Multi-band spectral subtraction algorithm dedicated for noise reduction in cochlear implants

Fathi KALLEL, Ahmed BEN HAMIDA

Department of Electrical Engineering, University of Sfax, Tunisia, E-mail: fathikallel@yahoo.fr
Department of Electrical Engineering,University of Gabes, Tunisia, E-mail: ahmed.benhamidag@gmail.com

Christian BERGER-VACHON
Department of Electrical Engineering, University of Sfax, Tunisia, E-mail: christian.berger-vachon@univ-lyon1.fr

Abstract: *In this work, a new multi-band spectral subtraction based noise reduction algorithm is proposed for cochlear implants (CI). The enhanced speech signal is estimated on each stimulation channel. For performance evaluation, some objective speech assessment tests relying on Perceptual Evaluation of Speech Quality (PESQ) score and speech Itakura-Saito (IS) distortion measurement were performed to choose the best noise estimation algorithm. In order to evaluate the speech intelligibility, subjective listening tests were assessed with 50 normal hearing listeners using a specific CI simulator and three cochlear implantees. Experimental results, obtained using French Lafon database corrupted by an additive babble noise and speech-shaped noise at different Signal-to-Noise Ratios (SNR), showed that proposed multi-band spectral subtraction algorithm produces significant improvements to speech recognition compared to the subjects' daily strategy.*

Key words: *cochlear implants , multi-band spectral subtraction, noise reduction.*

1. Introduction

Since the introduction of CI in the 1970s, implant users' speech perception in quiet has improved to the stage where many previously deaf patients can converse confidently over the telephone. These improvements have been largely produced by two major developments: the introduction of multichannel implants, and improvements in speech-processing strategies (Wilson et al. 1991). One major remaining challenge is to improve speech perception in noisy situations, where even the most successful CI users experience great difficulty. Although many cochlear implant (CI) users are capable of high degrees of speech understanding in quiet listening conditions, speech recognition falls sharply in the presence of background noise or competing speakers (Kiefer et al., 1997; Müller-Deile et al., 1995; Fetterman and Domico, 2002).

To reduce the effects of background noise, some single-microphone noise-reduction algorithms originally developed for normal hearing persons have been applied to CI speech processing (Yang and Fu, 2005; Loizou et al., 2005; Loizou, 2006; Van Hoesel and Clark, 1995; Wouters and Vanden Berghe, 2001; Müller-Deile et al., 1995). These algorithms were able to somewhat improve CI users' performance in noisy listening conditions. In general, single-microphone noise reduction algorithms are more desirable and cosmetically more appealing than the algorithms based on multiple-microphone inputs. A few single-microphone noise-reduction strategies (Weiss, 1993; Yang and Fu, 2005) have been proposed for cochlear implants, some of which were implemented on old cochlear implant processors based on feature extraction strategies (F0/F1/F2 and MPEAK strategies) and some of which were implemented on the latest processors. (Weiss, 1993) demonstrated that pre-processing the signal with a standard noise reduction algorithm could reduce the errors in formant extraction. The latest speech processors, however, are not based on feature extraction strategies but are based on vocoder-type strategies. Yang and Fu 2005 (Yang and Fu, 2005) evaluated a spectral-subtractive algorithm using the latest implant processors. Significant benefits in sentence recognition were observed for all subjects with the spectral subtractive algorithm, particularly for speech embedded in speech-shaped noise.

The preprocessing approach to noise reduction, however, has three main drawbacks. Firstly, preprocessing algorithms sometimes introduce unwanted distortion in the signal. Secondly, some algorithms like subspace algorithms and Wiener filter, are computationally complex and subsequently power hungry and do not integrate well with existing CI strategies. Finally, it is not easy to optimize the operation of a particular algorithm to individual users. Ideally, noise reduction algorithms should be easy to implement and be integrated into existing coding strategies.

Hu et al., (2007) proposed a noise reduction algorithm for cochlear implants that applies attenuation to the noisy envelopes inversely proportional to the estimated signal-to-noise ratio (SNR) in each channel. This algorithm could be easy integrated in existing coding strategies. The performance of the proposed noise reduction algorithm is evaluated with nine Clarion CII cochlear implant patients using IEEE sentences embedded in multi-talker babble and speech-shaped noise at 0–10 dB SNR. Results indicate that the the sigmoidal-shaped weighting function produces significant improvements to speech recognition compared to the subjects' daily strategy. Much of the success of the proposed noise reduction algorithm is attributed to the improved temporal envelope contrast.

Kasturi et Loizou (2007), proposed an acoustic-to-electric mapping function for cochlear implant users in noisy environments based on s-shaped mapping function. The proposed algorithm was expansive for low input levels up to a knee point level and compressive thereafter. The knee point of the mapping functions changed dynamically and was set proportional to the estimated noise floor level. The performance of the mapping function was evaluated on a sentence recognition task using IEEE sentences embedded in 5 to 10 dB SNR multi-talker babble and in 5 dB SNR speech shaped noise. Nine postlingually deafened cochlear implant users participated in the study. Significantly higher performance was achieved with the s-shaped mapping functions than the conventional log mapping function used by cochlear implant users in their daily strategy, in both multi-talker and continuous speech shaped conditions, especially when the s-shaped mapping function was optimized to individual cochlear implant users.

In brief, only a few studies were conducted to develop a new speech processing strategy for cochlear implant integrating a noise reduction algorithm. In this paper, we propose a simple noise reduction algorithm that can be easily integrated in existing strategies used in commercially available devices, especially Digisonic SP CI manufactured by Neurelec.

The current paper is outlined as follows. Section 2 provides theoretical overview of the proposed speech enhancement system for cochlear implant. Firstly, we propose a comparative study of different noise reduction algorithms based on Minimum Statistics (MS) approach. Best performance's algorithm will be then used for enhanced speech spectrum estimation. Section 3 evaluates the experimental results and gives an overall discussion of all obtained results. Section 4 concludes the paper

2. Noise reduction Algorithm

The proposed cochlear implant coding strategy, integrating the noise reduction algorithm, is illustrated in figure 2. Generally, the digital signal processor function for a cochlear prosthesis primarily consists in dividing an input speech signal into a number of frequency bands in order to extract the input signal energy in each band corresponding to each implanted electrode. This goal could be achieved by computing the Short Time Frequency Transform of the input signal (STFT), 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. When input speech signal is affected by external noise, a noise reduction algorithm is needed, and the signal energy in each channel is computed from enhanced speech signal. The proposed coding strategy contains then the following major parts:

- Computing the STFT of the input signal and grouping frequency bins into different channels
- Noise power spectrum estimation,
- Enhanced speech signal estimation and energy computing.

