Robust CIS Strategy for Cochlear Implants

D. Tochetto, D. Milone, H. Rufiner and L. Aronson.
Laboratorio de Cibernética, Universidad Nacional de Entre Ríos, Entre Ríos, Argentina.

Abstract

Cochlear prosthesis provide high scores of speech comprehension in quiet environment. However, these scores decrease in noisy environments. In this work, the design, execution and evaluation of a complementary denoising block for the continuous interleaved sampling (CIS) stimulation strategy is presented. The denoising module is implemented with a time delay neural network. For the recognition tests, performed on normal hearing subjects, an acoustic simulator of a cochlear implant was used. Speech was corrupted with babble noise at 0, 5 and 10 dB of signal to noise ratio. Results of tests administered to subjects show that the proposed robust strategy was better than standard and enhanced CIS strategies.

Introduction

Cochlear implants (CI) allow to partially restore hearing in profoundly deaf people. Some cochlear implanted patients (CIP) are able to reach almost the same speech recognition scores than that of the normal hearing subject (NHS) in quiet environments. Nevertheless, the benefits obtained at quiet environments are poorer than those obtained in real environments where background noises are mixed with the speech signal. NHS is able to recognize speech in environments with extremely high noise levels [1,2]. The most important factors for this degradation in CIP are: spectral resolution [3], compression function [4], stimulation frequency [5], intensity resolution [6] and stochastic resonance [2]. In order to solve this problem, several robust stimulation strategies have been developed with partially satisfactory results. In the present study, CIS strategy improvements by adding a denoising block implemented with a time delay neural network (TDNN) are presented. This type of neural network is capable of speech dynamics modelization [7]. In this way, the new strategy, named CIS filtered (CISF), makes a dynamic filtering of the output amplitudes of the CIS strategy, process that is learnt during a training stage. Recognition scores obtained with this strategy are compared with those obtained by original CIS, and CIS Enhanced (CISE) [8] strategies, using an acoustic simulator of a cochlear implant (ASCI).

Method

Acoustic Simulator of Cochlear Implant

The ASCI used in [9] was implemented. ASCI reproduces the procedures performed by the CI sound processor and generate an acoustic signal, which is used to stimulate a NHS. The signal is composed by a summation of sine waves with time varying amplitudes and fixed frequencies. The following equation represents the synthesized speech signal $\underline{s}(m,n)$ in a given interval (4 ms) or frame m:

$$\underline{s}(m,n)=\sum_{i=1}^{8} A_i(m)\sin(2pf_i n + j_i) \qquad (1)$$

with n from 1 to N_m (number of samples in the frame m), $A_i(m)$ the root mean square (RMS) of the output amplitude in frame m and j_i the central phase of the *i-th* filter at frequency f_i, estimated by fast Fourier transform applied to the input. Finally, these signals are Hamming windowed, with an overlapping of $N_m/2$, and then added in order to obtain a smoother signal.

The ASCI were implemented using 8 processing channels. This number of channels arose from preliminary tests performed in order to find the minimum number necessary to reach an asymptotic recognition score, in quiet environments. ASCI generates processed signals with CIS, CISE and CISF strategies for stimulating the NHS. In the next sections, the way in which each of strategies generates the output envelopes used by the ASCI, will be described.

CIS Strategy

The input signal is pre-emphasized with a high pass filter (2nd order with cut-off frequency of 1200 Hz) and it is also decomposed with 8 band-pass filters (6th order), all distributed in logarithmic scale in the range from 300 to 5500 Hz. Then, the envelopes of these signals are extracted by full wave rectification and low pass filtering (2nd order with cut off frequency of 400 Hz). In CI, envelopes are compressed within the dynamic range of stimulation and are used for modulating electrical stimulus amplitude. But at this point, ASCI use this uncompressed amplitudes to generate stimulus in the way above mentioned.

