AN INVESTIGATION OF NOISE RESISTANCE OF SPEECH STRATEGIES IN COCHLEAR INPLANTS

Yahya Öztürk* Gülden Köktürk**
e-mail: yahyaozturk@birim.com.tr *e-mail: gulden.kokturk@eee.deu.edu.tr*
**Birim Information Technologies, Atatürk Cad., no. 380, Kat 7, Alsancak, İzmir, Türkiye*
***Dokuz Eylul University, Faculty of Engineering, Department of Electrical & Electronics Engineering,Campus of Kaynaklar, 35160,Buca, Izmir, Turkiye*

Key words: Cochlear implants, CIS strategy, N-of-M strategy

ABSTRACT

Cochlear implants (CI) improve partial hearing to profoundly deaf people. Many investigators from various disciplines can be made combined efforts for progression on these implants. The speech processing strategy in modern CI's extracts and encodes amplitude information in a number of frequency bands. This paper investigates noise resistance different speech strategies that include CIS and N-of-M strategies and it compares each others for cochlear implants.

I. INTRODUCTION

A particular percentage of the populations in developed countries encounter hearing impairment. CI has been developed to increase the hearing capacity for people. In recent years, adults and children had benefited by usage of CI with improvement of implant techniques. Although these devices permit increasing performance, a significant gap in speech recognition still remain between CI listener and people which possess normal listening capability.

CI system often consists of the following modules: a microphone, a speech processor, a transmitter, a receiver and an electrode array (Figure 1) [1]. The speech processor is outside the body, and responsible for extracting spectral features from input speech signals in order to generate electrical stimulation pulses for each electrode. In response to the stimulation currents, auditory nerve fibers innervating along the cochlea will be activated, and then evoke sense of hearing. Actually, CI is an imitation of the normal human cochlea in terms of physiology.

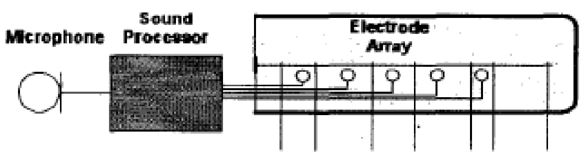

Figure 1 : Block diagram of CI system

Since William House and his associated developed the firs single channel implant, it responds to coarse temporal fluctuations as much as frequency characteristic [1]. Furthermore, speech recognition was restricted to transmitted frequency information and it was inadequate in comprehensibility. When multi-channel implants were introduced in the 1980s, several questions were raised regarding multi channel stimulation. Most important question was: "What kind of information should be transmitted to each electrode?" Depending on how researchers tried to address these questions, different types of signal processing techniques were developed. The various signal processing strategies developed for multi-channel cochlear prosthesis, can be divided into three categories: waveform strategies, feature-extraction strategies and "N-of-M" strategies [2][3]. These strategies differ in the way that information, is extracted from the speech signal and presented to the electrodes.

Speech coding strategies play an extremely important role in maximizing the user's overall communicative potential, and different speech processing strategies developed over the past two decades aim to mimic firing patters inside the cochlea as naturally as possible. In this paper, we introduce for two type strategies which are waveform strategies and "N-of-M" strategies and it organized as follows: Section 2 and Section 3 describe the noise resistance speech strategies in CI's. Section 4 presents the objective evaluation of the examined strategies and Section 5 gives our conclusions.

II. CIS SPEECH PROCESSOR FOR COCHLEAR IMPLANTS

CIS speech processor is developed by Wilson in 1991 [4][5][6]. Figure 2 describes the detailed configuration of a CIS speech processor. The input speech signal is first pre-emphasized for frequencies above 1.2 kHz at 6dB/Oct and then separated into several bands by a bank of band-pass (BP) filters. In each band envelope signal can be obtained after a rectifier (Rect.) and a lowpass filter (LPF). At the end, in order to generate stimulation pulses, the envelope signal is dynamically compressed into a proper scope with logarithmic or square law [7].

For human cochlea, the frequency-position function can be described as the following equation

$$f = A\left(10^{ax} - k\right) \tag{1}$$

Where ***f*** represents frequency in Hz, x is expressed as a proportion of basilar length (from 0 to l) A=165.4 and *a*=2.1, k=0.88.

