

COVER PAGE

Document downloaded by @DAEL

Thu Nov 26 16:16:51 2026
For personal use

When automatic English translation is provided, only the original document is authentic.
The EAA cannot be held responsible of any translation error

Bibliographical reference

Exploring Wavelet Transform Based Methodology for Cochlear Prosthesis Advanced Speech Processing Strategy, Mohamed Ghorbel, Amira Derbel, Fathi Kallel, Mounir Samet and Ahmed Ben Hamida, Acta Acustica **vol. 100** (Number 2), 2014, pp. 341-352

DOI
https://doi.org/10.3813/AAA.918713

ACTA ACUSTICA UNITED WITH ACUSTICA
Vol. 100 (2014) 341 – 352

DOI 10.3813/AAA.918713

Exploring Wavelet Transform Based Methodology for Cochlear Prosthesis Advanced Speech Processing Strategy

Mohamed Ghorbel[1)], Amira Derbel[1)], Fathi Kallel[1,2)], Mounir Samet[1)], Ahmed Ben Hamida[1)]

[1)] ATMS Research Unit, Advanced Technologies for Medicine and Signals, ATMS – ENIS, University of Sfax, Tunisia. amiraderbel06@yahoo.com

[2)] PACS Team, INSERM Unit 1028: "Cognition and Brain Dynamics", Lyon Neurosciences Centre, EPU-ISTIL, Claude Bernard University, Boulevard du 11 Novembre 1918, 69622 Villeurbanne, France

Summary

In this paper, a novel strategy for cochlear electrical stimulation based on non-linear human auditory physiological properties was developed. This strategy used the continuous Wavelet Transform (WT) as a speech processing tool. The most distinguished characteristic of the WT tool is its efficient time–frequency resolution. Such advantageous feature for cochlear stimulation could be adaptively adjusted not only by its position parameter that arranges frequency division but also by its scale parameter that could adjust the bandpass width. Following human auditory physiological properties in order to be close to the cochlear behavior, our conceived filter bank structure was divided into 21 frequency bands according to the Equivalent Rectangular Bandwidth (ERB) model. A comparative study was performed between the proposed WT speech processing tool and the classical Short Time Fourier Transform (STFT) speech processing tool. For performance evaluation, some objective speech assessment tests relying on spectral leakage (frequency selectivity) and tracking temporal feature (sensitivity) were performed. In order to evaluate the speech intelligibility, subjective listening tests were also assessed with 50 normal hearing listeners using a specific Cochlear Implant simulator. Experimental results showed that the WT speech processing tool better improves speech understanding compared to the STFT speech processing tool. As for more concretization, and in order to prove the validity of the previously mentioned comparative study, a real-time implementation testing of our strategy performed on a specific Digital Signal Processor (DSP) platform was also described. Using a 720 MHz TMS320C6416 DSP board, the execution of our proposed algorithm took only 10.5 µs per sample. This execution time is sufficient to process the speech signal in real-time.

PACS numbers: 43.25.Ts, 43.71.Ky, 43.72.Ar

1. Introduction

Total or even profound deafness could not be usually remedied or rehabilitated through the use of conventional hearing aids even in the case of important sound amplification. Such a phenomenon could be essentially due to the common reason of the loss of an important range of hair cells in the inner ear, i.e the cochlea. The cochlear prosthesis device, sometimes referred to as Cochlear Implant (CI), is an important innovation, "a miracle", which proved its success in possible deafness rehabilitation when stimulating the cochlea remaining hearing cells so as to restore the patient's hearing capacity even partially. Thus, CI operates by directly stimulating the auditory nerve using an electrical pulse, thereby bypassing the normal hearing mechanism [1].

This cochlear prosthesis apparatus is made up of two essential parts: an external speech processing part that is Radio Frequency (RF) linked to an internal part essentially for stimulation. For the external part, which could be considered the main one, different speech processing strategies are currently in use to encode the incoming acoustic signal into an appropriate set of electrical stimulating pulses to be applied by the internal part. All strategies were almost based on STFT spectral or filter bank approaches: a bank of bandpass filtering estimates the signal spectral content to be numerically encoded and then to be RF transmitted for each CI stimulation channel which covers specific range frequencies along the cochlea's tonotopy [2].

Recently, major research efforts have been invested in seeking effective strategies for extracting important sound features which could be encoded into a usable way for CI devices [3, 4]. Another criterion to be carefully studied is the real-time implementation considering that our target was to build a flexible and user-friendly algorithm to be implemented into the speech processor. Fine spectral resolution is highly desirable for the transmission of fine spectral cues ("fine structure") to the implant recipients. Fine resolution is also needed for the implementation of

Received 21 October 2011,
accepted 13 October 2013.

© S. Hirzel Verlag · EAA

focused current stimulation strategies [5]. Hence, the implementation needs to provide a good temporal resolution, which in turn requires high stimulation rates.

Clearly, the aforementioned factors are interrelated. For instance, the use of filter banks for spectral decomposition provides excellent temporal resolution, but inescapably limits the frequency resolution (due to the number of channels). The use of STFT provides excellent frequency resolution. Nevertheless, such use restrains the temporal resolution (because of its frame rate update) [6]. The WT tool provides an accurate resolution both in frequency and time, as it is one of the most appropriate tools to analyze such non-stationary signals as speech [7, 8, 9]. These WT tool properties might suggest, from a theoretical point of view, that such a tool could be suitable for cochlear prosthesis devices [10, 11]. In fact, the WT for speech processing in CIs has been subject of many recent studies such as [12, 13, 14, 15, 16, 17, 18]. In another challenge, designers of CI prosthesis were faced with the developing signal processing techniques that would mimic the normal cochlea function [19].

Although a great deal of research has been published on the use of a method of signal processing based on an auditory model [4, 10, 11, 20, 21, 22], only few studies actually tried to link the wavelet transform to the non-linear auditory model. In the present paper, we therefore propose to explore the use of an auditory model-based WT for frequency decomposition, in such a way as to investigate a user-friendly and flexible speech processing strategy for CI. To spotlight the importance of our proposed strategy, a comparative study was carried out with an auditory based STFT approach. We evaluated and compared the performance for objective measures in terms of frequency selectivity and sensitivity as well as subjective measures in terms of intelligibility. This research on cochlear prosthesis speech processing was then extended to report on a real-time implementation on a DSP platform to validate the practicality of our proposed strategy.

The remaining of this paper is organized as follows: An overview of common speech processing strategies is provided in section 2. Section 3 describes WT tool based CIS strategy. Section 4 describes a comparative study between Matlab implementation of STFT and WT tool speech processing. Section 5 covers the DSP implementation aspect and the real time results of our proposed strategy.

2. Common cochlear prosthesis speech processing strategies

The cochlear implant prosthesis is based on the idea of bypassing the normal hearing mechanism, i.e outer, middle and part of the inner ear including the hair cells. Therefore, its design usually includes two main parts: an external unit, the DSP sound analyzer, and an internal surgically implanted unit, the implant [1]. These two essential parts are connected by an inductive link as illustrated in Figure 1.

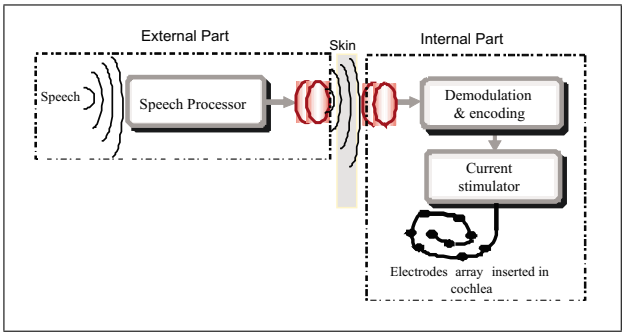

Figure 1. Cochlear prosthesis common basic architecture.

