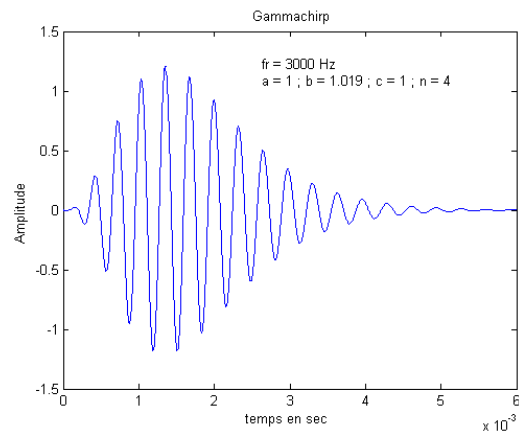An example of the temporal response of the Gammachirp filter is illustrated by the Fig.3.

Fig. 3. Temporal answer of a Gammachirp function centred on 3000 Hz, with a =1, b=1.019, c=1, n=4 and f=0.

The Frequency response of the filter bank Gammachirp model [15] is illustrated by Fig.4. It shows exactly the ear anatomy with multiple bands. The envelope curve

represents the hearing threshold. The channels number, the bandwidth and the central frequencies of each channel are illustrated in table 1. They are computed according the last expressions (8), (9) and (10).
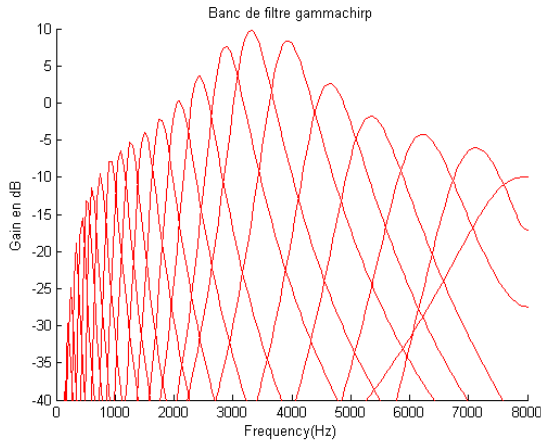

Fig. 4. Frequency response of the Gammachirp model

The following table represents an example of a filter bank decomposition with 25 GFB critical bands.

Table 1. Normalization of critical bands in Bark scale

Central frequency	bandwidth	Channel N°
100	100	1
150	100	2
250	100	3
340	100	4
450	110	5
570	120	6
700	140	7
840	150	8
1000	160	9
1175	190	10
1370	210	11
1600	240	12
1850	280	13
2150	320	14
2500	380	15
2900	450	16
3400	550	17
4000	700	18
4800	900	19
5800	1100	20
7000	1300	21
8500	1800	22
10500	2500	23
13500	3500	24
19500	3500	25

This kind of filter represents a good approximation of the inner ear and especially the cochlea and gives a good estimation of pitch and formants by using the psycho-acoustical experiences [7].
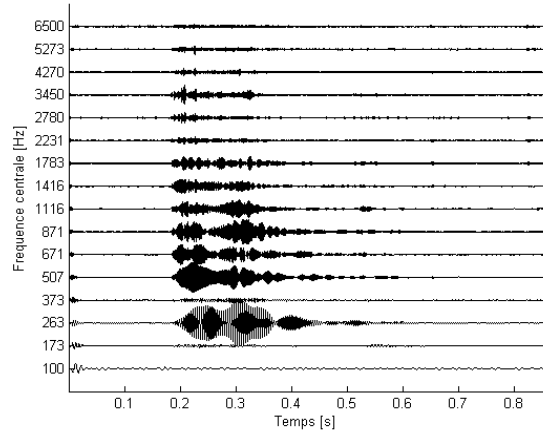

Fig. 5. Temporal representation of the gammachirp filterbank outputs for the **vowel /a/** (N=16 channels)
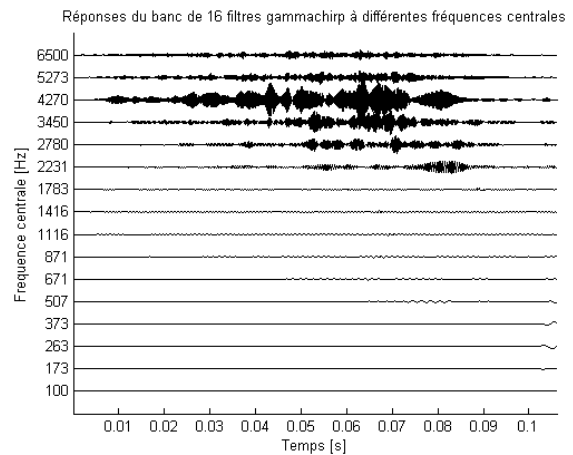

