SPEECH ANALYSIS USING WAVELET TRANSFORMS DEDICATED TO COCHLEAR PROSTHESIS STIMULATION STRATEGY

Imène CHEIKHROUHOU (1), Rabi BEN ATITALLAH (1), Kais OUNI (2),
Ahmed BEN HAMIDA (1), Nouri MAMOUDI (1) & Noureddine ELLOUZE (2)

(1) Laboratoire d'Electronique et des Technologies de l'Information (LETI)
ENIS, B.P.W, 3038 Sfax, Tunisia Phone: (216)74.27.40.88 Fax: (216)74.27.55.95
(2) Laboratory of Systems and Speech processing (LSTS), ENIT Tunisia
imene.cheikhrouhou@laposte.net a_benhamida@yahoo.com

ABSTRACT

In this article, we studied speech analysis using wavelet transforms in order to conceive a flexible stimulation strategy for cochlear prostheses. In our first experiment, we used two types of wavelets that are convenient for such speech analysis. This was based on parameterisation techniques of word by using bank filtering methodology. However, the constructed filters had the considered wavelets' shape impulse response. In fact, as a first experimentation, Morlet wavelet shape was tested and compared to Gammachirp wavelet shape. Two filter bank models were then conceived according to these two wavelet shapes in order to apply numerical filtering. Our purpose is to exploit such filtering for constructing a completely flexible cochlea's stimulation strategy. So, this would be very useful for specifying performances of each of these two analyses and for providing comparisons that would be useful for an ulterior strategy implementation on the prosthesis external part.

1. INTRODUCTION

Cochlear prosthesis is a recent apparatus intended to deep and total deafness rehabilitation where conventional hearing aids are not efficient. It is a scientific progress permitting to remedy to such handicap that causes social disintegration. [1].Two essential parts compose such apparatus: An external part generally piloted by a dedicated processor, a DSP, and an internal part which is commanded by the external one for generating adequate stimuli to the cochlea. Different strategies of stimulation have been conceived during the past few years [1].These were generally conceived around a numeric filtering as the 'RIF' type or around a spectral analysis based on FFT transforms [1]. In our research, we propose a speech analysis based on wavelet transforms 'W.T' in view to conceive an original stimulation strategy completely flexible. We conceived then 21 numeric filters, i.e a filter bank, that are constructed around two wavelet shapes.

In fact, a comparative study was then established between an analysis using the Morlet wavelet and another analysis using the Gammachirp wavelet [1]. These two shapes gave important results permitting then a great variety of choices when adjusting the stimulation.

2. PREVIEW ON USED WAVELET TRANSFORM

The W.T is an atomic decomposition whose atoms are descended of the same function, 'the mother wavelet' by applying translation'b' and dilation 'a' operations [2]:

$$W_\psi(a,b) = \int_{-\infty}^{+\infty} \psi^*_{a,b}(t)\,.f(t)dt. \tag{1}$$

This analysis consist in using a family of $\psi_{a,b}(t)$ constructed from a function ψ called the mother wavelet:

$$\psi_{a,b}(t) = \frac{1}{\sqrt{a}}\,.\,\psi(\frac{t-b}{a}),\ \text{where}\ a \in \Re^+,\ b \in \Re. \tag{2}$$

The CWT (continuous wavelet transform) consists on giving wavelet coefficients $W_\psi(a,b)$ defined by:

$$W_\psi(a,b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{+\infty} f(t)\overline{\psi}(\frac{t-b}{a})dt \tag{3}$$

$$f(t) = \frac{1}{C_\psi} \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} W_\psi(a,b)\psi_{a,b}(t)\frac{da}{a^2} \tag{4}$$

The discrete wavelet transform DWT is an analysing functions issued from the family $\psi_{m,n}$ given by:

$$\psi_{m,n}(t) = a^{-\frac{m}{2}}\psi(a^{-m}t - nb) \tag{5}$$

In our first tentative, we used the Morlet wavelet shape. It is a simple analytic shape, very resolute in time and in frequency, regular, locally periodic, with non compact support. It is formed by complex values of the shape of a complex sinus modulated by a gaussian envelope [2].