The external part includes a microphone that picks up the sound and a digital signal processing unit that converts the sound into a specific electronic code based on the adopted stimulation strategy. The internal part demodulates the code and generates electrical stimuli that drive the electrode array which is usually placed close to the residual nerve fibers inside the cochlea [23]. In a single-channel implant, only one electrode is used for cochlea stimulation. In the 'multichannel' cochlear implant systems; an electrode array is usually inserted into the cochlea so that different auditory nerve fibers can be stimulated at different spots in the cochlea, thereby exploiting the place mechanism for encoding frequencies. Electrodes are stimulated depending on the frequency of the signal. Electrodes near the base of the cochlea are stimulated by high frequency signals while electrodes near the apex are stimulated by low frequencies. The signal processor is responsible for splitting the input signal into different frequency bands or channels and delivering the filtered signals to the appropriate electrodes. The DSP function for a cochlear prosthesis primarily consists in dividing an input speech signal into a number of frequency bands (12 to 22) in order to extract the signal energy in each band corresponding to each implanted electrode. The DSP is programmed to fit the stimulation patterns to the function of the inner ear. The most common strategies used in cochlear implants are the Continuous Interleaved Sampling (CIS) and Advanced Combination Encoder (ACE) [24]. Both of these strategies can be achieved using either an analog filter bank approach or an STFT approach.

When using the filter bank approach, a set of bandpass filters is taken to divide the signal into a number of frequency bands or channels. Then, the filtered output would be full wave rectified and passed through low-pass filters to extract channel envelopes which determine the amplitudes of current pulses on corresponding electrodes. Figure 2a illustrates the CIS strategy where the input signal is passed through a bandpass filter bank. The computational complexity of this approach increases as the number of channels and the order of the filters increase. The same output could be achieved by computing the STFT of the input signal, grouping frequency bins into different channels, and then summing up the power of adjacent frequency bins falling in a channel to obtain the signal energy in that channel (Figure 2b).

3. Using WT methodology for CIS speech coding strategy

In this work, our first motivation was to explore an auditory model-based Wavelet Transform approach. The goal of this research was to take advantage of its time frequency resolution for speech signals in order to generate stimuli for cochlear prostheses. The CIS strategy principle can be used when filter responses are represented by Wavelet forms. For this CIS-WT approach, the pre-emphasized input speech signal is arranged into a set of frequency bands to extract the corresponding squared magnitude relative to each channel. These magnitudes are then added to the frequency range of a channel to produce N envelopes. After normalization, each channel envelope is selected for stimulation and compressed using a logarithmic compression map to follow physiological properties of the cochlea. Following this procedure, the output signal in any given channel is compressed and mapped to fit the threshold level (THR) and the patients' most comfortable level (MCL) [25]. The block architecture of the wavelet based CIS strategy is shown in Figure 3.

More precisely, our suggested CIS-WT analysis stage uses a continuous wavelet transform based on a non-linear auditory model that can be applied to a cochlear prosthesis speech processing strategy. This model is the equivalent of a band pass filter bank where respective frequency bands would be designed from the critical bandwidth (CB). In fact, analogous to the human auditory system, our proposed filter bank spectrum design is divided according to the CB model.

3.1. Filter bank for CB model

An appropriate division of the speech frequency domain should be taken into account because such a feature could be considered as an important prerequisite to yield a patient's stimulated comfort as well as an optimal performance with the cochlear prosthesis. Using a bank of nonlinear band pass filters adequately ranging along the considered speech frequency domain makes it possible to simulate the physiological cochlear frequency mapping [26, 27] which is known as the critical bandwidth [28]. The Equivalent Rectangular Bandwidth (ERB) might be regarded as a measure of the CB. This measure is equal to the bandwidth of a perfect rectangular filter, whose pass-band is equal to the maximum transmission of the specified filter. Several mathematical models formulate the ERB as a function of the centre frequency f_0 using moderate sound levels for young people with normal hearing [29, 30].

Among these models the Munich Equivalent Rectangular Bandwidth $\mathrm{ERB_M}$ could be written as a nonlinear formula as

$$\mathrm{ERB}_M = 25 + 75(1 + 1.4 f_0^2)^{0.69}. \quad (1)$$

Another model is a linear formula referring to Cambridge Equivalent Rectangular Bandwidth ERBC,

$$\mathrm{ERB}_C = 24.7(1 + 4.3 f_0), \quad (2)$$

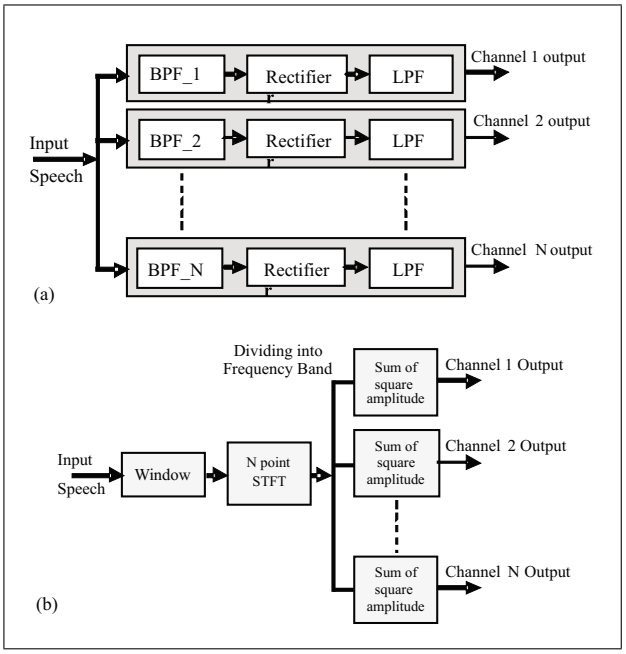

Figure 2. Common analysis stage strategies for cochlear prosthesis based on: (a) Filter-bank approach, and (b) STFT approach.

where f_0 is the central frequency of the band (in kHz).

A comparative study, described in [26], between two coupled filters based successively on approximate functions presented by equations (1) and (2) showed that the Munich Equivalent Rectangular Bandwidth is closer to the critical bands than the Cambridge Equivalent Rectangular Bandwidth. Therefore, it is the Munich formula that the present work adopts. In the present application, a 16 kHz sampling frequency was used so that the considered spectrum was limited to 8 kHz, to be divided adequately according to the adopted critical bands. Consequently, with 21 non-linear bandpass filters designed according to this Munich Equivalent Rectangular Bandwidth model, we could target a cochlear prostheses apparatus that could have up to 21 stimulation channels. Using this ERB model mapped on the normal auditory system, the appropriate filter bank outputs were calculated with their corresponding energy. Such extracted energies serve to encode the stimulating pulse levels to be generated in each of the 21 stimulation channels [26].

3.2. Auditory model based on WT model

Wavelet functions are efficient and powerful tools in signal analysis especially for a time frequency description and a multi-resolution advantage. Such functions can be generated from a single basic wavelet $\psi(t)$, the so-called 'mother wavelet', by scaling and translation,

$$\psi_{a,b}(t) = \frac{1}{\sqrt{a}}\,\psi\big((t-b)/a\big), \quad (3)$$

where the positive factor "a" represents the scale parameter, "b" represents the position parameter and the factor $1/\sqrt{a}$ represents energy normalization across the different

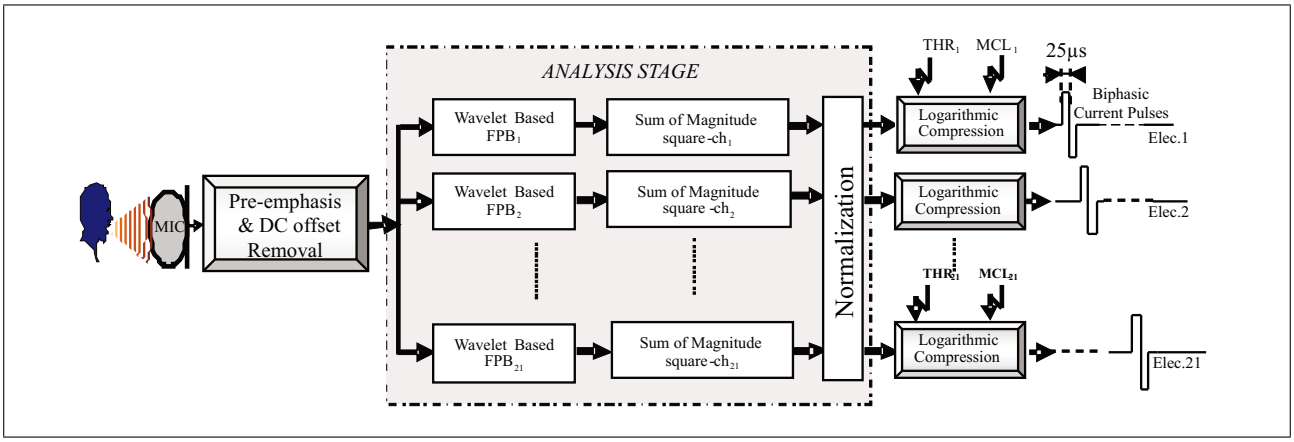

Figure 3. Architecture of our proposed stimulation strategy.