2.1. Noise power spectrum estimation

The noise estimator is, a very important component of the overall speech enhancement system, especially if the algorithm should be capable of handling non-stationary noise. In fact the noise estimator has a major impact on the overall quality of the speech enhancement. The simplest approach is to estimate and update the noise spectrum during the silent (e.g., during pauses) segments of the signal using a voice-activity detection (VAD) algorithm (e.g., Sohn et al., 1999). Although, such an approach might work satisfactorily in stationary noise (e.g., white noise), it will not work well in more realistic environments (e.g., in a restaurant) where the spectral characteristics of the noise might be changing constantly. Hence there is a need to update the noise spectrum continuously over time and this can be done using noise-estimation algorithms. A useful noise estimation approach, known as the MS, is to track the minima values of a smoothed power estimate of the noisy signal, and multiply the result by a factor that compensates the bias (Martin , 1994). Martin (Martin , 2001) proposed a method for estimating the noise spectrum based on tracking the minimum of the noisy speech over a finite window. A different non-linear rule is used Farsi, 2010 for tracking the minimum of the noisy speech by continuously averaging past spectral values. As the minimum is typically smaller than the mean, unbiased estimates of noise spectrum were computed by introducing a bias factor based on the statistics of the minimum estimates. While for some cases exact expressions for the bias are available, approximations are required in general. Martin (Martin, 2006) present approximations which allow an efficient computation and compensation of the bias. Doblinger (1995) updated the noise estimate by continuously tracking the minimum of the noisy speech in each frequency bin. Rangachari et al., (2006) introduced a noise-estimation algorithm (which updates the noise estimate faster than the above methods and also avoids overestimation of the noise level. The noise estimate was updated in each frame based on voice-activity detection. If speech was absent in a specific frame, the noise estimate was updated with a constant smoothing factor. The speech-presence decision made in each speech frame was based on the ratio of noisy speech spectrum to its local minimum. Rangachari et al., (2006) proposed an improved noise estimation algorithm which updates the noise estimate in each frame using a time–frequency dependent smoothing factor computed based on the speech-presence probability.

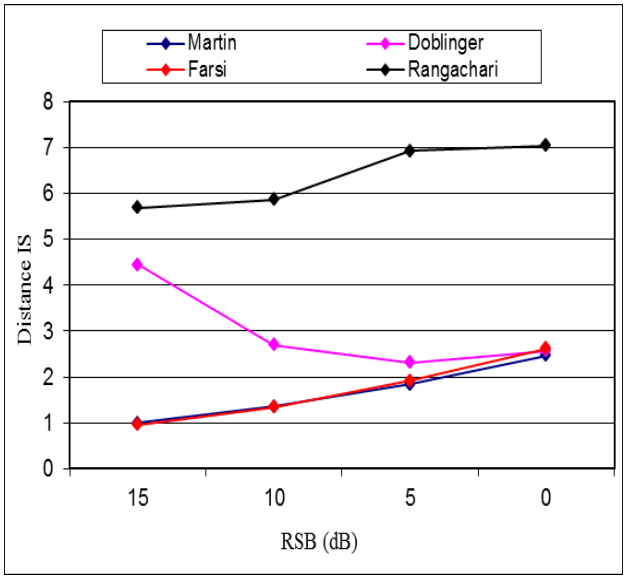
In this paragraph, we present a comparative study of four noise reduction algorithms proposed by Farsi (Farsi, 2010), Martin (Martin, 2006), Doblinger (Doblinger, 1995) and Rangachari (Rangachari, 2006).

The performances of the previously considered noise power spectrum estimation are evaluated with only one interfering babble noise source at different SNR levels varying from 0 to 15dB with 5dB step. The two following objective measurements are considered for performance assessment:

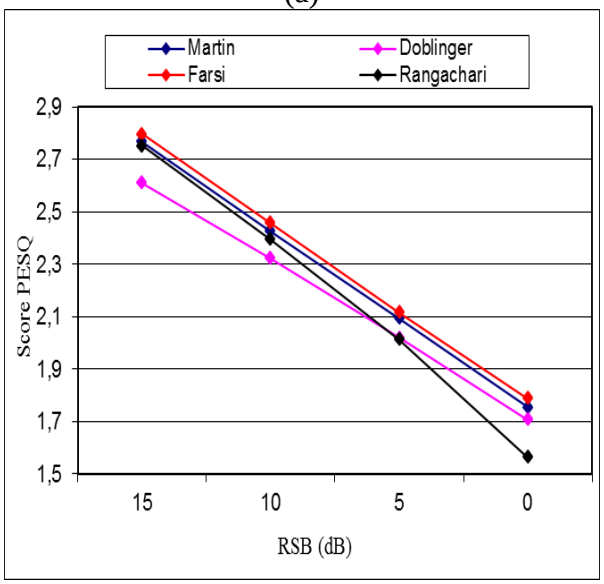
– Perceptual Evaluation of Speech Quality (PESQ) score which ranges from 0.5 (for the worst case) to 4.5 (for the best case) according to the ITU-T Recommendation P. 862 standard (P.862., 2001).

– Itakura-Saito (IS) distance which is based on the similarity or difference between the all-pole model of the clean and the enhanced speech signals (Quackenbush et al., 1988).

Results are shown in Fig.1a and Fig.1b respectively indicating the mean PESQ score and IS distance for different MS based noise spectrum estimation algorithms.



(a)



(b)

Fig. 1. A comparative study of MS based noise spectrum estimation algorithm for different SNR levels: (a) Mean IS values (b) Mean PESQ scores

It is clear from both figures that best performances are noted, for different SNR levels, when Farsi algorithm is considered for noise power spectrum estimation. In fact, low IS distance, indicating better performances, are observed for Farsi and Martin algorithms with a little superiority for Farsi algorithm at 0dB SNR level. On the other hand, highest PESQ scores are obtained for different SNR levels when Farsi algorithm is considered for noise spectrum estimation.