Fig. 6. Temporal representation of the gammachirp filterbank outputs for the **consonant /sh/** (N=16 channels)
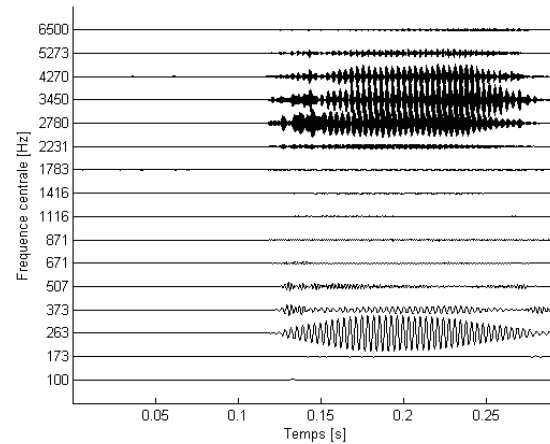

Fig. 7. Temporal representation of the gammachirp filterbank outputs for the **vowel /i/** (N=16 channels)
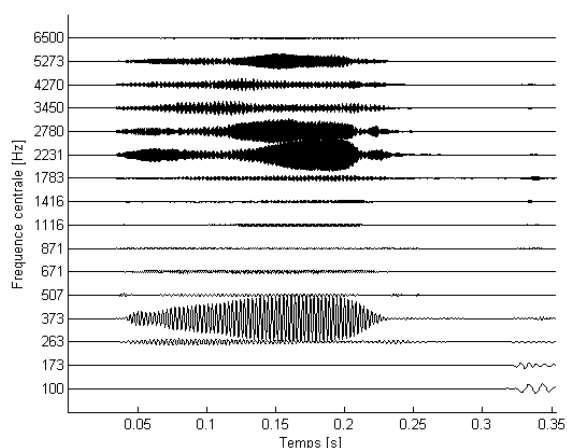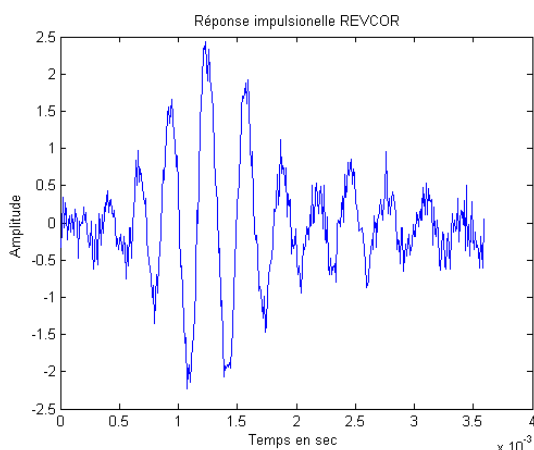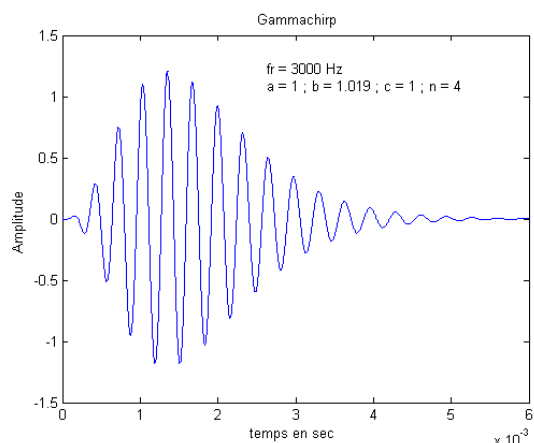
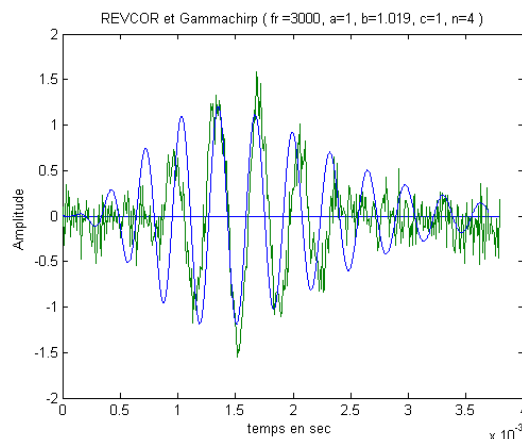Fig. 8. Temporal representation of the gammachirp filterbank outputs for the **vowel /u/** (N=16 channels)