Table I. Number of FFT lines related to each frequency analysis band and its associated center frequencies in Hz.

Analysis frequency band 'z'	1	2	3	4	5	6	7	8	9	10	11
Center frequency f_c [Hz]	50	150	250	350	450	570	700	840	1000	1170	1370
Number of lines N_z	3	3	3	3	4	4	4	5	5	6	7
Starting line $nstart_z$	1	4	7	10	13	17	21	25	30	35	41

Analysis frequency band 'z'	12	13	14	15	16	17	18	19	20	21
Center frequency f_c [Hz]	1600	1850	2150	2500	2900	3400	4000	4800	5800	7000
Number of lines N_z	8	8	11	12	14	18	22	29	35	42
Starting line $nstart_z$	48	56	64	75	87	101	119	141	170	205

scales. Hence, these Wavelets are a class of functions used to locate a given function in both space and scaling. The 'Daughter wavelets' $\psi_{a,b}$ can also be formed by translation "b" and contraction "a" [31],

$$\begin{aligned} \mathrm{WT}_{a,b}(s) &= \frac{1}{\sqrt{a}} \int_{-\infty}^{+\infty} \psi\big((t-b)/a\big) s(t)\,\mathrm{d}t, \\ &= \int_{-\infty}^{+\infty} \psi_{a,b}(t)\, s(t)\,\mathrm{d}t. \end{aligned} \tag{4}$$

The WT model characteristics are suitable for the needs of the multi-scale analysis of one speech signal 's'. That is why we propose to filter this speech signal 's' according to this multi-scale analysis model for cochlear prosthesis external processing [23]. For this adopted WT model, the scale parameter 'a' can be calculated using the central frequency 'f_0' of the appropriate filter band. In this approach, we can use different mother wavelets such as Morlet and Marr wavelet [31]. Morlet wavelet form could be the most convenient one for this processing because of the easy selection of the central frequency, the stable quality factor 'Q' as well as the exponential mathematical formulation. Morlet wavelet can be then defined as

$$\psi(t) = \frac{1}{\sigma\sqrt{2\pi}}\, \mathrm{e}^{-t^2/(2\sigma^2)}\, \mathrm{e}^{-\mathrm{i}2\pi f_0 t}, \tag{5}$$

where f_0 is the central frequency of Morlet wavelet (in kHz). The decay parameter 'σ' in the Morlet wavelet links time and frequency resolutions. The mother wavelet envelope could be described by the term

$$\tilde{\psi}(t) = \frac{1}{\sigma\sqrt{2\pi}}\, \mathrm{e}^{-t^2/(2\sigma^2)}. \tag{6}$$

WT functions were proposed to be explored in this research for the basic idea of mimicking the Auditory Physiological Model according to the ERB frequency distribution. In fact, the temporal envelope of the mother wavelet varies according to the characteristics of the target ERB model by adjusting the decay parameter 'σ'.

The result of the mathematical demonstration to find the relation between decay parameter 'σ' and the band width B in each corresponding frequency is summarized as [26]

$$\sigma = \sqrt{0.7}/(\pi B). \tag{7}$$

Based on the Munich Equivalent Rectangular Bandwidth formula, and using f_0 as central frequency for each filter band, the final expression of the Munich based Morlet wavelet is represented by [26]

$$\begin{aligned} \psi(t) = {} & \sqrt{\frac{\pi}{1.4}}\big(25 + 75(1 + 1.4 f_0^2)^{0.69}\big) \\ & \dot{\mathrm{e}}\left[-\left(25+75(1+1.4 f_0^2)^{0.69}\right)^2 \pi^2 t^2\right]/1.4\, \mathrm{e}^{-\mathrm{i}2\pi f_0 t}. \end{aligned} \tag{8}$$

The above formula is essential to the exploration for our proposed cochlear prosthesis speech processing strategy.

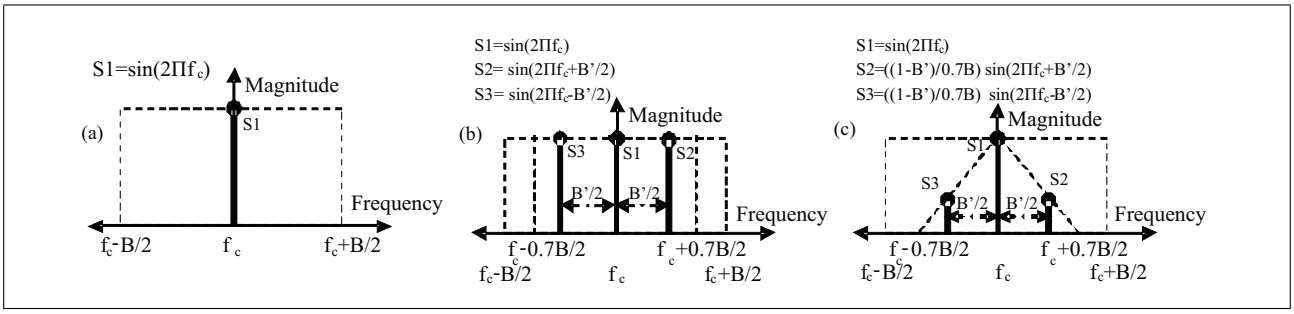

Figure 4. Representation of (a) Input = S1, (b) Input = S1 +S2 +S3 (c) Input = S1 +S2 +S3.

4. Carrying out WT speech processing strategy for cochlear prostheses: Comparative study with STFT approach

This section compares the performance of our proposed WT based speech processing strategy with the conventional STFT strategy. Comparisons were performed for an objective measurement in terms of frequency selectivity and sensitivity (temporal feature) as well as a subjective measurement in terms of intelligibility. The speech signal was initially sampled at 16 kHz. Each frame contained 512 points and lasted 32 ms, leading to 256 spectrum lines.

The bandwidth of the two considered filters, based on STFT and WT respectively, matched the frequency range according to the Munich Equivalent Rectangular Bandwidth model. Table I presents the number of lines (N_z), the starting line ($nstart_z$) assigned to each frequency band 'z' ($z = 1 \ldots 21$) and its associated center frequency (f_c).

This Comparative study was performed at two analysis rates, 250 Hz and 500 Hz. These analysis rates were controlled by adjusting the overlap between the successive analysis windows. The first analysis rate fixed at 250 Hz was realized with a 50% overlap between two successive analysis frames. The second analysis rate, fixed at 500 Hz, was achieved with a 75% overlap.

4.1. Objective measures

4.1.1. Spectral Leakage as a frequency selectivity measure between considered bands: Simulation of channel selectivity

The main function of the speech processing strategy is to divide the spectrum into specific frequency bands relative to the targeted cochlear implant stimulation channels. For these frequency bands, appropriate energy values are calculated in order to subsequently be encoded before they are sent to the stimulation channels. A good measure of performance could be the Leakage Factor 'LF' (spectral leakage) that indicates how much or what percentage of the energy in one band could be leaked into adjacent bands [7].

When exploring the spectral leakage, we conducted a test in which a synthetic signal (e.g., pure tone) of known frequency was used as an input signal. This signal was then filtered within 21 ERB frequency bands using either the WT or the STFT approach. The wavelet mother length and the number of STFT lines are equal to 512. Energies were processed accordingly: LF was evaluated by calculating the ratio between the bands distributed energies, except for the considered band (input tone), over the total sum of energies [7]. Such a measure evaluates how much energy leaked into other bands: The lower LF indicates best performance in terms of providing good frequency specificities. The LF for channel "ch" was calculated using the following equation:

$$\mathrm{LF}_{ch} = 100 \cdot \frac{\sum_{n=1}^{N} \sum_{k\neq ch}^{\text{channels}} \mathrm{out}(n,k)^2}{\sum_{n=1}^{N} \sum_{k=1}^{\text{channels}} \mathrm{out}(n,k)^2}, \tag{9}$$

Where out(n, k) is the analysis channel output of channel k at a time instant n and N is the total number of samples in the test input.

