A neural network model for optimizing vowel recognition by cochlear implant listeners

Chung-Hwa Chang , Gary T. Anderson, *Member IEEE,* and Philipos C. Loizou, *Member IEEE*

Abstract**-- Due to the variability in performance among cochlear implant (CI) patients, it is becoming increasingly important to find ways to optimally fit patients with speech processing strategies. This paper proposes an approach based on neural networks, which can be used to automatically optimize the performance of CI patients. The neural network model is implemented in two stages. In the first stage, a neural network is trained to mimic the CI patient's performance on the vowel identification task. The trained neural network is then used in the second stage to adjust a free parameter to improve vowel recognition performance for each individual patient. The parameter examined in this study was a weighting function applied to the compressed channel amplitudes extracted from a 6-channel Continuous Interleaved Sampling (CIS) strategy. Two types of weighting functions were examined, one which assumed channel interaction and one which assumed no interaction between channels. Results showed that the neural network models closely matched the performance of 5 Med-El/CIS-Link implant patients. The resulting weighting functions obtained after neural network training improved vowel performance, with the larger improvement (4%) attained by the weighting function which modeled channel interaction.**

Index terms**—neural networks for cochlear implants, cochlear implants.**

I. INTRODUCTION

There are presently several million people in the United States suffering from profound hearing loss, and for years they had to rely on conventional hearing aids. Conventional hearing aids, however, provide little, if any, benefit for profoundly deaf patients with damaged sensory cells. Today, a prosthetic device, called cochlear implant, can be implanted in the inner ear and can restore partial hearing to profoundly deaf people. Some individuals with implants can now communicate without lip-reading, and some can communicate over the telephone. Part of the success of cochlear implants can be attributed to the signal processing techniques developed over the years for extracting electrical stimuli from speech [1][2].

Some of these techniques are aimed at preserving waveform information. They process the speech signal through a small number of frequency bands (4-8), and (simultaneously) stimulate the electrodes with analog amplitude information extracted from the bandpass-filtered waveforms. Such a technique was originally used in the Ineraid device (now discontinued) and is currently employed in the Clarion device [2]. Other speech processing strategies are aimed at preserving speech envelope information. Speech is processed through a small number of bands (8-12), and the envelopes of the filtered waveforms are extracted through rectification and low-pass filtering (200-400 Hz). The envelope amplitudes are then used to modulate biphasic pulses and stimulate the electrodes in an interleaved (i.e., non-simultaneous) fashion. This strategy, known as the Continuous Interleaved Sampling (CIS) strategy [3], is currently employed in the Med-El, Clarion and Nucleus-24 devices. And, finally, other speech processing strategies are aimed at preserving spectral information, such as spectral peaks. The SPEAK strategy (used in the Nucleus device) for instance processes speech through a large number of frequency bands (16-20), and extracts the envelopes of the filtered waveforms using rectification and low-pass filtering. Of the 20 extracted envelope amplitudes, only the 6-10 maximum amplitudes are used for stimulation in each cycle [4]. The 6-10 maximum amplitudes are used to modulate biphasic pulses and stimulate the electrodes in an interleaved (i.e., non-simultaneous) fashion.

Most of the above speech processing strategies can be configured using a multitude of parameters, which can be easily modified by the audiologist using existing fitting software. In the SPEAK strategy (Nucleus 22), for example, one can change the pulse width, the number of maxima selected, the pulse rate, the filter allocation table, etc. In the CIS strategy, currently supported by all implant devices (Nucleus 24, Med-El and Clarion) in the United States, one can change the pulse width, the pulse rate, the electrode stimulation order and the compression function that maps the acoustical input to electrical output. A series of studies [5-8] has shown that the CIS parameters (e.g., pulse rate, pulse width, etc.) can be varied to optimize individual subjects' performance. This optimization, however, can be time consuming, given the large number of speech processing parameters available. It is therefore becoming

C. Chang and G. Anderson are with the Applied Science Department at the University of Arkansas at Little Rock, and P. Loizou is with the Electrical Engineering Department, University of Texas at Dallas, Richardson, TX 75083-0688 (e-mail: loizou@utdallas.edu).

increasingly more important to identify the set of parameters that are most likely to affect speech recognition. This is an extremely important issue particularly for fitting implant patients, since in a clinical setting the audiologists cannot devote too much time to select these parameters.