2.2. Enhanced speech signal estimation and energy computing

The given signal bandwidth ranged from 300 to 6055 Hz. Then, several ways of allocating the filters in the frequency domain were considered. An approximate analytical expression for describing the conversion from linear frequency f (in Hz), into the critical band unit b (in barks) is given by the following equation (Tranmüller, 1990) which was used in the current study:

$$b(f) \approx 6.7 * ArgHyperbolic \sin(\frac{f-20}{600}) \quad (1)$$

where f is the frequency bin index.

In CI, many coding strategies exist, but, with few exceptions, they are all variants of the Continuous Interleaving Sampling (CIS) or Advanced Combination Encoder (ACE) methods (Cooper et al., 2006). They split the input speech signal into short time-segments (frames) and use a filter bank to yield a M-band spectral representation. Next, 'N' bands (N<M for ACE, N=M for CIS) having the largest amplitude are selected and compressed, in order to match the narrow dynamic range of electrical hearing stimulation. In the current study, the ACE strategy was adopted with M=20 and N=8. Practically, the considered frequency range was divided into 'M' bands. This frequency range spanned from 3 to 20 barks; the spacing step was 0.85 bark (table.1) Sampling rate was fixed at 16 kHz. The picked up speech signal could be expressed in temporal domain as below:

$$y(l) = s(l) + d(l) \quad (1)$$

Where s(l), d(l), and y(l) represent respectively speech signal, noise signal, and noisy speech signal. l is the discrete-time-index. The noisy speech signal, y(l), was divided into overlapping frames by the application of a window function and analysed using the Short-Time Fast Fourier Transform (STFFT) classically given by:

$$Y(f,\lambda)=\sum_{n=0}^{L-1} y(l+\lambda.U).w(l).e^{-j\frac{2\pi}{L}l.f} \quad (2)$$

where λ is the time frame index, U is the frame length in time and w is the analysis window (Hanning window) of size L given by equation.3.

$$w(l)=0.5\left(1-\cos(2.\pi\frac{l}{L})\right) \quad (3)$$

The spectrum bands were then grouped according to the filter widths. The bandwidths of the filters were the frequency range leading to the simulated electrical stimulation. The window length (frame) was set to L=128 samples, corresponding to 64 spectrum bins. This value is a good compromise between spectral resolution and time resolution. The FFT bins were then combined to provide the required number of analysis channels and table 1 indicates the number of the FFT bins assigned to each analysis channel and the characteristics of the frequency channels.

Table 1: FFT bins attributed to the frequency channels and their corresponding frequencies in Hz

Frequency Channels 'm'	Number of bins Nm	Starting bin nstartm	Center frequencies fcenter (Hz)	Cutoff frequencies bm-em (Hz)
1	1	3	375	300-387
2	1	4	438	387-480
3	1	4	500	480-581
4	1	5	625	581-690
5	1	6	750	690-810
6	1	7	875	810-944
7	1	8	1000	944-1093
8	2	9	1187	1093-1259
9	1	11	1375	1259-1445
10	2	12	1562	1445-1655
11	2	14	1812	1655-1891
12	2	16	2062	1891-2158
13	2	18	2312	2158-2459
14	3	20	2625	2459-2801
15	3	23	3000	2801-3188
16	4	26	3437	3188-3626
17	3	30	3875	3626-4123
18	5	33	4375	4123-4688
19	5	38	5000	4688-5328
20	11	43	5687	5328-6055

Since Boll's original work (Boll, 1979), many different variations of the spectral subtraction have been proposed (Berouti et al., 1979; Virag, 1999; Soon et al., 2000; He and Zweig, 1999). Most, if not all, implementations of the spectral subtraction approach are variants of the approach proposed by Berouti et.al. (Berouti et al., 1979). In this implementation, the estimate of the clean speech spectrum is obtained as:

$$\left|\hat{S}(f)\right|^2 \approx |Y(f)|^2 - \alpha.\left|\hat{D}(f)\right|^2 \quad (8)$$

Since the noise spectrum cannot be directly obtained, an estimate $\hat{D}_m(f)$ is calculated using an appropriate noise estimation algorithm. The over-subtraction factor α is frequency dependent and computed as a function of the segmental SNR (SSNR). The over-subtraction method assumes that the noise affects the speech spectrum uniformly and the over-subtraction factor subtracts an overestimate of the noise over the whole spectrum. However, with real-world noise (in real environments), the noise spectrum is not uniform for all frequencies. In order to reduce the speech distortion and the musical noise caused by large values of α, its value is adapted from frame to frame. The basic idea is to take into account the fact that the subtraction process must depend on the SSNR of the frame in order to apply less subtraction with high SSNR and vice versa. That's why; a multi-band spectral subtraction approach was proposed in Kamath et al, (2002) and adopted by Udrea et al. (2008). In multi-band spectral subtraction approach, the noisy spectrum is divided into L non-overlapping bands, and spectral subtraction is performed independently in each frequency band.

In our work, multi-band spectral subtraction algorithm is integrated into adopted CI coding strategy. Thus, the estimate of the clean speech spectrum in the m-th band is given by equation (10).

$$\left|\hat{S}_m(f)\right|^2 = \left|Y_m(f)\right|^2 - \alpha_m\delta\left|\hat{D}_m(f)\right|^2 \qquad b_m \le f \le e_m \quad (10)$$

Where bm and em are the beginning and ending frequencies bins of the m-th frequency band, αm is the over-subtraction factor of the m-th band and δ is a tweaking factor that can be individually set for each frequency band to customize the noise removal properties. The band specific over-subtraction factor αm is a

function of the SSNRm of the m-th frequency band which is calculated as:

$$SSNR_m(dB) = 10.\log_{10}\left(\frac{\sum_{f=b_m}^{e_m}|Y_m(f)|^2}{\sum_{f=b_m}^{e_m}|\hat{D}_m(f)|^2}\right) \quad (11)$$

Where Ym (f) is the spectrum of the noisy speech and $\hat{D}_m(f)$ is the estimated power spectrum of the noise signal in the m-th frequency band. According to the SSNRmvalue calculated in equation (11), the over-subtraction factor αm is given by equation (12).