Three different sets of input were used to measure spectral leakage. The first set was a sinusoidal signal with the frequency coinciding with the considered central band frequency as indicated in Figure 4a. The second was the sum of three sinusoidal signals: one signal had a frequency equal to that of the central band; the two other sinusoids were distributed within 70% of the considered bandwidth as shown in Figure 4b. The amplitudes of all the three signals were considered to be of the same value. The third set was similar to the second one, but with the signals' amplitudes varied inversely proportional to the distance from the central frequency as shown in Figure 4c. Different simulations were carried out for the second and third sets for all the 21 ERB frequency bands.

After averaging the LF factor over all simulations for our WT speech processing strategy as well as for the STFT strategy, Figure 5 illustrates the real performance of our proposed strategy where the LF factor was clearly reduced.

4.1.2. Tracking temporal features

Speech processing for hearing-impaired people is a really sophisticated task that requires special attention to the information delivered to the malfunctioning target ear. Trying to understand speech in the presence of background sounds, it becomes noticeable that hearing-impaired people perform more poorly than normal-hearing people when listening in a steady background sound, but also perform considerably more poorly when listening in a fluctuating background sound. Normal- hearing people get a considerable benefit from temporal dips in background sounds,

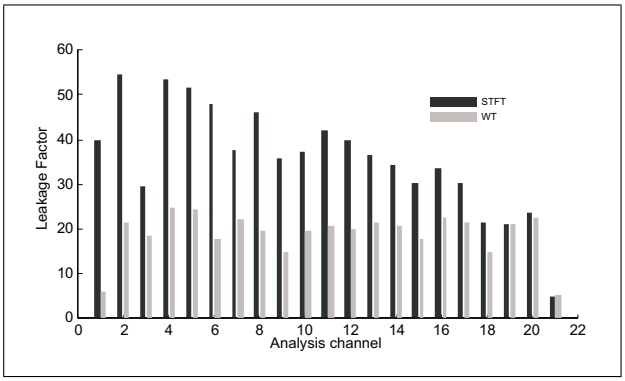

Figure 5. LF Comparison using our WT based strategy and STFT based strategy.

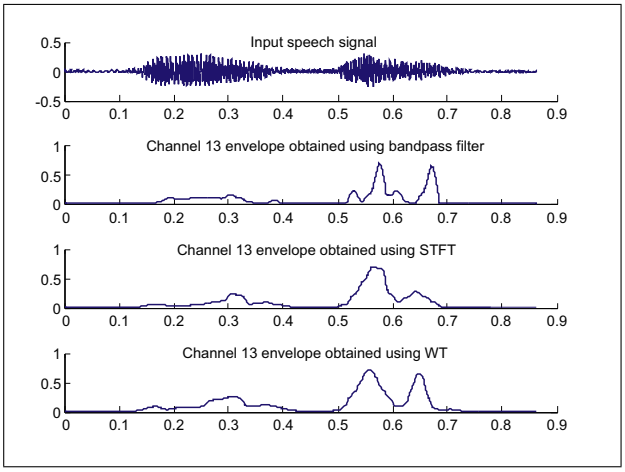

Figure 6. Input speech (syllable "buze"), band pass filter envelope for channel 13, STFT approach envelope for channel 13 and WT approach envelope for channel 13.

but the hearing-impaired seem to have little or no ability to "listen in the dips" to enhance speech perception, even when appropriate amplification could be provided to ensure that the sounds in the dips are above the absolute threshold [7]. A reduction in the ability to resolve temporal features of complex sounds could be a contributing factor to such a difficulty. Here, we tried to show that a reduced ability to process the sound's temporal features plays an important role. In fact, temporal features could retain all the information that could distinguish a target sound from a background sound.

Generally, speech sounds could be in fact processed by filter banks carefully distributed onto N adjacent frequency bands. For each specific band, the filtered signal would be processed in order to extract energies and to sometimes preserve the temporal envelope [24]. From our experimental tests (reported in Figure 6), an illustration of a sample envelope output of channel 13 is presented. In this test, the syllable "buze" spoken by a male speaker taken from the Fournier database [32] was used. The extracted envelopes using the bandpass filter, STFT and WT strategies are superimposed for comparison. As can be seen, the wavelet-based strategy performs on par with the other strategies as far as being able to track temporal envelope features. Using WT strategy will, therefore, address the limitations of the STFT approach which has excellent frequency resolution, but limited time resolution. This is due to the STFT has a fixed time-frequency window, making it inaccurate to analyze signals but the wavelet transform (WT) has a flexible time-frequency window. In fact, our preliminary results indicate that the WT based strategy has a better compromise between time and frequency resolutions. This time-frequency compromise seems important to analyze the fast phonetic variations of the speech signal and improves its intelligibility.

Similar experimental tests were performed for all the considered channels; thus, validating our objective measures. In the following section, we will present the subjective experimental tests in order to complete and validate our WT approach.

4.2. Subjective measures

In this section, two speech processing strategies based on the WT and the STFT dedicated to cochlear implant are evaluated using subjective listening tests. Subjective measures can be of great importance as they help in assessing the WT speech processing strategy for the cochlear prosthesis apparatus. Access to such intelligibility measures could be of high significance as they can be used to guide the development of new speech encoding algorithms for cochlear prostheses [33].

4.2.1. Listeners

Performance evaluation was carried out with a population of fifty normal hearing subjects. Their audiograms were tested in the ORL department prior to the experiment. All participants had normal-hearing thresholds (pure-tone air conducted thresholds between 250 and 8000 Hz $<$ 20 dB HL) without history of auditory pathology. Normal-hearing subjects' ages ranged between 17 and 35 years. All participants were native speakers of French. Tests were performed in the anechoic room of the Edouard-Herriot hospital of Lyon, France.

4.2.2. Phonetic material

The phonetic material used is the French Fournier set [32], which contains thirty nine lists and each of the lists is made up of 10 disyllabic words pronounced by a single male talker. This is a commonly used speech stimulus for intelligibility assessment in French. The sound level was calibrated to 70 dB SPL. All these lists are processed in the anechoic room of the ORL department of the Edouard-Herriot Hospital of Lyon-France in a silent environment first and then in a noisy environment in the presence of an additive babble noise. The signal-to-noise ratio (SNR) varied from 6 dB to -6 dB with 6 dB steps.

4.2.3. Cochlear prosthesis simulator

In order to investigate the behaviour of the above speech processing strategy in the case of cochlear implant, vocoder simulations were used. It has been pointed out by many

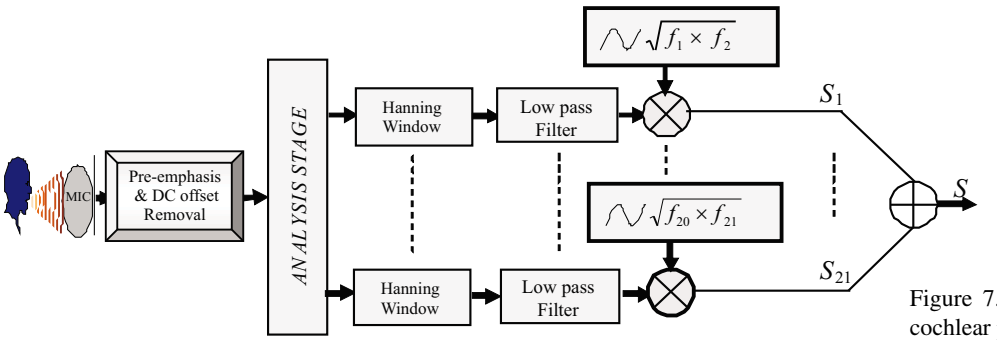

Figure 7. Block diagram of the cochlear prosthesis simulator.

studies (e.g. [34]) that these simulations provide results consistent with the outcome of cochlear implants and that vocoded speech signals can be presented to normal hearing listeners in the absence of confounding factors associated with cochlear implants.