Fig.5, 6, 7 and 8 illustrate a simulation of the temporal responses of the 16 Gammashirp filter channels for the three vowel audio input signal (/a/,/u/, /i/) and an unvoiced audio signal (consonant (/sh/).

(a): experimental response

(b) Gammachirp model response

(c) Comparison between simulation and experimental results

Fig. 9. Superposition of the Revcor impulse response of an auditory nerve with Fr frequency = 3000 Hz and the Gammachirp function centred on Fr, with a =1, b=1.019, c=1, n=4

B. Comparison of Revcor and Gammachirp results

In order to validate the Gammashirp model and to optimize its parameters, a comparison between the REVCOR experimental results and the simulated responses of the analytic model is presented. The validation is conducted by computing at every time the mean quadratic error MQE and the SNR (signal to noise ratio) in order to optimize the recover parameters of the Gammachirp function that gives the minimal value of MQE. Fig.8 represents the superposition of Revcor and the simulated Gammachirp model.

C. Simulation results and parameter optimization

Tables 2 and 3 give the MQE values between the Revcor impulse response| of an auditory nerve and the gammachirp auditory model with the same central frequency.

Table 2. parameter optimization

Fr (Hz)	a	B	c	n	MQE
	2	3	1	4	17.35
	1	3	1	4	5.29
	1	**8**	**3**	**4**	**1.44**
	1	8	5	4	1.45
	1	8	1	4	1.45
1016	1	3	3	4	0.137
	2	3	3	4	0.485
	1	4	3	4	0.033
	1	7	3	4	0.018
	1	**8**	**3**	**4**	**0.017**
	1	8	3	5	0.018
	1	3	1	4	0.132
1523	1	1	1	4	17.72
	2	1	1	4	70.91
	2	3	1	4	0.052
	1	3	1	4	0.019
	2	3	3	4	0.050

Fr	a	B	c	n	MQE
		8	**3**	**4**	**0.008**
	1	8	1	4	0.0081
2035	1	3	2	4	0.5197
	3	3	1	5	0.5217
	1	**2**	**1**	**4**	**0.5084**
	1	3	1	4	0.5108
	2	1.019	1	4	0.5094
	1	8	3	4	0.5217
	3	1.019	1	4	0.5062
2500	1	1	1	4	1.0305
	2	1	1	4	4.1163
	1	2	1	4	0.0094
	1	8	1	4	0.0012
	1	**8**	**3**	**4**	**0.0012**
	1	3	3	4	0.0016
3000	1	1	1	4	0.5005
	1	1.019	1	4	0.4674
	1	**8**	**3**	**4**	**0.2431**
	1	1.02	1	4	0.4658
	1	2	1	4	0.2492
	1	3	1	4	0.2437
	1	3	2	4	0.2445

Fr	a	B	c	n	MQE
	1	5	1	4	$6.9351 \ 10^{-5}$
	1	8	3	4	$6.9104 \ 10^{-5}$
	1	**8**	**3**	**5**	$\mathbf{6.9010 \ 10^{-5}}$
	2	8	3	5	$6.9013 \ 10^{-5}$
	1	1	8	4	0.0116
6406	1	3	1	4	$7.0485 \ 10^{-4}$
	1	3	2	4	$7.0836 \ 10^{-4}$
	1	3	1	5	$7.0796 \ 10^{-4}$
	1	3	8	4	$7.0813 \ 10^{-4}$
	2	**3**	**1**	**4**	$\mathbf{7.0365 \ 10^{-4}}$
	5	3	1	4	$7.1152 \ 10^{-4}$
	1	8	1	4	$7.0813 \ 10^{-4}$
	1	8	3	4	$7.0813 \ 10^{-4}$