The main objective of the present study is to develop a method to automatically optimize the speech processing parameters for a given implant patient. The parameter examined in this study was the weighting function applied to the CIS channel amplitudes. The main question addressed here is: "Can the channel amplitudes, extracted using the CIS strategy, be optimally adjusted (or weighted) for a particular patient to improve speech recognition?" To answer that question, we trained a neural network to mimic the patient's performance on the vowel identification task. The trained neural network model of the patient was then used to adjust the channel amplitudes to optimize the subject's performance. Results are presented using vowel identification data previously collected for 5 Med-El/CIS-Link cochlear implant patients [9].

II. MODELING VOWEL PERCEPTION BY USERS OF MULTICHANNEL COCHLEAR IMPLANTS

Despite the benefits obtained with current CI speech-processing strategies, our understanding of the perceptual and cognitive mechanisms employed by CI users for speech understanding is very limited. We still do not know, for example, how CI users utilize the limited (and impoverished) spectral information provided by the implant to understand speech. Understanding the underlying perceptual mechanisms employed by CI users will help us (1) develop better speech processing strategies, and (2) optimally fit CI users. A mathematical modeling approach was taken by Svirsky [10] to explain how CI patients correctly identify some vowels and how they misidentify others. In Svirsky's model, the CI user's response to a particular sound stimulus is represented as a point in a multidimensional space and is modeled with a multi-dimensional Gaussian distribution. Overlap between the distributions in the multidimensional space suggests that the CI user will make an identification error. The inputs to Svirsky's model were the ratio of channel amplitudes and the formant frequencies, while the varying parameter was the just-noticeable differences (jnd) of the formant frequencies. In practice, the jnds can be measured for each subject using psychophysical methods.

In this paper, we propose a different approach for modeling vowel perception by CI users. We believe that the proposed approach is simpler as it does not require extensive psychophysical measurements as input. It only requires the input-output response matrix of each CI user. It also has the added flexibility that it can be used to optimize individual CI user's performance (see following section III). An adaptive neural network structure (Figure 1a) is used to model vowel perception by cochlear implant users. Such an adaptive structure has been used in the past for system identification applications [11], where the input and output signals of a linear system (e.g. FIR or IIR) are known and one needs to identify the unknown system that generated the output signal. In our application, the unknown system was the patient's auditory system, which maps the acoustic signal into phonetic categories (the 11 vowels, in our case). The inputs to this system are the vowels processed through the CIS strategy, and the outputs of the system are the responses of the patients to the input vowel stimuli. A neural network is used in this study as a crude model of the subject's auditory system. The CIS channel amplitudes are used to train the neural network so that the output of the neural network matches the response of the patient.

A. *Speech material*

Data for training and testing were taken from the vowel identification task in the Loizou *et al.* study [9] with cochlear implant patients. The stimuli for the vowel identification test consisted of the vowels in the words 'had', 'hod', 'head', 'hayed', 'heard', 'hid', 'heed', 'hoed', 'hood', 'hud', and 'who'd' produced by 11 men and 11 women. These words were taken from recordings made by Hillenbrand *et al.* [12]. The vowel stimuli were presented to 5 cochlear implant patients for identification in ten sessions [9]. The vowels were completely randomized in each session and presented to the patients three times. Overall, a total of 660 vowels (= 10 sessions x 11 vowels x 3 repetitions x 2 speaker groups) were presented for identification. The patient's responses to each stimulus (in terms of confusion matrices) were saved and used in this study to train the neural network models.