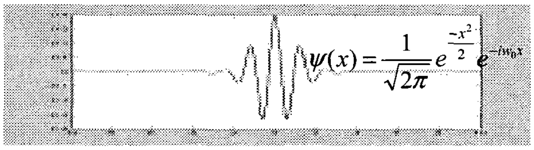

Fig. 1 : *Morlet Mother wavelet*

0-7803-8379-6/04/$20.00 ©2004 IEEE.

As second test shape, we used Gammachirp Wavelet [3]. Irino was the first one who proposed the Gammachirp filter that was a filtering temporal model deducted of an impulse response modelling, by using electric impulses' measurement of cat internal ear nervous fibers. It presents a non symmetrical envelope dependent of the level of intensity of the applied resonant stimulus [4]:

$$g(t) = \lambda_n t^{n-1} e^{-2\pi\alpha t} . e^{j(2\pi f_0 t + c \ln(t) + \varphi)} \qquad (6)$$

- n : the order of the corresponding filter.
- f_0 : the modulation frequency of the Gamma function.
- φ : the initial phase.
-λ_n : a normalising parameter.
-α : equivalent rectangular bandwidth of the cochlea
- c : parameter of asymmetry filters.

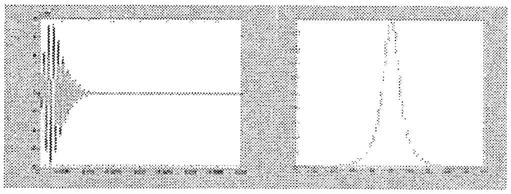

Fig. 2 : I*mpulse response of 'Gammachirp' shape with its corresponding frequency illustration for 1000Hz* [1]

3. COCHLEAR FILTERS' CONCEPTION

Our cochlear prosthesis stimulation strategy could be summarized by the following figure:

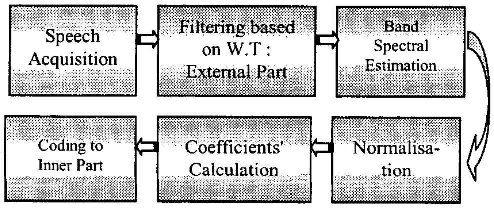

Fig.3: S*peech coding to the implant* [1]

In the case of filters' conception from the studied wavelet transform, it was necessary to take account of the critical bands: Two pure sounds having different frequency and presented simultaneously have a constant sonie. The width limit of the gap between frequencies represents the critical band [5]. The peripheral auditory system is equivalent to a band pass filterbank of which respective bands would have the critical bandwidth [6]. These last are formulated under shape of several mathematic models among these critical Munich bands: [7] $B_M = 25 + 75(1+1.4F^2)^{0.69}$

and Cambridge bands: [8] $B_C = 24.7\ (1+4.3F)$

Where F is the central frequency of the band (in KHz). Munich equation was then adopted in our work since that there is a similarity between widths of these bands with those of critical bands to corresponding frequencies [1].

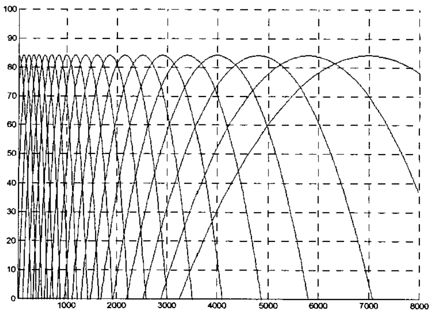

Fig.4: *Morlet Wavelet's filterbank for Munich bands* [1]

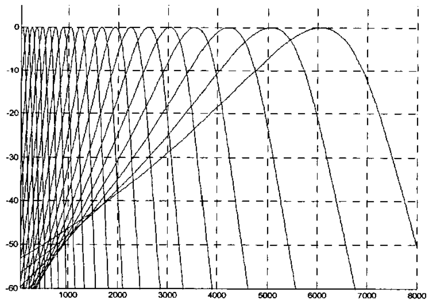

Fig.5: *Morlet Wavelet's filterbank for Cambridge bands* [1]