$$\alpha_m = \begin{cases} 5 & SSNR_m < -5dB \\ \alpha_0 - \frac{3}{20}(SSNR_m) & -5 \le SSNR_m \le 20dB \\ 1 & SSNR_m > 20dB \end{cases} \quad (12)$$

Where α0=4 is the desired value at 0dB SSNR.

within each band, since most of the speech energy is present in the lower frequencies, smaller δ values were used for the low frequency bands in order to minimize speech distortion. The values of δ were empirically determined and set to following values (Udrea et al., 2008).

$$\delta = \begin{cases} 1 & 60Hz \le f_l \le 300Hz \\ 1.3 & 300Hz \le f_l \le 1KHz \\ 1.6 & 1KHz \le f_l \le 2KHz \\ 1.8 & 2KHz \le f_l \le 3KHz \\ 1.3 & 3KHz \le f_l \le 8KHz \end{cases} \quad (13)$$

Factors, αm and δ could be adjusted for each band for different speech conditions to get better speech quality. The negative values in the enhanced spectrum in equation (14) were floored to the noisy spectrum as:

$$|\hat{S}_m(f)|^2 = \begin{cases} |\hat{S}_m(f)|^2 & if \quad |\hat{S}_m(f)|^2 > \beta.|\hat{D}_m(f)|^2 \\ \beta.|\hat{D}_m(f)|^2 & otherwise \end{cases} \quad (14)$$

Where the spectral floor parameter was set to β= .002.

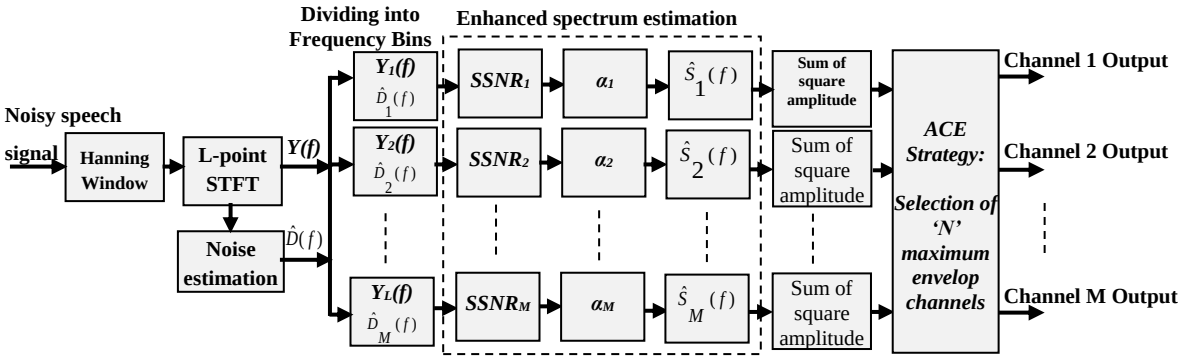


Fig. 2. Block diagram of the proposed noise reduction algorithm integrated into ACE coding strat

While the use of the over-subtraction factor αm provides a degree of control over the noise subtraction level in each band, the use of multiple frequency bands and the use of the δ weights provide an additional degree of control

The processed enhanced signals are then processed using the following power estimation equation:

$$E(m) = \frac{\sum_{f=nstart_m}^{f=nstart_m+N_m-1} \left|\hat{S}_m(f)\right|}{N_m}, m = 1...M \quad (15)$$

Where E(m) is the power of the considered band.

The output of this analysis is a vector of power values for each frame of data. According to the ACE strategy and for each frame, only the first 'N' channels presenting the most important power levels are used. The other channels were set to zero.

In order to investigate the behavior of the above described speech enhancement algorithm in the case of cochlear implant, vocoder stimulations were used. However, it was shown by many (e.g. (Whitmal et al., 2007)) that these simulations provide results consistent with the outcome of cochlear implants and that vocoded speech signals could be presented to normal hearing listeners in the absence of confounding factors associated with cochlear implants.

To reconstruct the acoustic signal, for each frequency band 'm', a Hanning window 'w(l)' was weighted by its related power value 'E(m)' to get the envelope signal, Env(m,l), according to Equation (6).

$$Env(m,l) = E(m).w(l), \quad m = 1:M, \quad l = 1:L \quad (6)$$

To prevent sharp variations, a further low pass filtering (cut off at 150 Hz) was applied on the envelope (smoothing). Then, a white noise was shaped to fill each frequency band (3rd order Butterworth filter) having the same band-pass as the selected frequency band. For each frequency band, the speech signal was synthesized by the multiplication of the filtered narrow band noise signal by the corresponding smoothed envelop. Finally, all signals coming from the different channels were summed up and the signal power of the processed speech signal was normalised so that the speech was reproduced at the same sound

pressure level as measured when the original speech was recorded (70 dB). Figure 3 present the envelop signal on channel 3 for clean signal, noisy signal, enhanced signal using both sigmoidal and proposed approaches.

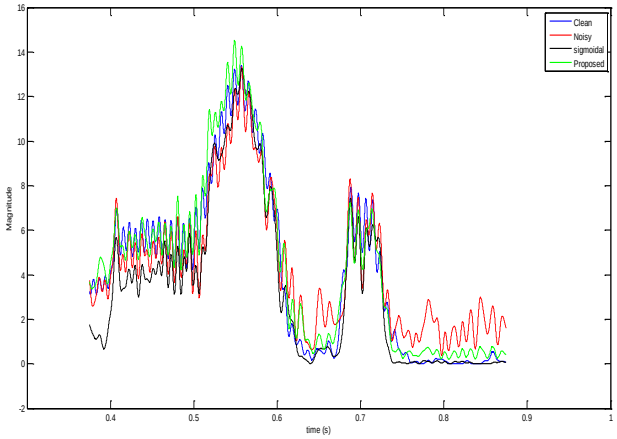


Fig. 3: temporal variation of clean, noisy and enhanced signals

3. Performance evaluations

In this experiment, we investigate the potential benefits of processing the noisy speech signal with the proposed speech enhancement algorithm. Therefore, to assess the performance of the aforementioned noise reduction algorithm, comprehensive phoneme's recognition tests were conducted with simulated CI users.