Recent researches on mechanism of human cochlea revealed that sound vibration propagates in the form of traveling wave in cochlea. Along the basilar membrane of cochlea, the apical section corresponds to the low frequency receptor, and the basal corresponds to the high frequency section. In other words, cochlea is a frequency-analyzer in space domain. Thus, cochlea can be considered as a parallel bank of band-pass filters with almost constant quality (Q)-factors [8][9].

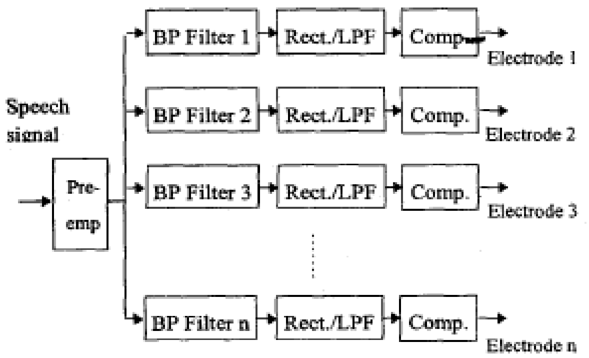

Figure 2. Block diagram of CIS speech processor

III. N-OF-M SPEECH PROCESSOR FOR COCHLEAR IMPLANTS

In these strategies, the signal is filtered into m frequency bands, and the processor selects, out of m envelope outputs, the n (n<m) envelope outputs with the largest energy (Figure 3). Only the electrodes corresponding to the n selected outputs are stimulated at each cycle. For example, in a 6-of-22 strategy, from a maximum of twenty two channel outputs, only the six channel outputs with the largest amplitudes are selected for stimulation at each cycle. The "N-of-M" strategy can be considered to be a hybrid strategy in that it combines a feature representation with a waveform representation.

IV. NOISE THEORY AND PERFORMANCE CRITERIA

Assuming that the speech signal, X, and the noise, N, are additive, the noisy speech, y, is modeled as

$$Y = X + N \tag{2}$$

It is generally adopted that the speech is not correlated with noise; this is a reasonable assumption in most cases when the signal and noise are generated by independent sources. We can write easily noise equation as

$$N = Y - X \tag{3}$$

The performance criteria is SNR value which is estimated by this formula

$$SNR\left(Y, \hat{Y}\right) = 10 \log \left(\frac{\|Y\|_2^2}{\left\|Y - \hat{Y}\right\|_2^2} \right) \quad [dB] \tag{4}$$

where Y input signal, $\hat{Y}$ output signal and related transfer block as shown figure 3.

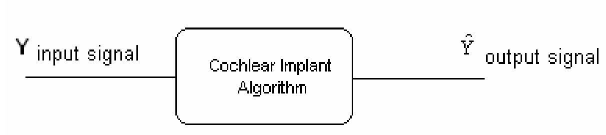

Figure 3. Block diagram of relation between Y and $\hat{Y}$

We assume $\hat{Y}$ approximately equals original signal X therefore Y - $\hat{Y}$ equals N.

Generally, the form of noise is classified as white noise and colored noise.

IV.I White Noise

Pure white noise is a theoretical concept, since it would need to have infinite power to cover an infinite range of frequencies. Furthermore, a discrete-time signal by necessity has to be band-limited, with its highest frequency less than half the sampling rate. A more practical concept is band-limited white noise, defined as a noise with a flat spectrum in a limited bandwidth. The spectrum of band-limited white noise with a bandwidth of B Hz is given by

$$P_{NN}(f) = \begin{cases} \sigma^2, & |f| \le B \\ 0, & \text{otherwise} \end{cases} \tag{5}$$

IV.II Cloured Noise

Although the concept of white noise provides a reasonably realistic and mathematically convenient and useful approximation to some predominant noise processes encountered in telecommunications systems, many other noise processes are nonwhite. The term 'coloured noise' refers to any broadband noise with a nonwhite spectrum. For example most audio frequency noise, such as the noise from moving cars, noise from computer fans, electric drill noise and people talking in the background, has a nonwhite predominantly low-frequency spectrum. Also, a white noise passing through a channel is ‘coloured’ by the shape of the frequency response of the channel. Two classic varieties of coloured noise are so-called ‘pink noise’ and ‘brown noise’, shown in Figures 4 and 5.