CIS Enhanced Strategy

This strategy, proposed by Loisou and Li [8], improves speech recognition enhancing contrast among the output amplitudes of the CIS strategy. The method comes out by observing how noise decreases dynamic range within output channels, reducing spectral pick-to-valley ratio but preserving their location. The input signal is processed similarly to CIS strategy, and then over $A_i(m)$ the contrast enhancement is applied as follows:

$$B_i(m) = \left[\frac{A_i(m) - A_{\min}(m)}{A_{\max}(m) - A_{\min}(m)}\right] C\; A_i(m) \tag{2}$$

where $A_{max}(m)$ and $A_{min}(m)$ are the corresponding maximum and minimum amplitudes of all the channels, $B_i(m)$ is the enhanced amplitude of *i-th* channel and C is a positive gain constant experimentally determined.

CIS Filtered Strategy

A TDNN is trained in order to perform a dynamic filtering of the CIS strategy output amplitudes. The training method forces the net to discover the underlying temporal and spectral characteristics that manage the filtering process in order to obtain clean signals starting from the noisy ones. Fig. 1 depicts the diagram of the simulated strategy.

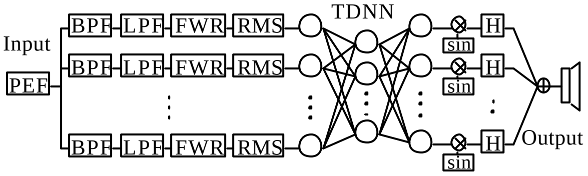

Figure 1. CISF strategy block diagram. PEF: pre-emphasis filter, BPF: band-pass filter, LPF: low-pass filter, FWR: full-wave rectification, H: Hamming window (time delays are not showed).

The dimension of input data was fixed in 8, which is the number of processing channels. Likewise, in order to maintain spectral resolution of the original CIS strategy, the number of neurons of the output layer was equal to the number of channels. TDNN was implemented with two hidden layers. The number of neurons of each one was obtained by means of objective measurements of the improvements produced by the net. Measurements consisted of evaluating the SNR rising, with CISF processed signal, by varying the number of hidden neurons. Therefore, the number of neurons of the first hidden layer was fixed equal to 8,

while number of neurons for the second hidden layer was fixed equal to 24. The number of delays for first hidden layer was six and two delays were selected for second hidden layer.
Training consisted of feeding noisy speech processed by CIS strategy to the TDNN and propagating it forward to the output layer. At this point the squared difference between target speech and the network output was calculated. Target output is obtained by processing the same but clean speech with CIS strategy. The error gradient is calculated and propagated back modifying the weights of each layer (Fig. 2). The process is repeated for the different training signals until reaching finalization criteria. The learning algorithm used was TDNN adapted back-propagation [7]. To carry out training stage, the input and target output should be normalized to values between 0 and 1.

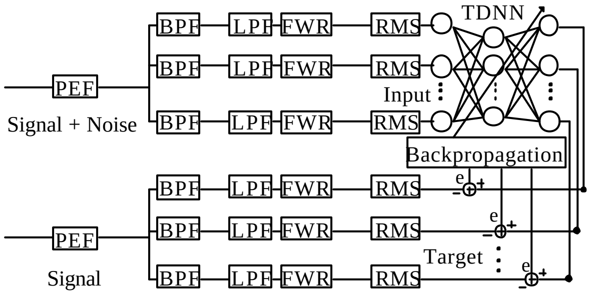

Figure 2. Block diagram for training TDNN (references as in Figure 1).

The ending criteria established were: maximum number of training epochs, minimum error gradient and minimum squared error (whatever first happens). To determine the maximum number of epochs, a training series was carried out, with a 0 dB SNR signal. Then, a test was accomplished with other signals at 10, 5 and 0 dB SNR. The generalization peak was reached at 700 training epochs. The values of minimum gradient and minimum square error were also experimentally obtained and they were set to 10^{-6} and 10^{-5} respectively.
Once the TDNN was trained, the signals used on the recognition tests were generated. For this, CIS output amplitudes are normalized and propagated to the output layer. Then, signals are denormalized and taken by the ASCI for stimulus generation.