3.1. Phonetic material

The used phonetic material was the French Lafon set which contains twenty lists composed of 17 three-phoneme words pronounced by a single male talker. This is the most commonly used speech stimuli for intelligibility assessment in French. Sound level was calibrated to 70 dB SPL. All these lists were recorded in the anechoic room of the ORL department of the Edouard-Herriot Hospital of Lyon-France with an additional babble noise and then speech-shaped noise. The Signal to Noise Ratio (SNR) was varied from -3dB to 6dB in 3 dB steps.

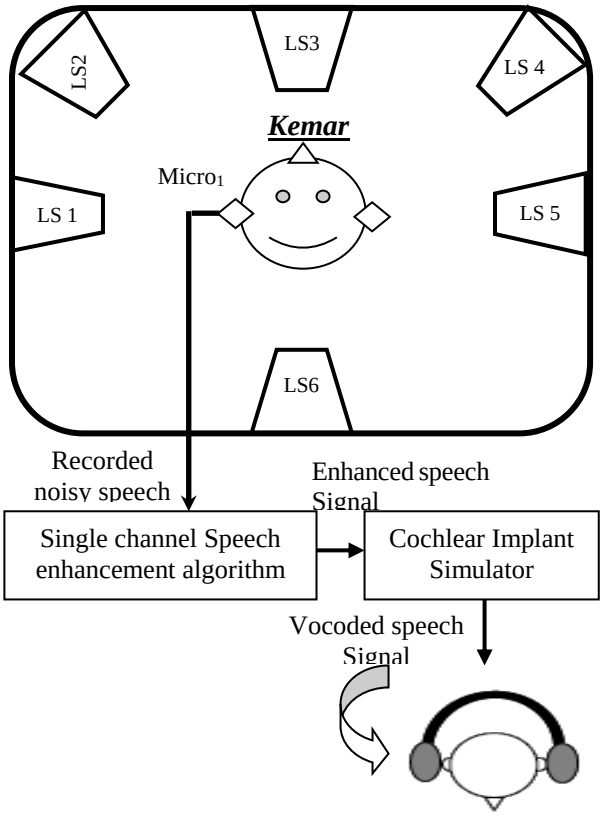


Figure 4. Experimental setup (Anechoic room) LS3=speech signal, LS2, LS4 =noise signal, LS1, LS5 and LS6 were not used

The experimental setup is presented in figure 4. A CD player (PHILIPS-CD723) was connected to an audiometer (MADSEN-Orbiter 922) for intensity level adjustment. The target speech signal was always placed directly in front of the listener (an artificial head) at 0° azimuth (LS3 position). Clean speech signal is corrupted by multi-talker babble and speech-shaped noise. Two interfering noise sources were placed asymmetrically either across both hemifields (-60° and 60° corresponding respectively to LS2 and LS4 positions).

3.2. Subjects

Performance evaluation of the proposed speech enhancement algorithms was done with a population of fifty normal hearing subjects; their audiogram was tested in the ORL department prior to the experiment. The age of the normal hearing subjects ranged from 18 to 32 years. All participants were native speakers of French. CI subjects are fitted with the binaural Digisonic SP multichannel implant device manufactured by Neurelec Corporation (France) and their biographical data are indicated in Table 3. All participants were native French speaking subjects. Listening sessions took place in the Cochlear Implant Room of the Edouard-Herriot Hospital. Subjects' hearing is checked prior to the experiment. Tests with all the subjects were done in Cochlear Implant Room of Edouard-Herriot Hospital of Lyon These tests were approved by the Lyon Hospitals Ethical Committee.

Table 2: Biographical data of recruited CI implantees

	Subject 1	Subject 2	Subject 3
Age (years)	38	52	45
Gender	F	M	M
Past surgery (years)	5	5	5`

3.3. Listening session conditions

After listening to each sentence, subjects were instructed to repeat what they heard. Before each situation, subjects were given a practice session containing ten random words processed according to that situation. None of the sentences used in test was used in the practice. No score was calculated for these practice sets. To minimize any order effect in the experience such as learning or fatigue, all conditions were randomized among subjects. Different sets of sentences were used in each condition. A sequential test order, starting with sentences processed in quiet and in noise from the highest SNR level (6dB) and to the lowest SNR level (-3dB) was employed. We took this sequential approach in order to give the subjects some time to adapt to the listening in noisy conditions. At the end of each listening session, the responses of each individual were collected, stored and scored off-line with the number of correctly identified phonemes. All phonemes were scored. The percentage of correctly repeated phonemes was then calculated (out of two Lafon's lists,

corresponding to 102 phonemes). Experiments were performed using a PC equipped with a conexant AC-link audio soundcard. To evaluate subjectively the proposed speech enhancement algorithm, three methods were considered:

- Noisy speech signal (non-processed signal) considered as a reference condition.
- Enhanced speech signal using the proposed speech enhancement algorithm.
- Enhanced speech signal using sigmoidal-shaped function proposed by Hu et al. (Hu et al., 2005).

All these experiments were conducted using speech babble and then speech shaped noise as interfering signals. Interfering noise signals are coming from three noise sources placed asymmetrically either across both hemifields (-60°, 60°, and 90° correspond respectively to LS2, LS4 and LS5 positions). Each subject listened to a total of 48 lists (three speech enhancement algorithms at two different noise types at 4 SNR levels). These lists were played on the CD player in a random order. For each situation, the subjects listened unilaterally to the sentences using a closed professional 'Sennheiser' HD250 linear headphones at a comfortable level calibrated to 70 dB SPL. All the situations were tested with the same subjects for the different SNRs. The speech stimuli were processed offline with MATLAB software.

3.4. Results

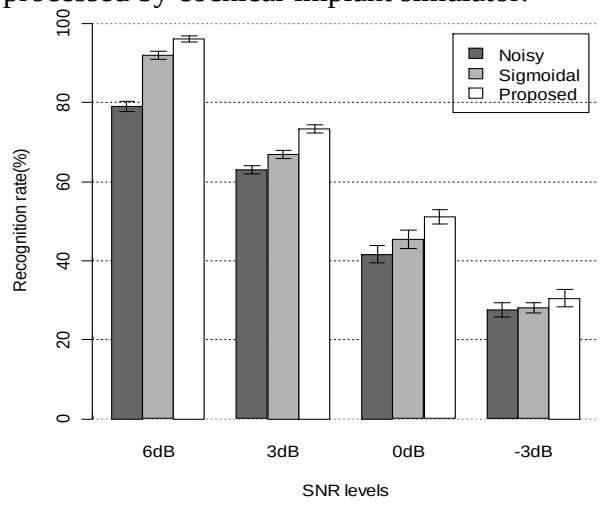
To determine the impact of speech enhancement algorithms on speech intelligibility with cochlear implant, the theory of linear mixed-effects model (Bates, 2007) was considered. We used the lmer program of the lme4 package in the R environment (R Development Core Team, 2007). Intelligibility scores for this experiment were derived from the percentage of correctly repeated phonemes per situation.