Speech signals were digitized at 16 kHz sampling frequency with 8 kHz speech bandwidth. These signals were then filtered using either the WT approach (with two values for position parameters "b=8" and "b=16" respectively) or the STFT approach. The cut-off frequencies and the technical details regarding stimulus generation are provided in the last section describing the ERB model. The use of these cut-off frequencies also ensured that the amount of spectral information provided by the envelope stimuli was similar for both normal-hearing and hearing-impaired subjects.

These twentyone envelopes were then processed using a Hanning window followed by a specific low-pass filtering. Each filtered envelope could be then modulated in amplitude using one specific sine wave. This wave is taken with a frequency equal to the geometric mean of the upper and lower frequency band limits and with a random starting phase. Then, the twentyone stimuli were added to obtain one single stimulus that contained the envelope information. The presentations of these details are given in Figure 7.

Our stimuli were first presented to normal-hearing subjects. Then, a comparison was performed between the WT approach and the STFT approach taking into consideration all the pre-established conditions – Algorithms, SNR and analysis rates.

4.2.4. Listening session procedure

The listening tests were conducted using a personal computer connected to a CD player (PHILIPS-CD723). An audiometer (MADSEN-Orbiter 922) was used for the calibration and intensity level adjustments. Stimuli were presented to the subjects using closed professional 'Sennheiser' HD250 linear headphones at a comfortable level calibrated to 70 dB SPL. Prior to the formal testing, a training session containing 10 random words was administered in order to familiarize each subject with the stimuli. No scores are calculated in the training session. Following the practice session, the subjects were tested in the various experimental conditions. During the testing session, the subjects were instructed to repeat the words they heard. In total, there were 24 testing conditions involving three different SNR levels (−6, 0 and 6 dB), two analysis rates (250 Hz and 500 Hz) and three different algorithms (STFT, and WT with two different position parameters "b"). At the end of each listening session, the responses of each subject were collected, stored and scored off-line with the percentage of correctly repeated syllables.

4.2.5. Results and Discussions

For each speech analysis algorithm, the mean scores of syllables recognition across all subjects, taken in quiet (No Added Noise: NAN) or noisy environments with all SNRs (6 dB, 0 dB and −6 dB) were calculated.

As depicted in Figure 8, for each SNR, a group of three bars is presented. The first bar indicates the mean percent score obtained when processing speech by STFT. The second and third bars stand for mean percent scores obtained when processing speech by the WT strategy with their two values for position parameters (b=8 "WT8" and b=16 "WT16"). Figure 8 reports the mean percent scores of syllables recognition obtained with the three algorithms (WT8, WT16 and STFT) for perception in a quiet "NAN" and noisy environments with the SNRs considered (6 dB, 0 dB and −6 dB) with the two analysis rates 250 Hz and 500 Hz, respectively.

As shown in this figure, results indicate that the score of syllables recognition depended on both the speech analysis algorithm and the analysis rate in quiet and in noisy environments at different SNRs. In fact, results of normal-hearing subjects indicated that the WT based speech processing strategy was constantly superior to STFT for syllables perception. For syllables recognition in a quiet environment, the WT strategy achieved a higher correct rate of perception of about 80% as compared to a 60%–70% of success rate with STFT strategy at different analysis rates. Similarly, syllables recognition score was strongly affected by the SBR. Indeed, in a noisy environment, the WT strategy performed better than STFT. Overall SNR levels, the average recognition score for the WT16 strategy was 14% better than the STFT based strategy.

In order to assess the performance of the above speech processing strategy, Chi2 tests with the following parame-

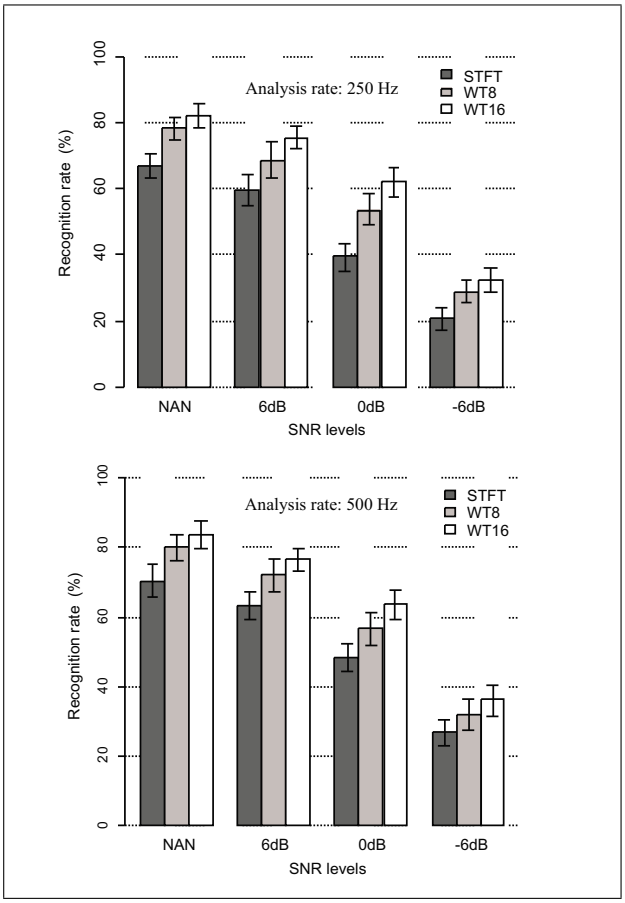

Figure 8. Mean recognition score in quiet and in noisy environments at different SNRs for two stimulation rates.

ters were performed to examine the effects of the considered factors:

- Repeated measures (the same subjects underwent all the situations),
- Dependent variable (Recognition score in percentage),
- Three factors:
  - Speech analysis algorithm (WT8, WT16 and STFT),
  - Analysis rate (250 Hz and 500 Hz),
  - SNR (NAN, 6 dB, 0 dB and -6 dB, this last factor was taken at random; the first two factors were fixed).

This model, with both fixed and random effects, is known as the mixed-effect model, and is a standard for data analysis [35]. We fit our models using the 'lmer' program of the 'lme4' package [36] of the statistical programming language and software environment [37].

The analysis indicated that there were significant effects of the speech analysis algorithm (Chi2[2] = 21, $p \leq 0.001$) and the SNR (Chi2[3] = 167, $p \leq 0.001$), but non-significant effect of the analysis rates (Chi2[1] = 1.7, p = 0.2) could be noticed. However, there was non-significant interaction between the considered speech analysis algorithms and the SNR (Chi2[6] = 0.54, p = 1). Similarly, non-significant interactions between SNR and analysis rates (Chi2[3] = 0.08, p = 0.9) and between the analysis rates and the speech analysis algorithm (Chi2[2] = 0.28, p = 0.8) were revealed. Chi-2 tests confirmed the existence of a main effect of speech analysis algorithm on speech recognition.

Post hoc comparisons were run to explore the differences obtained with the considered speech analysis algorithm at different SNR levels. We used the Tukey HSD test via the general linear hypothesis test 'glht' function from multiple comparison 'multcomp' package of R.

In quiet (NAN), syllables recognition was affected by the speech analysis algorithm. Post hoc comparison indicated that syllables recognition performance improved significantly using WT16 over STFT ($p < 0.005$) and the improvement was about 14%. When WT8 speech analysis algorithm was considered, syllables recognition improvement was limited to 10% and this difference was non-significant ($p > 0.005$).

When SNR was 6 dB, a non-significant improvement of syllables recognition was observed when WT16 speech analysis algorithm was considered ($p > 0.005$). In fact, the improvements in syllables recognition were about 13% and 8% when WT16 and WT8 were respectively adopted.

When signal and noise levels were identical, it can be noted that WT16 and WT8 performed significantly better than STFT ($p < 0.005$) and the improvement was about 19% and 8% respectively.

WT16 and WT8 speech analysis algorithm were also better than STFT at −6 dB and improvements were about 10% and 6% respectively. These improvements were non-significant ($p > 0.005$).

To sum up, our results show a loss of the ability of Normal-hearing subjects with STFT algorithm for speech perception. These results are in accordance with the previous studies performed with objective measures. We could then confirm that the WT16 based speech processing strategy for cochlear prostheses presents better sensitivity, frequency selectivity and intelligibility than the STFT and WT8 based strategies. This is justified by the fact that WT provides a time-frequency representation with a better resolution than the STFT.

This research on cochlear prosthesis speech processing was then extended to explore the real time DSP implementation for the cochlear prosthesis external processing.