4. SPEECH ANALYSIS USING WT

Cochlear prosthesis stimulation strategy was based on these two WT shapes since they were suitable to speech processing. We proceeded to a comparative survey at the time of this speech analysis, and however, we proposed a choice option between one or the other being given that the pathological cases were varied. We used the distance calculation to compare parameters' vectors [1]:

$$d = \frac{1}{21}\left(\sum_{k=1}^{21} \left| x_k - y_k \right|^2 \right)$$

4.1. Localized phonemes case (two different speakers)

Our system was tested by using phonemes of TIMIT Basis. This base gave us for every sentence and for every speaker several possibilities such as the female or the male voice as well as the segmentation possibility according to words or according to vowels and consonants. Indeed, it supplied a label of the beginning and the end for every segmentation type. Selectivity is always better in the case of Morlet. However, the supplementary bars that we notice in the case of Gammachirp, and that denote no selectivity, could generate enriching stimulations for certain cases of patients. In this parameterisation, vowels always kept low frequencies. Also, the maximum of parameters for every vowel generally kept the same frequency positions.

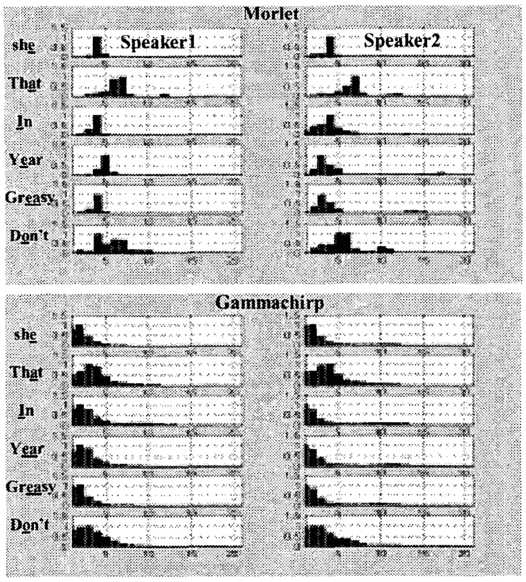

Fig. 7: *Representation of 21 parameters for different* ***vowels***

Table1. *Distance between different* ***vowels*** ***: Morlet***

Sp1/2	She	that	In	year	greasy	don't
She	**0.0005**	0.1211	**0.0008**	0.0660	**0.0007**	0.0570
that	0.1017	**0.0071**	0.1029	0.0875	0.1024	0.0545
in	**0.0234**	0.1156	**0.0182**	0.0664	**0.0193**	0.0471
year	**0.0543**	0.1298	**0.0408**	0.0736	**0.0437**	0.0992
greasy	**0.0531**	0.1351	**0.0387**	0.0856	0.0415	0.0993
don't	0.1143	0.0752	0.1163	0.0545	0.1171	**0.0516**

Table2. *Distance between different* ***vowels*** ***: Gammachirp***

Sp1/2	she	that	In	year	greasy	don't
She	**0.0006**	0.0587	**0.0086**	**0.0115**	**0.0002**	0.0363
That	0.0599	**0.0009**	0.0330	0.0326	0.0653	0.0108
In	**0.0077**	0.0698	**0.0199**	0.0258	**0.0058**	0.0531
Year	**0.0040**	0.0642	**0.0147**	**0.0190**	**0.0027**	0.0450
Greasy	**0.0040**	0.0654	**0.0148**	**0.0196**	**0.0024**	0.0458
don't	0.0300	0.0161	0.0109	0.0114	0.0331	**0.0015**

The system demonstrated a modification of the minimums' distance path. We could realize that the fundamental frequency is variable.

In case of consonants, they cover all bands of frequencies but don't keep the same positions for maxima parameters.