The effects were studied through a mixed ANOVA analysis with the following parameters:

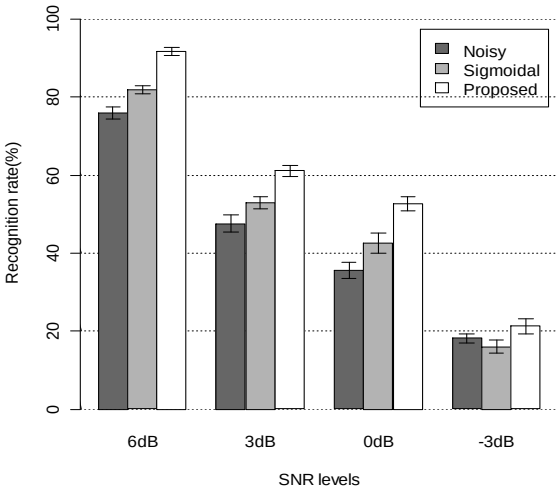
- Repeated measure (the same subjects under went all the situations)
- Dependant variable: the recognition rate in percent (score)
- Three factors:
  - Speech enhancement algorithm (Noisy, proposed algorithm, sigmoidal-shaped function)
  - Interfering noise type (speech babble or speech shaped noise)
  - Noise level (6 to -3 dB with 3 dB step, this last factor was taken as random; the first two factors were fixed).

3.4.1. Results with normal hearing subjects

Figure.5 shows the mean score as a function of speech enhancement algorithm at four SNR levels (6dB, 3dB, 0dB, -3dB) in the presence of two interfering noise type. Figure.5 (a) and Figure.5 (b) gives the mean score when a speech shaped noise and babble noise are considered respectively. First, second and third fourth line of figure.6 gives the mean score at 6dB, 3dB, 0dB, -3dB SNR levels respectively. A group of four values is presented at each SNR level and at each interfering noise type. The first value indicates the mean percent score obtained with vocoded noisy speech signal. Second and third bars indicated the mean percent score obtained with enhanced speech signal processed using sigmoidal-shaped function and proposed MBSS speech enhancement algorithms respectively and then processed by cochlear implant simulator.



(a)



(b)

Figure 5. Recognition score (mean ± standard error) with noisy, speech sigmoidal shaped function and MBSS enhancement algorithms in the presence of (a) speech shaped noise (b) babble noise

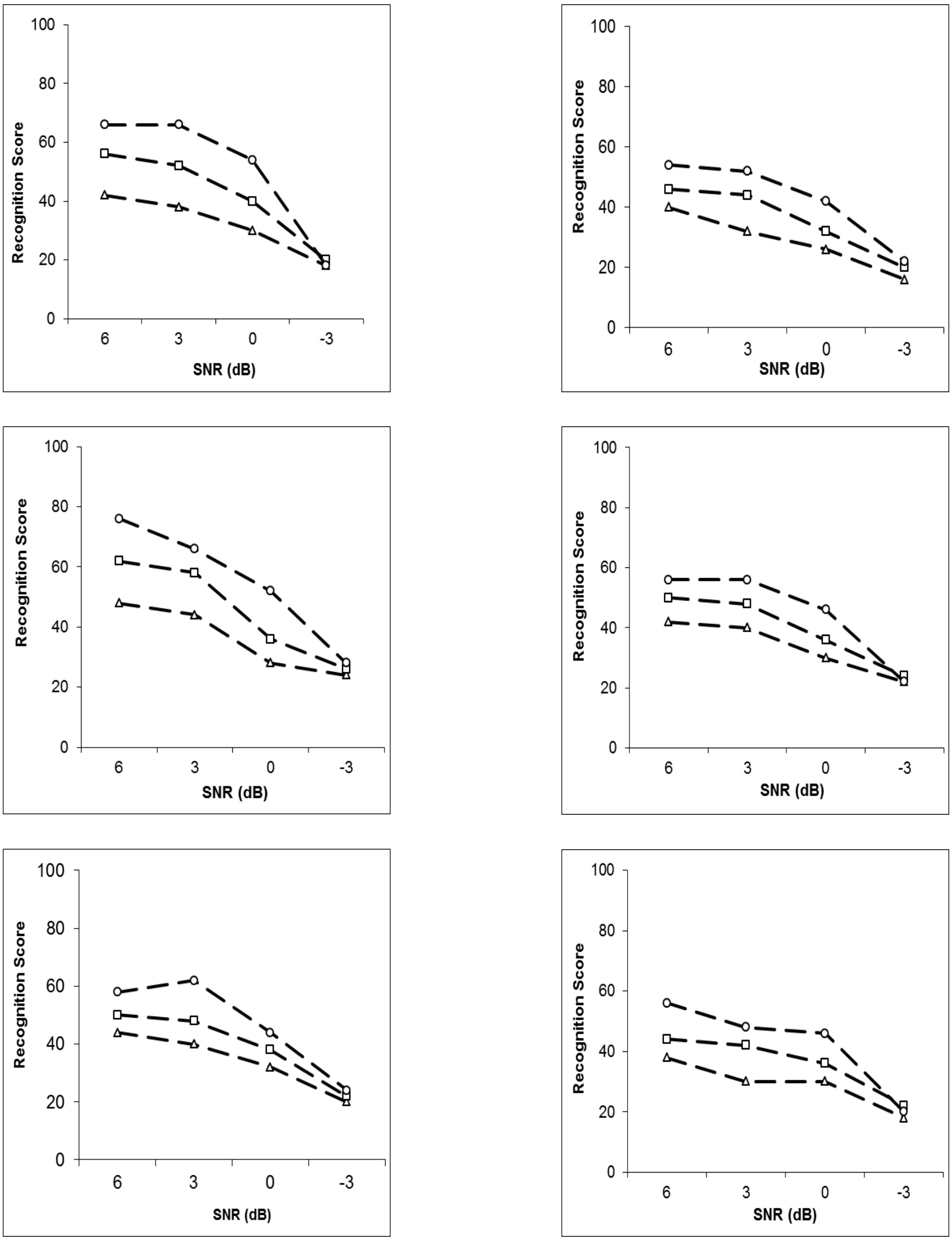
It is clear from figure.5 that better performances are obtained when proposed MBSS speech enhancement algorithm is considered. A decrease in performance is noted when the SNR is decreased and particularly in the presence of interfering babble noise. Statistical analysis results indicate a main effects of the speech enhancement algorithm (Chi2[2]=690, p<0.001), of the SNR (Chi2[3]=14086, p<0.001) and of the interfering noise type (Chi2[1]=502, p<0.001). Furthermore, there were a significant interaction between the speech enhancement algorithm and the SNR (Chi2[6]=322, p<0.001), a significant interaction between the speech enhancement algorithm and the interfering noise type (Chi2[2]=19, p<0.001) and between interfering noise type and the SNR (Chi2[3]=142, p<0.001).