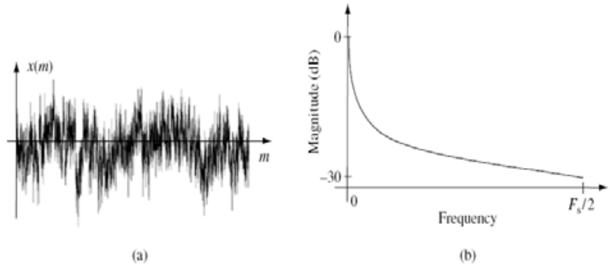

Figure 4 (a) A pink noise signal and (b) its magnitude spectrum

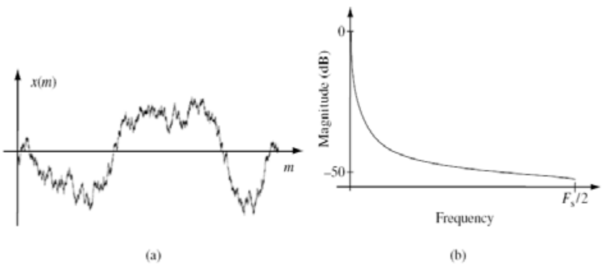

Figure 5 (a) A brown noise signal and (b) its magnitude spectrum

IV. RESULTS AND DISCUSSION

In this study, the CI speech algorithms have been tested on noisy signals. The speech signals corrupted by additive colored noise with various signal to noise ratio (SNR) were used for performance evaluation. Speech signals have the 16 KHz sampling rate. We used several types coloured noise which are F-16 cockpit noise, factory noise and Volvo noise that generally used researches on cochlea in TIMIT database.

We separate the signal into 22 subbands then apply CIS and N-of-M cochlear speech signal strategies. Afterwards, we calculate new SNR values using below equation [4].

We use custom program using MATLAB 6.5 for CI speech strategies and computing SNR values. Noise levels for different five sentences were listed in Table 1. These sentences which used are “Good service should be rewarded by big tips”, “Draw every outer line first, then fill in the interior”, “The high security prison was surrounded by barbed wire”, “The fifth jar contains big, juicy peaches”, and “Biblical scholars argue history”.

Table 1. Noise levels for different five sentences

Sentence	F16 cockpit Noise CIS	F16 cockpit Noise N-of-M	Factory Noise CIS	Factory Noise N-of-M	Volvo Noise CIS	Volvo Noise N-of-M
1	-0,4867	1,3196	-0,3431	1,4968	-0,2665	1,5653
2	-0,3869	1,7894	-0,2392	1,9598	-0,2067	1,9763
3	-0,4807	1,5776	-0,2905	1,7685	-0,2737	1,8250
4	-0,9193	0,9376	-0,7454	1,1981	-0,7021	1,2667
5	-0,2739	2,0085	-0,2412	2,0706	-0,1729	2,0736

Figure 7, Figure 8 and Figure 9 give the SNR values for F16 cockpit noise, Volvo noise and factory noise, respectively. According to these figures, it is clearly seen that N-of-M strategy gives better results than CIS strategy. In addition to, average SNR values are expounded for comparison (Figure 10), and Figure 11 show the SNR for repeated many experiments.

V. CONCLUSION

In this paper we have addressed the issue of comparison of two different strategies for noisy speech. To summarize, we used five different sentence and three different noise level f16 cockpit noise, factory noise, Volvo noise respectively in our study. We investigated noise resistance different speech strategies that include CIS and N-of-M strategies and compared each others for CI. We show that N-of-M strategy gives better results than CIS strategy. These results prove that the N-of-M strategy is more efficient in a noisily environment.