Signals and subjects

In the experiments, the vowels /a/, /e/, /i/, /o/ and /u/, and the consonants /b/, /c/, /d/, /f/, /g/, /j/, /l/, /m/, /n/, /p/, /r/, /s/ and /t/ in vowel context /a/-consonant-/a/ were used. A native Argentinean female speaker produced all the utterances in Spanish. The signals were recorded using a Shure SM58 microphone, a Turtle Beach professional sound card and the acquisition program Scope [10]. Afterwards, noisy signals were generated at 0, 5 and 10 dB SNR using Babble noise from NOISEX-92. For the reproduction the same sound card and a high quality Thelefonic auricular were used.
Seventeen Argentinean NHS participated in this study. Each subject had a period of training with the speech material. The training consisted of two random presentations of each vowel and consonant with different noise conditions using the three strategies. In each presentation, the same signal was administered three times and the correct answer was showed in the computer screen. The tests of recognition in noise were performed at the end of the training period. These tests included 10 random presentations of each vowel and consonant, repeated three times, using the three strategies and the three noisy conditions. The speech materials utilized here were different to those used during the training stage. The subjects introduced answers by the keyboard choosing among the different alternatives showed in the screen.

Results and Discussion

The recognition for vowels obtained with the CIS strategy was 61,23%, 72,00% and 72,92%, with the CISF was 73,23%, 79,69% and 76,62%, and with the CISE was 59,69%, 68,62% and 74,15%, all for 0, 5 and 10 dB SNR. The recognition for consonants obtained with the CIS strategy was 39,78%, 50,77% and 53,19%, with the CISF was 40,00%, 52,53% and 54,95%, and with the CISE was 36,05%, 52,20% and 57,69%, all for 0, 5 and 10 dB SNR.

Results obtained with the CISE differ to that obtained in [8]. A reason could be the selected value of the enhancement constant *C* in (2). In [8], tests with consonants were not performed.

The CISF strategy improved markedly vowels recognition rates for all the noisy conditions. Nevertheless, for consonants recognition tests, the improvements were lower.

Conclusions

We could conclude that the CISF strategy reduces noise considerably, as well for vowels as for consonants, as it could be observed in objective measures. However, the subjective results indicate that in some cases this great reduction does not improve the speech intelligibility, but, on the contrary, add some distortions. One of the causes could be the complex temporal behavior of the consonants. Another cause could be the non-linearity in the filtering, generated by the sigmoidal transfer functions of the neurons.

References

[1] Huettel L., Collins L., A theoretical comparison of information transmission in the peripheral auditory system: Normal and impaired frequency discrimination, Speech Communication, 39:5-21, 2003

[2] Zeng, F-G., Fu, Q-J., Morse, R.P, Human hearing enhanced by noise, Brain Research, 869:251-255, 2000

[3] Fu, Q-J., Shannon, R., Wang, X., Effects of noise and spectral resolution on vowel and consonant recognition: Acoustic and electric hearing, J. Acust. Soc. Am. 104(6):3586-3596, 1998

[4] Wilson, B., Lawson, D., Zerbi, M., Wolford, R., Speech processors for auditory prostheses, Third Quarterly Progress Report, 1999, available at <http://www.rti.org>

[5] Aronson L., Pallares N., Effect of the Stimulation Rate on the Speech Perception in Patients with Cochlear Implants using the CIS Strategy, submitted to Medical Engineering and Physics, 2002

[6] Zeng, F-G., Galvin, J., Amplitude mapping and phoneme recognition in cochlear implant listeners, Ear Hear, 20:60-74, 1999

[7] Waibel, A., Hanazawa, T., Hinton, G., Shikano, K., Lang, K., Phoneme recognition using time-delay neural networks, IEEE Transaction on Acoustic, Speech and Signal Processing, 37(3):328-338, 1989

[8] Loizou, P., Liu, X., Improving vowel recognition in noise using the CIS strategy, 29th Annual Neural Prosthesis Workshop, NIH, Bethesda, MD, USA, 1998

[9] Dorman, M., Loizou, P., J. Fitzke, Tu, Z., The recognition of monosyllabic words by cochlear implant patients and by normal-hearing subjects listening to words processed through cochlear implant signal processing strategies, Annals of Otology, Rhinology and Laryngology, 109(12), Suppl. 185, 64-66, 2000

[10] Rufiner H., Martinez C., Sistema de Análisis de señales de voz de aplicación fonoaudiológica y lingüística, Anales del 1er Congreso Latinoamericano de Ingeniería Biomédica, Mazatlán 98, México, 1:741-744, 1998