Post-hoc comparisons were run to assess significant differences in scores between the scores obtained with considered speech enhancement algorithms. That's why we used the 'glht' function from 'multcomp' package of R to take a fitted response model and a matrix defining the hypotheses of interest to perform the multiple comparisons. Statistics results shown that phoneme recognition scores were significantly better with sigmoidal-shaped function and proposed MBSS speech enhancement algorithms when SNR levels are respectively fixed at 6dB (p<0.001) and 0dB (p=0.025), but there isn't a significant improvement when the proposed MBSS speech enhancement algorithm is considered, compared to sigmoidal-shaped function (p=0.9). The mean recognition scores achieved with both considered speech enhancement algorithms are significantly improved (p < 0.001) and significant higher performances are noted when the proposed MBSS speech enhancement is considered (p < 0.001) at 3dB SNR level. No significant improvement was seen at -3dB with both considered speech enhancement algorithms (p=0.95).

In the presence of speech shaped as an interfering noise, with proposed MBSS speech enhancement algorithm is considered, phoneme recognition score are significantly higher than those obtained with sigmoidal shaped function at 3dB SNR level (p < 0.001), but no significant improvement is observed at 6dB, 0dB and -3dB SNR levels (p >0.5). With clean speech signal embedded in 6dB, 3dB and -3dB SNR babble, a significant improvement in phoneme recognition score is observed (p < 0.001), but no significant improvement is noted at -3dB SNR level (p=0.6).

3.4.2. Results with CI implantees

Experimental results obtained with three CI implantees are presented in this section. Figure. 6a and Figure. 6b show the individual subject scores obtained with different considered speech enhancement algorithms when speech shaped noise and babble noise are respectively considered. Statistical tests indicate a main significant effect of the speech enhancement algorithm (Chi2[2] = 69, p < 0.001), of the SNR (Chi2[3] = 232, p < 0.001) and of interfering noise type (Chi2[1] = 14, p < 0.001). However, there is no significant interaction between the speech enhancement algorithm and the SNR (Chi2[6] = 10, p = 0.11) and between the speech enhancement algorithm and the interfering noise type (Chi2[2] = 1.5, p = 0.5).



(a) Speech shaped noise

(b) Babble noise

Figure 6.Phoneme's recognition scores with deafened CI subjects for different Speech enhancement algorithms at all SNRs.

4. Discussion

In this paper, we proposed a new speech processing strategy for cochlear implant integrating a noise reduction algorithm. Multi band spectral subtraction based proposed speech enhancement algorithm can be easily integrated in existing strategies used in commercially available devices. The above analysis clearly indicates that the proposed sigmoidal-shaped function provided significant benefits to CI users in nearly all conditions. We believe that much of the success of the proposed noise reduction algorithm can be attributed to the improved temporal envelope contrast.

As a first experimental study, performance of speech enhancement algorithms was tested with normal hearing subject using a CI simulator. When speech shaped was considered as an interfering noise, the mean improvement observed was 16% with sigmoidal-shaped function speech enhancement algorithm and by 21% with proposed multi-band speech enhancement algorithm. In the conditions where babble noise were present, we noted a mean improvement of 5% with sigmoidal-shaped function speech enhancement algorithm and 11% improvement with proposed multi-band speech enhancement algorithm.

Next, the same experiment was performed with three CI implantees. When the multi-band speech enhancement algorithm was applied, we observed a benefit in term of recognition score from 14% when speech babble is considered to around 19% in the presence of speech shaped noise. This improvement in performance was a bit smaller for the case of the sigmoidal-shaped function and the improvement in term of recognition score are variable from 7% when speech shaped noise were present to around 9% in the presence of speech babble.

The percentage point improvement is also variable as a function of the SNR level. In fact, at 6dB, 3dB, 0 dB and -3dB SNR levels, we noted an improvement of 13%, 12%, 15% and 2% respectively when sigmoidal-shaped function speech enhancement algorithm is considered. When the proposed multi-band speech enhancement algorithm is considered, we noted an improvement of 18%, 14, 20% and 11% at 6dB, 3dB 0 dB and -3dB SNR levels respectively. Overall, our algorithm compares favorably against other single-microphone methods proposed for cochlear implants (e.g., Yang and Fu, 2005; Loizou et al. 2005). A larger improvement in performance was obtained with our proposed method in the presence of babble noise compared to the improvement reported by the other preprocessing method. Other advantages of the proposed method include the lack of algorithmic delay associated with preprocessing techniques, low computational complexity, and ease of integration in existing CI strategies.

This behavior was confirmed with the CI implantees. In fact, an average improvement in recognition score of 9%, 11%, 7% and 3% was respectively observed at 6dB, 3dB, 0dB and -3dB when sigmoidal-shaped function was considered. Better performance was obtained when multi-band speech enhancement algorithm was considered and the average improvement in recognition score are respectively 19%, 21%, 18% and 3% at 6dB, 3dB and 0dB and -3dB.

Acknowledgement

This research work is supported by the Franco-Tunisian cooperation project 'PHC-Utique' managed by CMCU (2008-2011). We appreciate the efforts of all the subjects. We also would like to thank the Cochlear Implantation Center in the Edouard Herriot hospital in Lyon France for the technical support and for the assistance in the listening test process.