We processed the above vowel data through an implementation of the 6-channel CIS strategy and used the output CIS channel amplitudes as input to the neural network. To better represent the dynamic nature of vowels, we computed the channel amplitudes for three speech segments of the vowel extracted at: 10% (onset), 50% (middle) and 80% (offset) of the duration of the vowel. The channel amplitudes for each of the three vowel segments were computed as follows. The signal was first processed through a pre-emphasis filter (2000 Hz cutoff) with a 3 dB/octave roll-off, and then bandpassed into 6 frequency bands using sixth-order Butterworth filters. The center frequencies of the bandpass filters were 393, 639, 1037, 1685, 2736 and 4443 Hz, and the 3-dB bandwidths were 187, 304, 493, 801, 1301, and 2113 Hz respectively. The envelope of the signal was extracted by full-wave rectification and low-pass filtering (second-order Butterworth) with a 400 Hz cutoff frequency. The channel amplitudes were estimated by computing the root mean-square (rms) energy of the envelopes within a 10-ms window. Due to the limited electrical dynamic range, the envelope amplitudes, *s(n)*, were compressed using a logarithmic function of the form:

$$A(n) = C\log(s(n)) + D \quad n = 1,2,\cdots,6 \qquad (1)$$

where C and D are constants chosen to ensure that the electrical amplitudes fall between the threshold and most-comfortable levels. A total of 18 channel amplitudes (6 for the onset of the vowel, 6 for the middle portion of the vowel, and 6 for the offset of the vowel) were used as input to the neural network for each vowel.

B. *Neural Network Training and Results*

A three-layer feed-forward neural network was trained for each patient using the back-propagation learning algorithm [13]. The neural network had 18 nodes in its input layer (3 vowel segments $\times$ 6 channels), 3 nodes in its hidden layer and 11 nodes in the output layer (one for each vowel). The sigmoid function was used as the activation function in each layer. The neural network weights **W** (Figure 1a) were adjusted using a gradient descent algorithm to minimize the error between the patient's response to the vowel stimuli and the neural network's response. During training, the neural network's output was set to 0.9 to indicate the patient's response, while all other neural network outputs were set to 0.1. This range of 0.1-0.9 of neural network outputs was chosen because it often results in moderately-valued network weights, thus allowing for relatively fast training [14].

A total of 660 vowel stimuli were used for testing and training, with each vowel stimulus represented by 18 channel amplitudes. The vowel data were divided equally into three sets, with 220 stimuli in each set. The neural network was trained on two of the sets, and tested on the third. This was repeated two times so that all three sets were used for testing. So, in the first run data sets 2 and 3 were used for training and data set 3 was used for testing. In the second run, data sets 1 and 3 were used for training and data set 2 was used for testing. In the third run, data sets 1 and 2 were used for training and data set 3 was used for testing. The results were averaged across the three data sets.

Figure 2 shows the comparison between patient's performance and the trained network performance. As can be seen, the overall prediction rate of each of the networks closely matches that of the patient they were trained to model. It is also important to note that the trained network produced roughly the same response as the patient did for a given vowel. Note, however, that because the patient's responses were sometimes not consistent when repeatedly presented the same vowel, the neural network could not be trained to *completely* mimic the patient's performance. Tables 1 and 2 show, for example, the confusion matrices obtained by patient S5 and the neural network respectively. As shown in Table 2, the neural network produced roughly the same error patterns as patient S5. For instance, the dominant confusion for the input word "hod" was "had" for both patient and neural network. Similarly, the two dominant confusions for the input word "who'd" were "hoed" and "hood". This particular patient was most difficult to model because of the patient's inconsistency in identifying certain vowels reliably (e.g., see patient's response to the input word "hid" in Table 1). Despite the occasionally random responses of the patient, the model was able to capture the dominant confusion errors.

III. ADJUSTING THE CHANNEL AMPLITUDES FOR OPTIMIZED PERFORMANCE

After the neural network was trained to model the individual patient's performance, the network was fixed and used in the second stage (Figure 1b) to adjust the channel amplitudes for optimized performance. This was accomplished using a learning algorithm that minimized the error between the true response to an input vowel stimulus and the neural network's output (Figure 1b).