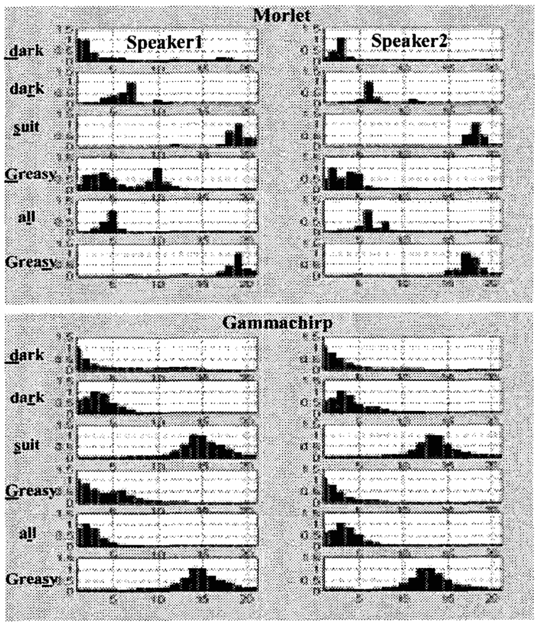

Fig.8: *Representation of 21 parameters for different* ***consonants***

Table 3. *Distance between* ***consonants Morlet***

Sp1/2	dark	dark	suit	greasy	all	greasy
Dark	**0.0703**	0.1168	0.1443	0.1004	0.0935	0.1220
Dark	0.1346	**0.0455**	0.1377	0.1610	0.0822	0.1149
Suit	0.1521	0.1361	**0.0416**	0.2197	0.1305	**0.0426**
Greasy	**0.0493**	0.1343	0.2069	**0.0654**	0.0689	0.1840
All	0.1376	**0.0582**	0.1455	0.1580	**0.0774**	0.1220
Greasy	0.1601	0.1489	**0.0808**	0.2315	0.1421	**0.0728**

Table 4 . *Distance between* ***consonants Gammachirp***

Sp1/2	**Dark**	dark	suit	greasy	all	greasy
Dark	**0.0067**	0.0457	0.2328	0.0229	**0.0093**	0.2270
Dark	0.0606	**0.0021**	0.2577	0.0313	0.0255	0.2490
Suit	0.1666	0.2553	**0.0364**	0.2387	0.2323	**0.0341**
Greasy	**0.0057**	0.0530	0.2327	**0.0251**	0.0115	0.2278
All	0.0608	**0.0018**	0.2678	0.0331	**0.0226**	0.2587
Greasy	0.1762	0.2635	**0.0724**	0.2451	0.2427	**0.0670**

There is a modification of the path of minimums' distance path. An overlap appears at the level of consonants.

4.2. Localized words case: Two different speakers

Words of TIMIT basis were tested (already used with phonemes) : Selectivity was always better for Morlet case, and noticed supplementary bars for Gammachirp denoting no selectivity for this case could generate enriching stimulations for certain cases of patients.

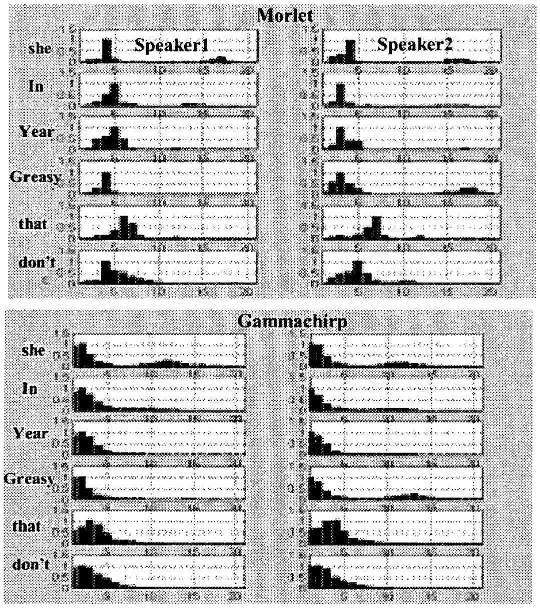

Fig. 9 : *Representation of 21 parameters for different words*

Table 5. Distance between different ***words*** ***Morlet***

Sp1/2	she1	in1	Year1	greasy1	that1	dont1
She2	**0.0050**	0.0564	0.0628	**0.0030**	0.1130	0.0378
In2	0.0740	**0.0852**	0.0762	**0.0586**	0.1088	0.1025
year2	0.0580	0.0526	**0.0424**	**0.0446**	0.1087	0.0732
greasy2	0.0643	0.0797	0.0706	**0.0564**	0.1294	0.0886
That2	0.1038	0.0890	0.0936	0.1017	**0.0201**	0.0719
dont2	0.0675	0.0127	**0.0060**	0.0656	0.0577	**0.0273**