REFERENCES

1. P. C. Loizou, Introduction to Cochlear Implants, IEEE Eng. Med. Biol. Magazine, pp. 32–41, January 1999.
2. P. C. Loizou, Signal - Processing Techniques for Cochlear Implants, IEEE Eng. Med. Biol. Magazine, pp. 34–46, May 1999.
3. Boëx C, Pelizzone M, Montandon P., Speech Recognition with a CIS strategy for the Ineraid Multichannel Cochlear Implant, American Journal of Otology, 17, 61-68, 1996.
4. B. S. Wilson, D. T. Lawson, M. Zerbi et al., New Processing Strategies in Cochlear Implantation, The American Journal of Otology, 1995 16, 669-675.
5. B. S. Wilson, C. C. Finley, D. T. Lawson et al., Better Speech Recognition with Cochlear Implants, Nature, 352:236-238, 1991.
6. B. S. Wilson, C. C. Finley, D. T. Lawson et al., Design and Evaluation of Continuous Interleaved Sampling (C1S) Processing Strategy for Multi Channel Cochlear Implants, J. of Rehab, Research and Development, 30:110-116, 1993.
7. K. Niet, N. Lant, S. Gao, Implementation of CIS Speech Processing Strategy for Cochlear Implants by using Wavelet Transform, Proceedings of ICSP, 1998.

8. Wang K and Shamma **SA.** Auditory analysis of Spectro-temporal information in acoustic signal. IEEE Eng. in Med.& Bio., **1995; 14(2): 186-1 94**
9. Goldstein MH ,Jr. Auditory periphery **as** speech signal processor. IEEE Eng. in Med.& Bio., **1994;** April: **186-1 96**
10. R. Lawrence, B. H. Juang, Fundamentals Speech Recognition, PTR Prentice Hall, 1993.