The values in bold correspond to the optimized parameters as they conduct to the minimal errors MQE. We can conclude that for the GBF filter, the optimal parameters of the auditory model are:

a=1	B=8	C=3	N=4

D. Noise effect (SNR)

To investigate the robustness of the auditory model, we calculated the number and the order of channels which will be excited in noisy environments. For example, table 4 and fig. 10 represent the number of excited channels (No) according to SNR. We can easily observe that the variation follows the following law:

$$No = N_{max} - 0.6 \ (SNR+5) \qquad (11)$$

Table 4. The noise effect on MQE and stimulated channels

SNR in dB	Number of stimulated channels	Mean Quadratic error: MQE	Maximum number of channels
15	4	0,0021	22
10	6	0,0037	22
5	11	0,0078	22
0	13	0,0095	22
-5	16	0,0110	22
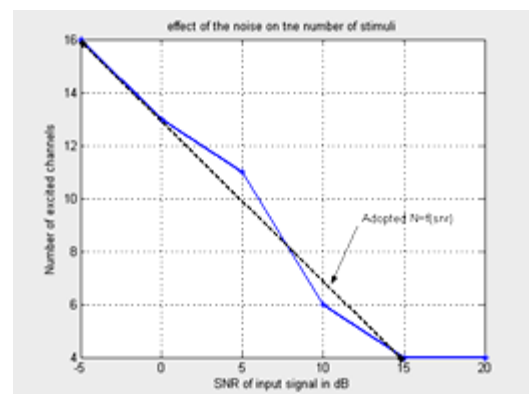
Fig. 10. Number of excited channels in function of SNR

Note that every (Fr) frequency corresponds to the Glasberg band of the GFB filter. Besides, according to the last Fig. 5, 6, 7 and 8, we can deduce that for the auditory model, only a reduced number of channels are excited. For example, for the analysed vowels, there are 3 to 5 (from 16 or 32) significant channels which are stimulated.

This result is very interesting for auditory prosthesis and cochlea implants because it simplifies the electronic implementation.

Table 3. Parameter optimization

Fr (Hz)	a	B	c	n	MQE
3593	1	3	3	4	$2.3267 \ 10^{-4}$
	2	3	3	4	$2.6708 \ 10^{-4}$
	1	3	1	4	$2.14469 \ 10^{-4}$
	1	**8**	**3**	**4**	$\mathbf{1.8913 \ 10^{-4}}$
	2	8	3	4	$1.8931 \ 10^{-4}$
	1	7	3	4	$1.8915 \ 10^{-4}$
3789	1	1	1	4	0.970
	1	2	1	4	0.901
	1	**4**	**1**	**4**	**0.900**
	1	3	1	4	0.901
	1	8	3	4	0.901
	1	3	2	4	0.902
	1	3	3	4	0.901
	1	1	1	4	0.970
4219	1	3	3	4	$7.4375 \ 10^{-5}$
	1	1	1	4	0.0345
	2	1	1	4	0.1375
	1	5	3	4	$5.7165 \ 10^{-5}$
	1	**8**	**3**	**4**	$\mathbf{5.6442 \ 10^{-5}}$
	1	7	3	4	$5.6524 \ 10^{-5}$
4961	1	1	1	4	0.0117
	1	2	1	4	$9.4297 \ 10^{-5}$

Expression (11) shows that with very noisy signals (SNR <-5dB), all channels (n=Nmax) must be stimulated

for high speech intelligibility and acceptable recognition rates. Generally, for each decrease of 5 dB, SNR must be compensated by increasing 3 additional stimulated channels. This result is very important because it solves the problem of performance degradation of the cochlear implant environment and gives priority to electrical stimulation compared to acoustic stimulation.
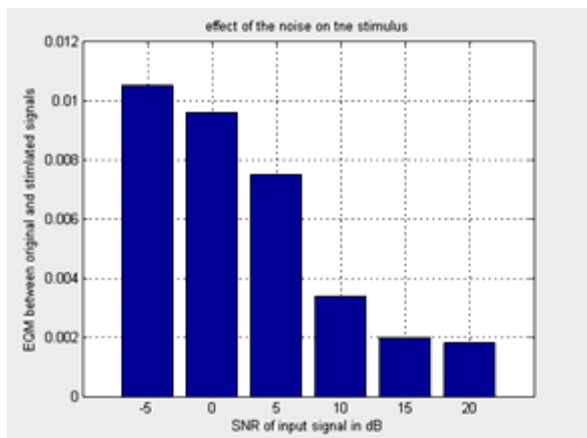