5. WT speech processing DSP implementation results

In this section, the task is mainly a proof of concretization and validation of the proposed methodology. Taking into consideration the fact that a cochlear prosthesis is driven by an external digital processor, usually a DSP, an attempt at real-time implementation of the proposed WT based approach is therefore sought in this section. The real-time implementation of this cochlear prosthesis stimulation strategy has been evaluated for various signal forms. The algorithm was implemented on a commercial DSP (TMS320C6416 development board) appropriate to perform our algorithms and to mimic the cochlear prostheses external processor. Using several optimization techniques as discussed in [38], we were able to achieve significant speedups and reduce the processing time to about

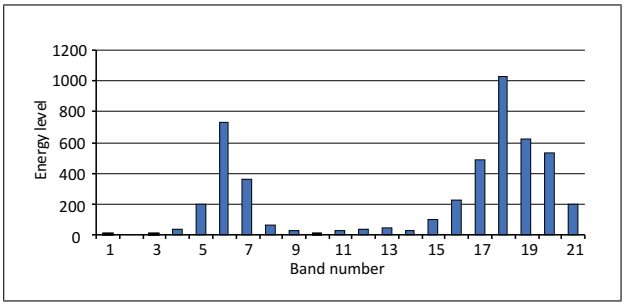

Figure 9. Energy level generated in real time relative to 550 +4000 Hz composite harmonic.

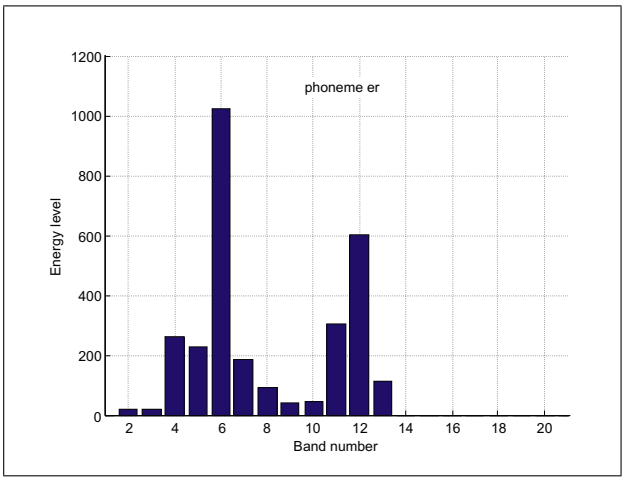

Figure 10. Energy level in Matlab simulation relative to the 'er' phoneme.

10.5 µs which is sufficient for real-time cochlear stimulation. Additional implementation details are described in the Appendix.

The validation of the real-time implementation was carried out using synthetic harmonic signals and complex tone signals extracted from the TIMIT database. These signals are generated with a Matlab-based software to be received by the AIC23 CODEC through the PC sound card.

Figure 9 presents the results of parameterization of a composite harmonic 550 Hz + 4000 Hz in real-time. We have noticed that maximum energies are reported at the sixth channel for the harmonic 550 Hz and at the eighteenth channel for the harmonic 4000 Hz. This result indicates that synthetic harmonic signals' energies are located in their corresponding frequency bands as shown in Table I.

In the second experiment, we have used the phoneme 'er' that was extracted from the TIMIT database. Results of our analysis algorithm are illustrated in Figures 10 and 11. It is noticeable that maximum output energies are located in the sixth and twelfth bands for both Matlab (Figure 10) and real-time implementation (Figure 11). Moreover, the other channel's outputs keep the same frequency positions compared with Matlab results. The real-time implementation has therefore been successfully tested, given

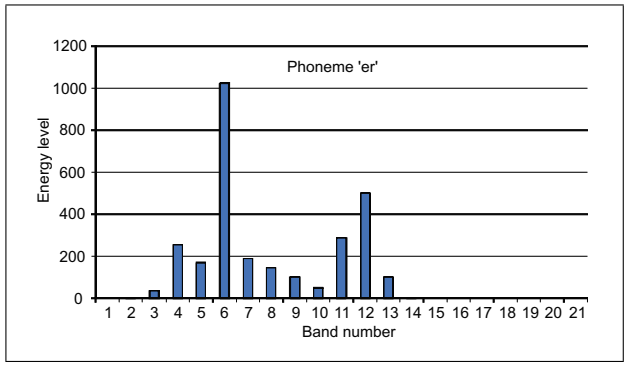

Figure 11. Energy level in real-time implementation relative to the 'er' phoneme.

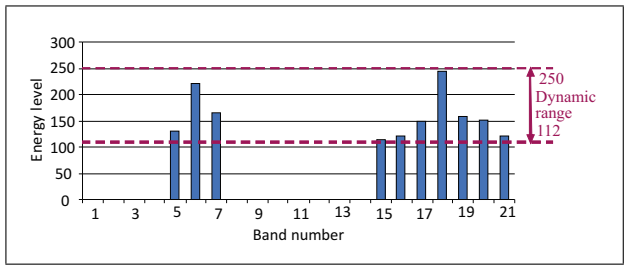

Figure 12. Compressed energy level generated in real-time relative to 550+4000 Hz composite harmonic.

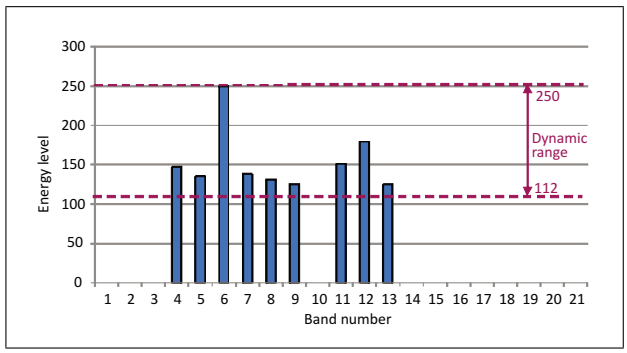

Figure 13. Compressed energy level generated in real-time relative to the "er" phoneme.

the fact that its test results are similar to those of the Matlab-based implementation.

To adapt the dynamic range of channel outputs to the acoustic dynamic range of each patient, a logarithmic compression is required. For additional information, see equation (A4) in the Appendix.

In the present study, the value of p = 0.0001 was used to obtain Log compression. Moreover, we took one random example of [THR, MCL] map, that varies between 112 and 250 for the active electrodes. The value 100 was chosen as an acoustic minimum (Ac_{Min}). This means that all energies lower than 100 are set to zero.

Figures 12 and 13 illustrate the computed energy levels after logarithmic compression of the composite harmonic 550 Hz + 4000 Hz and the phoneme 'er', respectively.

In Figure 12, only the energy bands 5–7 and 15–21 are compressed, all others are cancelled as their values are

less than the acoustic minimum ($Ac_{Min} = 100$). Similarly, Figure 13 retains energy bands 4–9 and 11–13 which are greater than 100 and cancels all other bands.

After applying the logarithmic compression, we obtained a successful validation test. Indeed, all the proposed strategy outputs ranged within the patient's implant dynamic range.

The experimental results of our approach based on WT tools for speech analysis and encoding dedicated to cochlear prostheses devices were satisfactory, specifically when the energy outputs of the synthetic harmonic and the 'er' phoneme are located in their corresponding frequency bands.

6. Conclusion

The presented research in our paper demonstrated that WT tool could be a successful procedure for analyzing speech signals that would be transformed into electrical stimulating impulses. A novel strategy for cochlear electrical stimulation based on a nonlinear human auditory physiological model has been then developed.

The most distinguished characteristic of the WT tool was its efficient time–frequency domain resolution. Such an advantageous feature for cochlear stimulation could be adaptively adjusted not only by its position parameter that arranges frequency division but also by its scale parameter that could adjust the bandpass width.

The human auditory physiological model was used to mimic the frequency mapping of the cochlea. The ERB model was mainly used to adapt the filter bank to the cochlear tonotopy. According this model, our conceived filter bank structure was divided into 21 non-linear bandpass filters.

This research involved a study that compared the performance of our proposed WT speech processing tool and the classical Short Time Fourier Transform STFT speech processing tool. This comparison proved our approach's superiority in terms of intelligibility, frequency selectivity (spectral leakage) and sensitivity (tracking temporal feature) that are highly relevant for cochlea stimulation.