.

References

1. Bates D. lme4: linear mixed-effects models using S4 classes. R package version 0.99875-8; 2007.

2. Berouti, M., Schwartz, R. and Makhoul, J., "Enhancement of speech corrupted by acoustic noise," Proc. IEEE Int. Conf. Acoust., Speech, Signal Process., pp. 208-211, Apr. 1979.

3. Boll, S.F. “Suppression of acoustic noise in speech using spectral subtraction,” IEEE Trans. Acoust., Speech, Signal Process., vol.27, pp. 113-120, Apr. 1979.

4. Cooper, H.R., Craddock, L.C., 2006. Cochlear Implants-A Practical Guide (2nd edition). West Sussex, England: Whurr Publishers Limited, 1-69.

5. Doblinger, G., 1995. Computationally efficient speech enhancement by spectral minimatracking in subbands. Dans Proc. Eurospeech, pages 1513–1516.

6. Farsi, H., “Improvement of minimum tracking in minimum statistics noise estimation method”, Int. Jour. Signal Process., vol. 4, pp.17–22, 2010.

7. Fetterman, B. L., Domico, E. H., 2002. Speech recognition in background noise of cochlear implant patients. Otolaryngology Head and Neck Surgery. 126, 257-263.

8. ITU-T Recommendation P.862. Perceptual evaluation of speech quality (PESQ): an objective method for end-to-end speech quality assessment of narrowband telephone networks and speech codecs; 2001.

9. Kamath, S., Loizou P. C., “A multi-band spectral subtraction method for enhancing speech corrupted by colored noise”, IEEE Int. Conf. ICASSP, pp. 4160–4164, 2002.

10. Kasturi K., Loizou P., « Use of s-shaped input-output functions for noise suppression in cochlear implants », Ear Hear, vol 28, 2007, p. 402-11.

11. Kiefer, J., Müller, J., Pfenningdorff, T., 1997. Speech understanding in quiet and in noise with the CIS speech-coding strategy (MED EL Combi-40) compared to the MPEAK and SPEAK strategies (Nucleus). Advances in Otorhinolaryngology. 52, 286–290.

12. He, C., and Zweig, G., “Adaptive two-band spectral subtraction with multi-window spectral estimation,” ICASSP, vol.2, pp. 793-796, 1999.

13. Hu Y., Loizou P., Li N., Kasturi K., « Use of a sigmoidal-shaped function for noise attenuation in cochlear implants », J Acoust Soc Am, vol 122, 2007, p. 128-34.

14. Loizou, P., Lobo, A., and Hu, Y. (2005). “Subspace algorithms for noise reduction in cochlear implants,” J. Acoust. Soc. Am. 118(5), 2791–2793.

15. Loizou, P. (2006). “Speech processing in vocoder-centric cochlear implants,” Adv. Oto-Rhino-Laryngol. 64, 109–143.

16. Martin, R., “Spectral subtraction based on minimum statistics”, Proc EUSIPCO, pp.1182–1185, 1994.

17. Martin, R., “Noise power spectral density estimation based on optimal smoothing and minimum statistics”, IEEE Trans. on Speech and Audio Process., vol. 9, pp. 504–512, 2001.

18. Martin, R., 2006. Bias compensation methods for minimum statistics noise power spectral density estimation. Signal Processing, 86 :1215–1229.

19. Müller-Deile, J., Schmidt, B.J., Rudert, H. 1995. Effects of noise on speech discrimination in cochlear implant patients. Annals of Otology Rhinology and Laryngology. 166, 303-306.

20. Quackenbush, S., T. Barnwell et M. Clements, 1988. Objective measures for speech quality testing. Dans Prentice-Hall.

21. Rangachari, S. et P. C. Loizou, 2006. A noise estimation algorithm for highly non-stationary environments. Speech Communication, 48 :220–231.

22. R Development Core Team, R: A language and environment for statistical computing, R foundation for statistical computing. Vienna, Austria; 2007. ISBN 3-900051-07-0. <http://www.R-project.org>.

23. Sohn, J., Kim N. S.., Sung W., “A statistical model-based voice activity detection”, IEEE Signal Process. Lett, vol. 6 , pp. 1–3, 1999.

24. Soon, I., Koh, S., and Yeo, C., “Selective magnitude subtraction for speech enhancement,” Proceedings. The Fourth International Conference/Exhibition on High Performance Computing in the Asia-Pacific Region, vol.2, pp. 692-695, 2000.

25. Tranmüller H. Analytical expression for the tonotopic sensory scale. J Acoust Soc Am 1990;85:97–100.

26. Udrea, R. M., Ciochina S. D., Vizireanu N., “Multi-band bark scale spectral over-subtraction for colored noise reduction”, Proceedings of ISSCS, 2005.

27. Yang, L., and Fu, Q. (2005). “Spectral subtraction-based speech enhancement for cochlear implant patients in background noise,” J. Acoust. Soc. Am. 117(3), 1001–1004.

28. Wilson, B., Finley, C., Lawson, D., Wolford, R., Eddington, D., and Rabinowitz, W., 1991. Better speech recognition with cochlear implants. Nature. 352. 236-238

29. Wouters, J., and Vanden Berghe, J. (2001). “Speech recognition in noise for cochlear implantees with a two-microphone monaural adaptive noise reduction system,” Ear Hear. 22(5), 420–430.

30. Van Hoesel, R., and Clark, G. (1995). “Evaluation of a portable two-microphone adaptive beamforming speech processor with cochlear implant patients,” J. Acoust. Soc. Am. 97(4), 2498–2503.

31. Virag, N., ”Single channel speech enhancement based on masking properties of the human auditory system,” IEEE Trans. Speech and Audio Processing, pp 126-137, vol. 7, March 1999.

32. Weiss, M. (1993). “Effects of noise and noise reduction processing on the operation of the nucleus 22 cochlear implant processor,” J. Rehabil. Res. Dev. 30, 117–128.

33. Whitmal NA, Poissant SF, Freyman RL, Helfer KS. Speech intelligibility in cochlear implant simulations: ffects of carrier type, interfering noise, and subject experience. J Acoust Soc Am 2007;122:2376–88.