As shown in Figure 1b, **W** are the network weights estimated in stage 1, and *A(n)* are the compressed channel amplitudes which are adjusted by the weighting function *K(n)* to improve the patient's performance. The optimal (in the least-squares sense) weighting function *K(n)* can be estimated adaptively as follows. The network outputs $y_{net}(n)$ are written as:

$$y_{net}(n) = f(x(n),W) = f(A(n)K(n),W) \qquad (2)$$

where **W** are the fixed (as estimated in the first stage) network weights and *f(.)* is the sigmoid function. With the network inputs and weights both fixed, improvements in the network outputs are made by adjusting the weighting function *K(n)* in a gradient-descent manner, i.e.,:

$$\Delta K(n) = -\mu \frac{\partial J}{\partial k(n)} \qquad (3)$$

where J is a cost function used to evaluate network performance and μ is an empirical constant that determines the learning rate. The cost function, J, is the mean squared-error of the system, and is computed as:

$$J = \frac{1}{2}(d(n) - y_{net}(n))^2 = \frac{1}{2}(d(n) - f(A(n)\times K(n),W)^2 \qquad (4)$$

where d(n) is the true response to a given vowel. Finally, the weighting function *K(n)* is adjusted according to (3) to minimize the above cost function.

A. *Weighting functions*

Two different types of weighting functions *K(n)* were considered, one which assumed no electrode interaction and one which assumed electrode interaction. When an array of multiple electrodes is inserted into the cochlea, the individual electrodes may or may not stimulate different populations of neurons depending on, among other factors, the distance between electrodes. If multiple neuron populations are stimulated, then we say that there is electrode interaction, which causes a summation of electrical fields. Such an interaction might distort the acoustic information presented to the electrodes, and is therefore not desirable. One of the weighting functions considered in this study accounted for the presence of electrode interactions, while the other did not.

The first weighting function *K(n)* weighs each channel amplitude *A(n)* independent of the other amplitudes, i.e.,

$$X(n) = K(n)A(n) \quad n = 1,2,\cdots,6 \tag{5}$$

where *X(n)* are the weighted channel amplitudes. Written in matrix form:

$$\begin{bmatrix} K(1) & 0 & 0 & 0 & 0 & 0 \\ 0 & K(2) & 0 & 0 & 0 & 0 \\ 0 & 0 & K(3) & 0 & 0 & 0 \\ 0 & 0 & 0 & K(4) & 0 & 0 \\ 0 & 0 & 0 & 0 & K(5) & 0 \\ 0 & 0 & 0 & 0 & 0 & K(6) \end{bmatrix} \begin{bmatrix} A(1) \\ A(2) \\ A(3) \\ A(4) \\ A(5) \\ A(6) \end{bmatrix} = \begin{bmatrix} X(1) \\ X(2) \\ X(3) \\ X(4) \\ X(5) \\ X(6) \end{bmatrix} \tag{6}$$

The above scheme assumes no interaction between the six electrodes; that is, the electrical stimulation for electrode 1, X(1), only depends on the channel amplitude A(1), the electrical stimulation for electrode 2, X(2), only depends on the channel amplitude A(2), and so forth.

The second weighting function considers possible interaction between the electrodes, and therefore weighs the channel amplitudes as follows:

$$X(n) = \sum_{i=1}^{6} K(n,i)A(i) \qquad n = 1,2,\cdots,6 \tag{6}$$

or, in matrix form

$$\begin{bmatrix} K(1,1) & K(1,2) & K(1,3) & K(1,4) & K(1,5) & K(1,6) \\ K(2,1) & K(2,2) & K(2,3) & K(2,4) & K(2,5) & K(2,6) \\ K(3,1) & K(3,2) & K(3,3) & K(3,4) & K(3,5) & K(3,6) \\ K(4,1) & K(4,2) & K(4,3) & K(4,4) & K(4,5) & K(4,6) \\ K(5,1) & K(5,2) & K(5,3) & K(5,4) & K(5,5) & K(5,6) \\ K(6,1) & K(6,2) & K(6,3) & K(6,4) & K(6,5) & K(6,6) \end{bmatrix} \begin{bmatrix} A(1) \\ A(2) \\ A(3) \\ A(4) \\ A(5) \\ A(6) \end{bmatrix} = \begin{bmatrix} X(1) \\ X(2) \\ X(3) \\ X(4) \\ X(5) \\ X(6) \end{bmatrix} \tag{7}$$

In this case, the electrical pulse amplitude *X(i)* of each channel is a function of the combined set of channel amplitudes. The electrical stimulation for electrode 1, for example, not only depends on the channel amplitude A(1), but also depends on the channel amplitudes A(2), A(3), A(4), A(5) and A(6) because of the assumed interaction between electrodes.