To be more concrete and in order to validate our approach, a real-time implementation on a Digital Signal Processor was also performed and tested. Using a 720 MHz TMS320C6416 DSP board, the real time execution of our implemented WT based approach was essentially tested using synthetic harmonic signals. Energy levels were successfully generated in real time and in the right corresponding position. Other specific adjustments could also be added in order to improve comfort during stimulation such as the logarithmic compression.

Appendix

Real-time implementation of wavelet transform

For any real-time implementation of a signal-processing algorithm, several factors need to be considered such as the DSP's architecture and properties. Algorithm length and structure should equally be considered. In fact, the complexity of the algorithm needs to be reduced to minimize the execution time for the real-time application.

The DSP platform chosen in this experimental simulation was the TMS320C6416. It uses VelociTI, high-performance Very Long Instruction Word (VLIW) architecture. The C6416 register file extends this by additionally supporting packed 8-bit types and 64-bit fixed point data types. Such a fixed point DSP was conceived with important properties that make it user-friendly and suitable for various applications [39, 40, 41].

As formulated by equation (4), each family of wavelets $\psi_{a,b}(t)$ is multiplied by an input data frame $s(t)$. This formula can be expressed in discrete time as

$$\mathrm{WT}_{a,b}(s) = \frac{1}{\sqrt{a}} \sum_{k=0}^{N-1} \psi\big((k-b)/a\big)s(k). \qquad \text{(A1)}$$

In fact, to process the WT approach in discrete time, the coefficient of the mother wavelet for each shifting "b" parameter is multiplied by an input data frame $s(k)$ (with k indicating the time index). This process could be repeated for each frequency "a" parameter ($a = 1 \ldots 21$).

However, this process was not found to be appropriate for a real-time processing since the input data frames are continuously generated on the recorded speech signal. Therefore, it was decided to leave the wavelet function fixed and to shift the input data frame according to each "b" parameter [42]. This processing method yielded the same results and is more suitable for a real-time WT implementation. The WT of an input speech signal $s(k)$ can be represented by

$$\mathrm{WT}_{a,b}(s) = \frac{1}{\sqrt{a}} \sum_{k=0}^{N-1} \psi(k/a)s(k-b), \qquad \text{(A2)}$$

where N is the mother wavelet length ($N = 512$) and b is the shifting parameter of the input speech signal.

WT process, as given in equation (11), could be expressed as a Finite Impulsion Response (FIR) Filter given in the following equation,

$$Y_a(b) = \sum_{k=0}^{N-1} \psi_a(k)s(k-b), \qquad \text{(A3)}$$

where $\psi_a(k) = \dfrac{1}{\sqrt{a}}\psi(k/a), \quad a = 1 \ldots 21.$

Coefficients of the mother wavelet could be calculated for each parameter "a" using Matlab software according to equation (8). All the computed values could subsequently be stored in a $(21 \times N)$ array in a header file included in the source file. Computing the WT outputs within a short duration window, with the input window being shifted by one sample at a time ($b = 1$), could allow the analysis rate to be equal to the input sampling rate. This has a positive effect for real time execution [43].

Figure A1 illustrates the DSP implementation of the considered WT based algorithm. In our implementation,

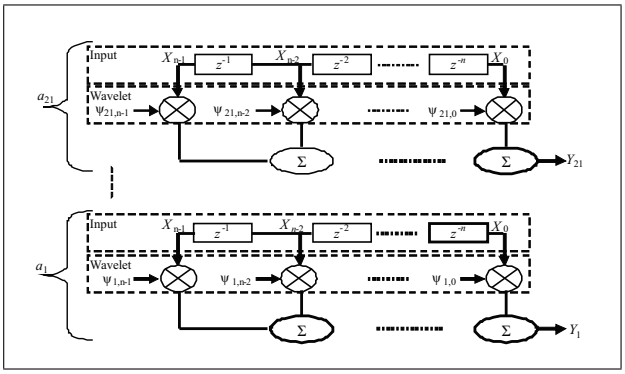

Figure A1. WT processing based implementation algorithm.

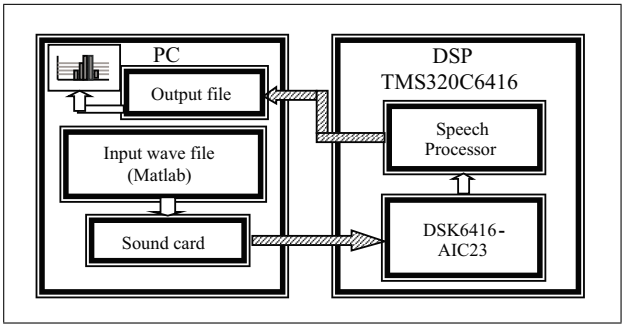

Figure A2. Experimental platform used to test the real-time speech processing.

we used a single 512-input sample window (X1…X512). For each input sample, placed at its specified level in the input window, we computed 21 wavelet transform outputs (Y1…Y21).

In real time, the stimulation rate could be maintained between 250 and 2400 pulses per second 'p/s' per channel depending on the patient's MAP [24, 44]. For the worst of the cases (2400 p/s), we could derive the minimum biphasic stimulation pulse that could be equal to $1/(2400N_a) = 20\,\mu s$, where $N_a = 21$ channels. Hence, we note that the processing time for a real time implementation must be less than 20 µs.

Using several optimization techniques as discussed in [38], we could achieve significant speedups and reduce the processing time to about 10.5 µs. We can note, then, that the reduced processing time was sufficient for real-time DSP processing. This processing time represents only about 52% of the minimum biphasic stimulation pulse width (20 µs). It is worth pointing out that the timing indicated includes the decomposition, sum of magnitude square, normalization and logarithmic compression for the implementation of 16-bit fixed-point with 21 channels analysis.

A specific platform based on the TMS320C6416 processor was successfully used to allow such experiment and real-time implementation (see Figure A2).

This platform consists of two essential parts: the PC and the chosen TMS320C6416 DSP board.

The speech input signal as taken from the input wave file was transferred to the sound card that was connected directly to the AIC23 codec. Then, the real-time speech processing by the DSP computed energies for each channel and sent them to the PC board as an output file.

During experimentation, the PC sound card could be selected as the audio signal source and we could set the sampling rate to 16 kHz. To sample the audio signals, one of the multichannel buffered serial ports (McBSPs) was configured to connect the AIC23 CODEC. To receive the audio input data from the AIC23 CODEC continuously, an interrupt service routine (ISR) was executed when one sample input was received. Thus, the data sampling and the necessary processing could be done simultaneously and no incoming signals could be missed.

A logarithmic compression is needed in order to comfortable adjust energy levels that would be transformed into electrical impulses. This transformation is necessary because the range of acoustic amplitudes in conversational speech is considerably larger than the patient implant dynamic range [41, 45]. The acoustic-to-electrical mapping could be written as

$$Y(n) = A\left[X(n)\right]^p + B, \quad \text{where } p < 1, \tag{A4}$$

and where $Y(n)$ indicates the electrical amplitude output (measured in clinical units or microamperes) at time n, $X(n)$ denotes the acoustic envelope amplitude, and the constants A and B are used to ensure that the acoustic amplitudes are mapped within the electrical dynamic range [THR, MCL] [25]. The power exponent 'p' was used in the present study to control the steepness of the compression function.

Acknowledgement
This study was funded by Egide (PHC Utique code: 09G 1421) and CNRS-DGRST (code: 09/R11-18) projects for French- Tunisian collaboration. We appreciate the efforts of all the subjects and we thank the normal-hearing listeners for their patience and persistence with these experiments. Authors are also grateful to Professor Moneim Ghorbel, head of ORL service CHU of Sfax Tunisia, and to Professor Christian Berger-Vachon from Lyon Neuroscience center, Claude Bernard University, Lyon-France, for their assistance in upgrading the manuscript and to Mr. Ali M. Amri from Higher Institute of Electronic and Communication of Sfax Tunisia, for his revision of the paper's English.

References

[1] K. Rouiha, D. Bachir, B. Ali: Analysis of speech processing strategies in cochlear implants. Journal of Computer Science **4** (2008) 372–374.

[2] P. C. Loizou, G. Stickney, L. Mishra, P. Assmann: Comparison of speech processing strategies used in clarion implant processor. Ear and Hearing **24** (2003) 12–19.