Fig. 11. Mean Quadratic Error vs SNR

E. Energy Channel distribution

To confirm the last results, we have illustrated in the next fig.12 the sixteen channel energy coefficients computed from vowels and consonants localized in several words and pronounced by the same female speaker. We can observe a similar localization of the similar processed speech around 3 to 5 channels. For example, the most energy channels (which will be selected and excited) for the vowel /a/ in the words /dark/ and /had/ are the 4th, 5th and 6th. This means that it is not necessary to stimulate all the channels but only the most 3 significant channels (cochlea electrodes) will be excited.

In fig.13 and fig.14, we can observe that the most significant channels are around the 13th, 14th, 15th and 16th channels contrarily to the vowels where the excited channels are between the 4th and the 8th channel. These results show a good correspondence between the vowels in different words and their positions. Besides, it enhances the discrimination between vowels and consonants and conduct to a better voice intelligibility.
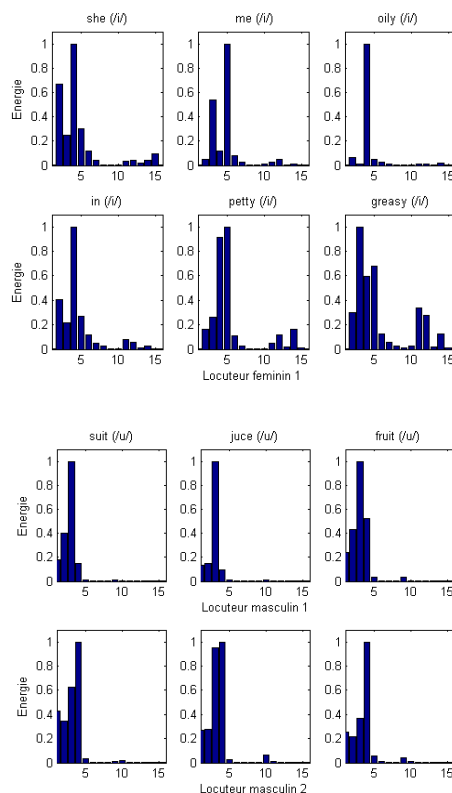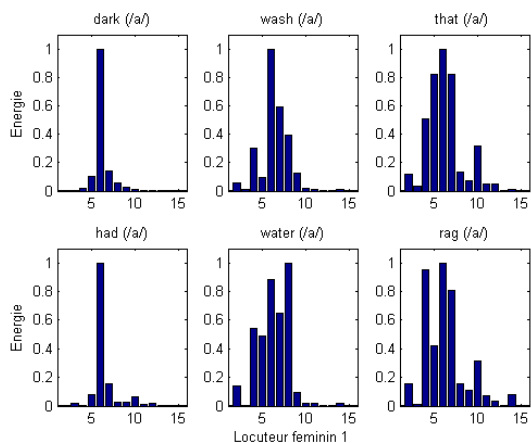

Fig. 12. Energy distribution of the three vowels /a/, /i/ and /u/ of the auditory channels in several words
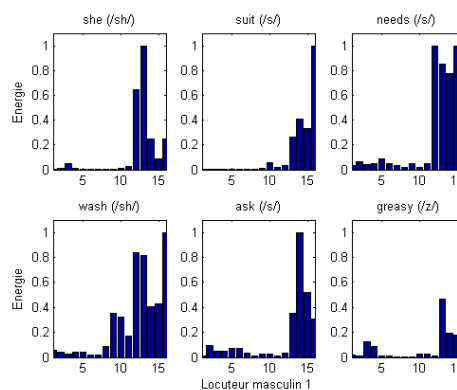

Fig. 13. Energy distribution of three unvoiced consonants /sh/, /s/ and /z/ of the auditory channels in several words (for a male speaker)
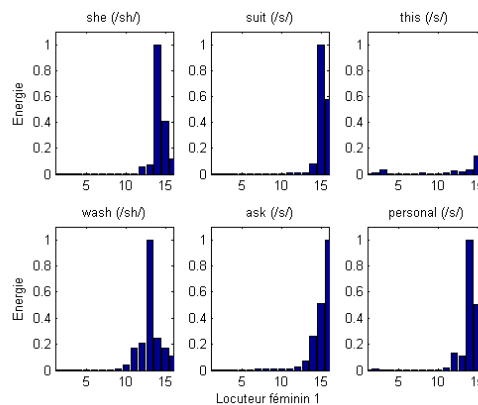