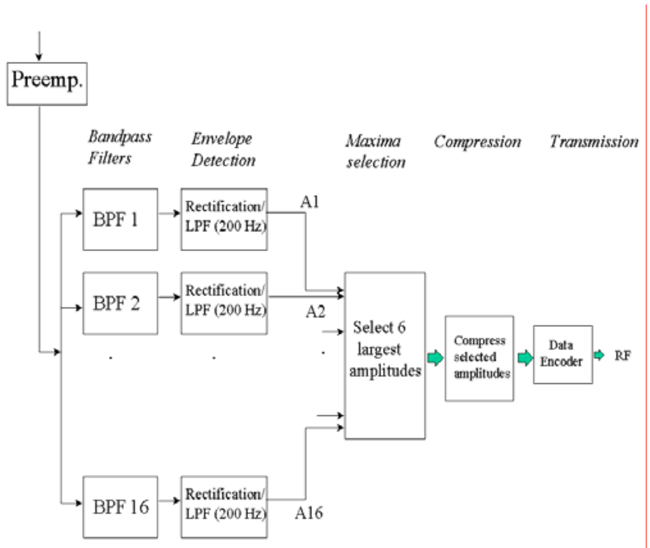

Figure 6. Block diagram of N-of-M speech processor

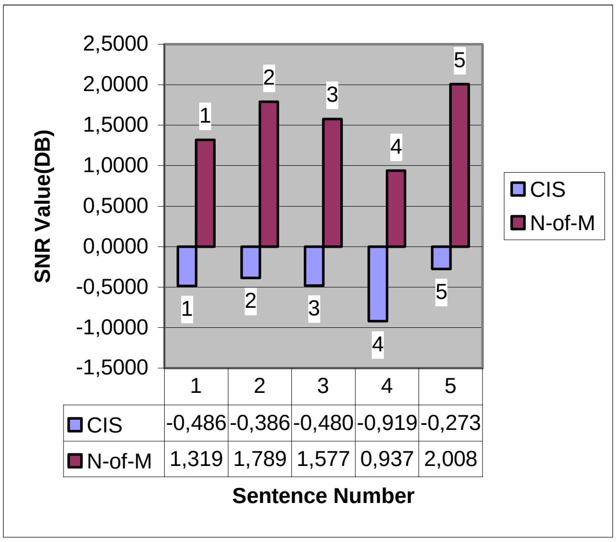

Figure 7. SNR values for F16 cockpit noise

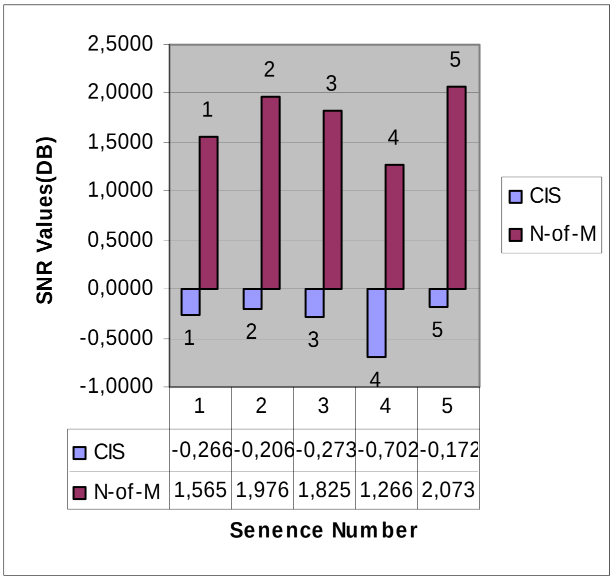

Figure 8. SNR values for Volvo noise

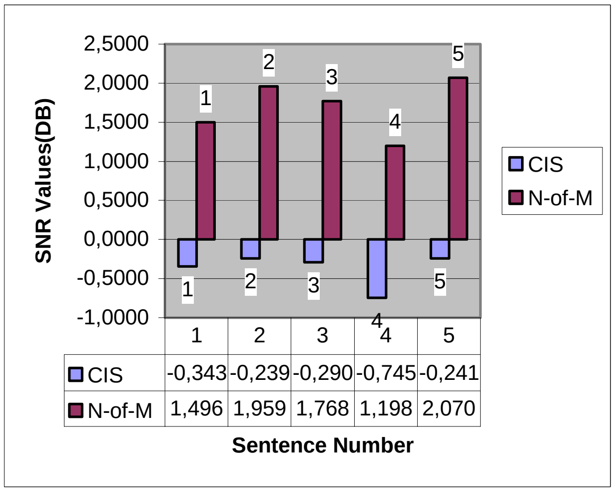

Figure 9. SNR values for factory noise

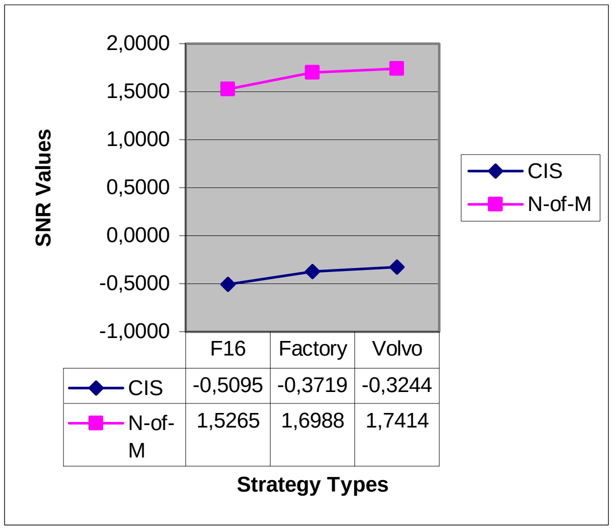

Figure 10. Average SNR values

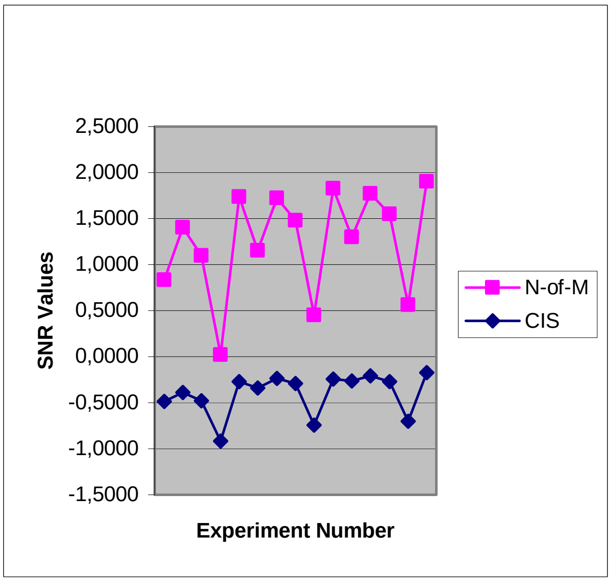

Figure 11. SNR values for both strategies