[3] F. Chen, Y. T. Zhang: A novel temporal fine structure-based synthesis model for cochlear implant. Elsevier Signal Processing **88** (2008) 2693–2699.

[4] D. B. Grayden, A. N. Burkitt, O. P. Kenny, J. C. Clarey, A. G. Paolini, G. M. Clark: A cochlear implant speech processing strategy based on an auditory model. IEEE ISSNIP (2004) 491–496.

[5] B. H. Bonham, L. M. Litvak: Current focusing and steering: Modeling, physiology and psychophysics. LM Hearing Research, Elsiever **242** (2008) 141–153.

[6] V. Gopalakrishna, N. Kehtarnavaz, P. C. Loizou: A recursive wavelet-based strategy for real-time cochlear implant speech processing on PDA platforms. IEEE transactions on Biomedical Engineering **57** (2010) 2053–2063.

[7] S. G. Mallat: A theory of multiresolution signal decomposition: the wavelet representation. IEEE PAMI **11** (1989) 674–694.

[8] A. F. Pérez-Rendón, R. Rafael: The convolution theorem for the continuous wavelet transform. Elsevier, Signal Processing **84** (2004) 55–67.

[9] S. Chand: SC- and SS-wavelet transform. Elsevier, Signal Processing **89** (2009) 305–313.

[10] X. Yang, K. Wang, S. A. Shamma: Auditory representations of acoustic signals. Information theory on IEEE Trans. **38** (1992) 824–839.

[11] R. Meddis, L. P. O'Mard, E. A. Lopez-Poveda: A computational algorithm for computing nonlinear auditory frequency selectivity. J. Acoust. Soc. Am. **109** (2001) 2852–2861.

[12] J. Yao, Y. T. Zhang: The application of bionic wavelet transform to speech signal processing in cochlear implants using neural network simulation. IEEE Trans. Biomed. Eng. **49** (2002) 1299–1309.

[13] I. Cheikhrouhou, R. Ben Atitallah, K. Ouni, B. H. Ahmed, M. Nouri, E. Noureddine: Speech analysis using wavelet transform dedicated to cochlear prosthesis stimulation strategy. 1st International Symposium on Communications, Control and Signal Processing, 2004, 639–642.

[14] K. Nie, N. Lan, S. Gao: Implementation of the CIS speech processing strategy for cochlear implants by using wavelet transform. ICSP **2** (1998) 1395–1398.

[15] A. Paglialonga, G. Tognola, F. Sibella, M. Parazzini, P. Ravazzani, F. Grandori, G. Baselli: Influence of cochlear implant-like operating conditions on wavelet speech processing. Comput. Biol. Med. **38** (2008) 799–804.

[16] A. Derbel, F. Kalel, A. B. Hamida, M. Samet: Wavelet filtering based on mellin transform dedicated to cochlear prostheses. IEEE EMBS (2007) 1900–1903.

[17] J. Yao, Y. T. Zhang: Bionic wavelet transform: a new time frequency method based on an auditory model. IEEE Trans. Biomed. Eng. **48** (2001) 856–863.

[18] G. Tian, Y. Shuli, Y. Datian: Application of wavelet in speech processing of cochlear implant. IEEE EMBC (2005) 5339–5342.

[19] P. C. Loizou: Mimicking the human ear: An overview of signal processing strategies for converting sound into electrical signals in cochlear implants. IEEE signal processing magazine **15** (1998) 101–130.

[20] T. Herzke, V. Hohmann: Improved numerical methods for gammatone filterbank analysis and synthesis. Acta Acustica united with Acustica **93** (2007) 498–500.

[21] M. Unoki, M. Akagi: A method of signal extraction from noisy signal based on auditory scene analysis. Speech Communication **27** (1999) 261–279.

[22] S. Strahl, A. Mertins: Analysis and design of gammatone signal models. J. Acoust. Soc. Am. **126** (2009) 2379–2389.

[23] N. Tönder, R. Hartmann, R. Klinke: A versatile system for the generation and the development of speech coding strategies in cochlear implants. IEEE Transactions on Biomedical Engineering **45** (1998) 773–782.

[24] C. Lorenzi, G. Gilbert, H. Carn, S. Garnier, B. C. J. Moore: Speech perception problems of the hearing impaired reflect inability to use temporal fine structure. PNAS **103** (2006) 18866–18869.

[25] P. C. Loizou: Speech processing in vocoder-centric cochlear implants. – In: Cochlear and Brainstem Implants, Adv Otorhinolaryngol. Basel, Karger **64** (2006) 109–143.

[26] A. Derbel, F. Kalel, A. B. Hamida, N. Ellouze: Wavelet-based parameterization of speech signal dedicated to cochlear prosthesis. Fourth International Multi-Conference on Systems, Signals and Devices, 2007.

[27] W. Nogueira, A. Katai, T. Harczos, F. Klefenz, A. Buechner, B. Edler: An auditory model based strategy for cochlear strategy. IEEE EMBS (2007) 4127–4130.

[28] E. Zwicker, E. Terhardt: Analytical expressions for critical band rate and critical bandwidth as a function of frequency. J. Acoustic. Soc. Am. **68** (1980) 975–982.

[29] W. M. Hartmann: Signal, sound and sensation. Springer-Verlag, 333 Meadowlands Parkway, Secaucus, NJ, 07094, 1997.

[30] B. C. J. Moore, B. R. Glasberg: Suggested formulae for calculating auditory-filter bandwidths and excitation patterns. J. Acoust. Soc. Am. **74** (1983) 750–753.

[31] I. Daubechies: Ten lectures on wavelets. Rutgers University and AT&T Bell Laboratories, CBMS-NSF, regional conference Series In Applied Mathematic, 1992.

[32] J. E. Fournier: Audiométrie vocale. Maloine, Paris, 1951.

[33] F. Chen, P. C. Loizou: Predicting the intelligibility of vocoded speech. Ear and hearing **32** (2011) 331–338.

[34] N. A. Whitmal, S. F. Poissant, R. L. Freyman, K. S. Helfer: Speech intelligibility in cochlear implant simulations: Effects of carrier type, interfering noise, and subject experience. J. Acoust. Soc. Am. **122** (2007) 2376–2388.

[35] R. H. Baayen: Analyzing linguistic data: A practical introduction to statistics. Cambridge University Press, 2008.

[36] D. Bates: Linear mixed model implementation in lme4. Department of statistics, University of Wisconsin-Madison, 2007.

[37] The R development core team, R: A language and environment for statistical computing. 2008.

[38] A. Derbel, M. Ghorbel, M. Samet, A. Ben Hamida, J. Tomas: Cochlear implant stimulation strategy based on wavelet transform: Toward DSP real time implementation. New Circuits and Systems Conference (NEWCAS), IEEE 9th International, Bordeaux, 2011, 117–121.

[39] Texas Instruments: TMS320C64x technical overview. Document technique: spru395b, 2001.

[40] E. Jennifer, B. Jeff: The evolution of DSP processors. Berkeley Design Technology Inc. BDTI, Insight • Analysis • Advice on Signal Processing Technology, 2000.

[41] R. Stan, R. Mike: Embedded edge. A Texas Instruments publication **2** (2001).

[42] A. Derbel, F. Kallel, M. Samet, A. Ben Hamida: Implementation of strategy based on auditory model based wavelet transform speech processing on DSP dedicated to cochlear prosthesis. 4th International Symposium on Computational Intelligence and Intelligent Informatics, 2009, 143–148.

[43] T. A. Morbiwala, M. Svirsky, M. El-Sharkway, M. Rizkalla: A PC-based processor for cochlear implant fitting that can be adjusted in real time. IEEE 48th International Midwest Symposium on Circuits and Systems (MWSCAS), 2005, 1310–1313.

[44] A. E. Vandali, L. A. Whitford, K. L. Plant, G. M. Clark: Speech perception as a function of electrical stimulation rate using the Nucleus 24 cochlear implant system. Ear Hear **21** (2000) 608–624.

[45] F. G. Zeng, G. Grant, J. Niparko, J. Galvin, R. Shannon, J. Opie: Speech dynamic range and its effect on cochlear implant performance. J. Acoustic. Soc. Am. **111** (2002) 377–386.