Fig. 14. Energy distribution of three unvoiced consonants /sh/, /s/ and /z/ of the auditory channels in several words (for a female speaker)

V. Conclusion

In this study, we presented a new auditory model based on the gammachirp filter-bank analysis. The parameters estimation of this model was conducted by comparison between the experimental Revcor strategy and the simulated auditory model responses. We succeeded to optimize the model parameters minimizing the mean quadratic error computed from the simulation results of the auditory model and the REVCOR experimental measurements. The implementation of this work was programmed with Matlab and C.

The simulation results have shown that a reduced number of 3 to 5 channels are sufficient to encode the speech signal. However, in noisy environments, it is necessary to compensate each SNR degradation of 5dB by adding three additional stimulation channels.

References

[1] M. Slaney Auditory, Auditory Toolbox, Technical Report 1998.

[2] M. J. Hewitt & R. Meddis: Implementation details of model computation of innerhair-cell for auditory-nerve synapse, journal of Acoustical Society of America, vol.87, no.4, p. 1813-1816, April 1990.

[3] M. Slaney: Lyons Cochlear Model, Apple Computer Technical report 13, 1988.

[4] S. Seneff: Mean-rate model of processing, journal speech auditory of Phonetics, pp.55-76, 1988.

[5] Carney, L.H., Megean j. Shkhter M, I.,"Frequency glides in the impulse responces of auditory-nerve fibers", J.Acoust.soc.Am, No.4, pp.2384-2391, 1999.

[6] Boer, E.,and Nuttall, A. L. ''The mechanical waveform of the basilar membrane. I. Frequency modulations in impulse responses and cross-correlation functions," Journal. Acoust. Am. 101, 3583–3592. 1997

[7] K. Ouni. Analysis of the vocal signal using of the acquaintances on the auditory perception and frequency time representation of signals. PHD Thesis ENIT, Tunisia 2003.

[8] H. Steven Colburn H. Laurel and Carney "Quantifying the implications of nonlinear cochlear tuning for auditory-filter estimates" JASA Journal, Vol 19. November 2001

[9] Earlab data viewer application, Hearing research center at Boston University 2009.

[10] T. Irino, D.Patterson. À time-domain, level dependent auditory filter : the gammachirp. *J.Acoust of Am.* 101(1): 12-419, January, 1997.

[11] M. H.Allerhand and Christian Giguere,"Time-domain modelling of peripheral auditory processing: À modular architecture and a software platform," Journal of the Acoustical Society of America, vol 98, pp 1890-1894, 1995.

[12] The Development System for Auditory Modeling http://www.essex.ac.uk/psychology/hearinglab/lutear/

[13] Patterson, R.D., Nimmo-Smith, I.,Wiber, D.L., and Milroy, R.,"The deterioration of hearing with age: Frequency selectivity, the critical ratio, the audiogram, and speech threshold", J.Acoust. Am., No.6, December 1982.

[14] Patterson, R.D., Nimmo-Smith, I.,"*Off-frequency listening and auditory-filter asymmetry", J.Acoust.* PloughshareAm., No.1, 1980.

[15] T. Irino, R.D. Patterson. Temporal asymmetry in the auditory system. *J. Acoust. Am.* 99(4):2316-2331, April, 1997.

Authors' Profiles

Mrs Bouafif Lamia has received…his Engineering diploma and M.S. degrees in Informatics from Science Faculty of Tunis. She obtained his PHD in Signal processing from the Engineering faculty of Tunis ENIT in 2010. Actually, she is an assistant professor at the High Institute of Biomedical Studies of Tunis where she teaches the signal and speech processing. Besides, she is a research member at the Signal and image processing laboratory of the High Institute of Engineering of Tunis (ENIT).

Mr Ellouze Noureddine is actually a senior professor at the High Institute of Engineering of Tunis. He received a Ph.D. degree in 1977 at INP (Toulouse- France). He is also the Director of the Research Laboratory LSTS at ENIT. Pr. Ellouze has directed multiple Masters and Thesis and published more than 300 scientific papers in journals and proceedings, in the domain of signal processing, speech and image processing, biomedical applications and pattern recognition.