Table 6. Distance between different ***words*** ***Gammachirp***

Sp1/2	she1	in1	Year1	greasy1	mthat1	dont1
She2	**0.0022**	0.0125	0.0133	**0.0040**	0.0540	0.0306
in2	0.0109	**0.0199**	0.0195	**0.0058**	0.0589	0.0395
year2	0.0082	0.0128	**0.0113**	**0.0018**	0.0500	0.0294
greasy2	**0.0051**	0.0217	0.0231	**0.0111**	0.0591	0.0405
That2	0.0632	0.0329	0.0353	0.0620	**0.0030**	0.0151
dont2	0.0209	0.0042	0.0038	0.0157	0.0155	**0.0016**

5. DISCUSSION

The comparison between the two wavelets' shape, Morlet and Gammachirp, cannot be an easy task, nor a decisive one. We tried all the long of this work to put in evidence advantages and inconveniences of every filtering type. On the other hand, in this research, we try to explore each type in order provide to clinicians a complete and flexible strategy capable of being adaptable to the great variety of pathological cases. As a summary, we could affirm that:

1- For phonemes pronounced by two different speakers:

For phonemes, one can say that for the two wavelets' types, maxima of parameters for every phoneme generally kept the same frequency positions.

Morlet wavelet shape provided more selective filtering whereas Gammachirp wavelet shape provided a minimal distance so better discern ability.

2- For words pronounced by two different speakers:

For the two wavelets' types, maxima of parameters for every word almost kept the same frequency positions.

The distance values were minimal for the two wavelets types which gave us the better discern ability.

6. CONCLUSION

Speech analysis using wavelet transforms was studied in this paper. Two wavelet shapes were tested for pure sounds, for phoneme and for words using the TIMIT basis. Our goal was to provide essential information in order to conceive a complete and flexible stimulation strategy for cochlear prostheses. Filter bank was then constructed for these two shapes of wavelets and testing on sounds was then performed. For the Morlet wavelet shape filter-bank testing, good selectivity was noticed. For the Gammachirp wavelet shape testing, differentiation of phonemes was noticed for two speakers' experimentation. Indeed, the distance by the Gammachirp wavelet is smaller than that of Morlet wavelet. This prove that this wavelet shape is less affected by the pitch variation, so that it could preferred in multi-speakers discourse.

7. REFERENCES

[1] Imène Cheikhrouhou, '*Analyse du signal sonore par transformée en Ondelettes en vue de concevoir une stratégie de stimulation pour la prothèse cochléaire'*, JSF_03, Tozeur, Décembre 2003 (en cours et accepté)

[2] I. Daubechies, '*Ten lectures on waveletes'*, SIAM, Philadelphia, PA, 1992

[3] P. IM. Johannesma,'*The pre-response stimulus ensemble of neurons in the cochlear nucleus'*, Symposium of Hearing Theory, pp., 1972.

[4] T. Irino, '*A Gammachirp function as an optimal auditory filter with meellin trasform*', IEEE ICASSP 96.

[5] K. Ouni., '*Formant Estmation using Gammachirp Filterbank'*, Eurospeech 2001 Scandinavia Conference, Denmark, September 2001

[6] J.O SmithIII & J.S Abel, '*Bark and ERB Bilinear transforms'*, IEEE Trans. On speech and Audio Processing, 1999.

[7] E. Zwicker & E. Terhardt, '*Analytical expressions for critical band rate and critical bandwidth as a function of frequency*', J.Acoust. Soc.Am, Vol 68, No 5, Nov, 1980

[8] B.C.J Moore & B.R. Glasberg, '*Suggested formulae for calculing auditory-filter bandwidths and excitation patterns*', J. Acoust. Soc.Am., Vol 74, No.3, 09/1983.