B. *Results and Discussion*

The performance of the neural network models using the different weighting function strategies is shown in Table 3. The initial performance of the neural network model before application of the weighting function is given in the row labeled 'No weighting'. Results for the weighting function described by Eq. 5 are given in the row labeled 'Weighting vector', while results for the weighting function described by Eq. 6 are given in the row labeled 'Weighting matrix'. As can be seen, adjusting the channel amplitudes yielded a small improvement in performance, ranging from 1% to 4%, with the largest improvement obtained using the weighting matrix. Although a small improvement, paired-samples (2-tailed) t-tests showed that the improvements made by the two weighting functions were statistically significant [weighting vector: $t(4)=-4.18$, $p=0.014$; weighting matrix: $t(4)=-4.99$, $p=0.008$]. As shown in Table 3, the amount of improvement in performance obtained was not the same for all subjects. The neural networks trained for subjects P2 and P5, for instance, improved performance by 4 percentage points, while the neural network trained for subject P3 improved performance by only 1 percentage point. We suspect that this is partially due to the inability of some subjects to identify the vowels consistently. That is, some subjects did not identify vowels consistently when presented the same vowel several times. Such inconsistencies make it extremely difficult to reliably train a neural network.

The weighting functions obtained for each subject are given in Appendix A. The first table lists the weighting functions (Eq. 5), which assume no electrode interaction. As can be seen, the values of the weighting functions for all subjects are close to 1, which explains why the improvement in performance was small. The fact that the weighting values were nearly 1 suggests that the compressed channel amplitudes were already "optimized" in some sense. Therefore, making the assumption that there is no electrode interaction imposes constraints on the amount of improvement that one can obtain by weighting the channel amplitudes. The second table lists the weighting matrices (Eq. 7), which assume electrode interaction. The weighting matrices are nearly diagonal with some variability in the off-the-diagonal values across subjects, suggesting that each subject exhibited a different pattern of channel interaction. The off-diagonal values are small, but not zero, suggesting the existence of channel interaction. Therefore, greater improvements in performance can be achieved if the assumption of channel interaction is made, which most likely better represents the situation in existing electrode arrays.

As mentioned above, the improvement in vowel recognition obtained by the neural network models was small. We believe that further improvement in performance could be obtained by: (1) increasing the amount of training data, and (2) modeling the male and female speakers separately. Increasing the amount of training data, either by sampling the vowels at more than three locations within the vowel or by including more speakers, will most likely improve the match between the neural network's response and the patients' response. Given the differences in formant frequencies between male and female speakers [12], it also seems reasonable to train different neural network models for male and female speakers. Further investigation is therefore needed to verify whether using more training data and male/female network models will improve vowel recognition performance.

The next step in the proposed approach is to verify whether cochlear implant subjects show similar improvements in vowel recognition as obtained by the trained neural networks when the channel amplitudes are weighted. We have not tested cochlear implant subjects yet with the weighted channel amplitudes, and this is left for future studies.

V. SUMMARY AND CONCLUSIONS

This paper presented a method based on neural networks for automatically adjusting channel amplitudes (estimated using the CIS strategy) to improve cochlear implant subject's performance on vowel identification. A neural network was first trained to model the patient's performance on vowel identification. Once the neural network was trained, it was used in place of the patient to adjust the channel amplitudes via a gradient-descent learning algorithm. Results showed that this adjustment, accomplished by weighting the channel amplitudes, yielded a small, yet significant, improvement in neural network performance. The type of weighting function applied to the channel amplitudes seemed to make a difference. Relatively larger improvements in performance were obtained using a weighting matrix that assumed that there was channel interaction compared to a weighting vector that assumed no interaction between channels.

The proposed approach seemed to be promising for optimizing speech-processing parameters associated with the CIS or SPEAK strategies. This is becoming a very important issue for fitting cochlear implant subjects, given the variability in performance between cochlear implant subjects caused by differences in etiology of hearing loss, neural survival patterns, etc. We hope that such an approach will be used in the future by clinicians to optimize patient's performance, and therefore make all cochlear implant patients "better-performing" patients.

ACKNOWLEDGMENTS

This research was supported in part by Grant No. R01 DC03421 awarded to the third author (PCL) from the National Institute of Deafness and other Communication Disorders, NIH.

REFERENCES

[1] Loizou, P. "Introduction to cochlear implants," *IEEE Eng. Med. Biol. Mag.*, vol. 18, no. 1, pp. 32-42, 1999.

[2] Loizou, P. "Signal-processing techniques for cochlear implants," *IEEE Eng. Med. Biol. Mag.*, vol. 18, no. 3, pp. 34-46, 1999.

[3] Wilson, B., Finley, C., Lawson, D., Wolford, R., Eddington, D. and Rabinowitz, W. "Better speech recognition with cochlear implants," *Nature*, vol. 352, pp. 236-238, 1991.

[4] McDermott, H., McKay, C. and Vandali, A. "A new portable sound processor for the University of Melbourne/Nucleus Limited multielectrode cochlear implant," *J. Acoust. Soc. Am.*, vol. 91, pp. 3367-3371, 1992.

[5] Wilson, B., Lawson, D. and Zerbi, M. "Speech Processors for Auditory Prostheses," NIH Project N01-DC-2-2401, Fifth Quarterly Progress Report, 1993.

[6] Wilson, B., Lawson, D. and Zerbi, M. "Advances in coding strategies for cochlear implants," *Advances in Otolaryngology – Head and Neck Surgery*, vol. 9, pp. 105-129, 1995.

[7] Dorman, M. and Loizou, P. "Changes in speech intelligibility as a function of time and signal processing strategy for an Ineraid patient fitted with CIS processors," *Ear Hearing*, vol. 18, no. 2, pp. 147-155, 1997.

[8] P. Loizou, O. Poroy and M. Dorman "The effect of parametric variations of cochlear implant processors on speech understanding," *J. Acoust. Soc. Am.*, vol. 108, no. 2, pp. 790-802, 2000.

[9] P. Loizou, M. Dorman, and V. Powell "The recognition of vowels produced by men, women, boys, and girls by cochlear implant patients using a six-channel CIS processor", *J. Acoust. Soc. Am.*, vol. 103, pp. 1141-1149, 1998.

[10] Svirsky, M. "Mathematical modeling of vowel perception by users of analog multichannel cochlear implants: Temporal and channel-amplitude cues," *J. Acoust. Soc. Am.*, vol. 107, no. 3, pp. 1521-1529, 2000.

[11] Widrow, B. and Stearns, S. *Adaptive Signal Processing*, Englewood Cliffs: Prentice Hall, 1985.

[12] J. Hillenbrand, L. A. Getty, M. J. Clark, and K. Wheeler "Acoustic Characteristics of American English Vowels", *J. Acoust. Soc. Am.*, vol. 97, no. 5, pp. 3099-111, 1995.

[13] Haykin, S., *Neural Networks: A Comprehensive Foundation*, Englewood Cliffs: Prentice Hall, 1999.

[14] Masters, T. *Practical neural network recipes in C++*, San Diego: Academic Press, p. 82, 1993.

APPENDIX

In this Appendix we list the weighting functions obtained for each patient after network training.

The following table lists the weighting function *K(n)*, n=1,2, .., 6, (Eq. 5) obtained after network training for each subject. Only the weighting functions obtained for testing set 2 are displayed. The weighting functions obtained for the other two testing sets were very similar.

Subject	K(1)	K(2)	K(3)	K(4)	K(5)	K(6)
S1	1.0012	0.9990	0.9998	0.9991	1.0012	0.9993
S2	0.9995	0.9981	0.9994	0.9995	0.9999	0.9985
S3	1.0001	1.0002	0.9994	0.9998	1.0002	0.9998
S4	1.0000	0.9999	1.0001	0.9999	1.0003	0.9998
S5	1.0002	0.9999	1.0005	1.0020	0.9976	1.0008

The following table lists the weighting function $K(i, j)$, i,j=1,2, .., 6, (Eq. 6) obtained after network training for each subject. Only the weighting functions $K(i, j)$ obtained for testing set 2 are displayed. The weighting functions obtained for the other two testing sets were very similar.

Subject	**Weights**						
S1		K(n,1)	K(n,2)	K(n,3)	K(n,4)	K(n,5)	K(n,6)
	K(1,m)	0.99952	0.00191	-0.00057	0.00161	-0.00156	-0.00155
	K(2,m)	-0.00064	0.99964	-0.00001	-0.00112	-0.00023	0.00035
	K(3,m)	-0.00133	-0.00096	1.00067	-0.00017	0.00089	0.00106
	K(4,m)	0.00152	0.00109	-0.00140	0.99984	-0.00036	-0.00083
	K(5,m)	0.00144	0.00125	-0.00073	-0.00214	1.00001	-0.00043
	K(6,m)	0.00000	-0.00237	0.00071	0.00098	0.00068	1.00028
S2		K(n,1)	K(n,2)	K(n,3)	K(n,4)	K(n,5)	K(n,6)
	K(1,m)	0.99745	0.00106	-0.00291	-0.00258	0.00451	0.00294
	K(2,m)	0.00157	1.00081	-0.00092	-0.00041	-0.00126	-0.00100
	K(3,m)	-0.00229	-0.00534	1.00547	0.00126	-0.00520	-0.00329
	K(4,m)	0.00256	0.00137	-0.00421	0.99632	-0.00071	-0.00088
	K(5,m)	-0.00051	0.00015	0.00168	0.00029	0.99986	0.00175
	K(6,m)	0.00078	-0.00002	0.00038	0.00010	0.00058	1.00043
S3		K(n,1)	K(n,2)	K(n,3)	K(n,4)	K(n,5)	K(n,6)
	K(1,m)	0.99193	0.00137	0.00516	0.00324	0.00095	-0.00455
	K(2,m)	-0.00352	0.99246	0.00450	0.00316	0.00736	-0.00114
	K(3,m)	-0.00052	0.00474	1.00113	0.00448	-0.00343	-0.00231
	K(4,m)	0.00797	0.00069	-0.00259	0.99874	-0.00313	0.00145
	K(5,m)	-0.00212	0.00451	-0.00297	-0.00152	1.00172	0.00123
	K(6,m)	-0.00054	-0.00260	-0.00027	-0.00113	0.00191	1.00022
S4		K(n,1)	K(n,2)	K(n,3)	K(n,4)	K(n,5)	K(n,6)
	K(1,m)	0.99898	-0.00029	0.00044	0.00069	0.00012	-0.00004
	K(2,m)	0.00116	1.00045	-0.00062	-0.00110	-0.00011	0.00009
	K(3,m)	-0.00106	-0.00060	1.00101	0.00094	0.00019	-0.00004
	K(4,m)	0.00110	0.00089	-0.00040	0.99950	0.00000	0.00025
	K(5,m)	-0.00028	-0.00063	-0.00046	0.00012	0.99986	0.00002
	K(6,m)	-0.00008	0.00008	-0.00009	0.00009	0.00001	0.99966
S5		K(n,1)	K(n,2)	K(n,3)	K(n,4)	K(n,5)	K(n,6)
	K(1,m)	0.99436	0.00698	0.00784	0.00020	-0.00292	-0.00371
	K(2,m)	0.00286	0.98253	0.00137	0.01163	-0.00002	0.00033
	K(3,m)	0.00691	0.00237	0.98833	0.00137	0.00266	0.00304
	K(4,m)	0.00038	0.00361	-0.00522	0.98996	0.00429	0.00451
	K(5,m)	-0.01107	-0.00677	0.00557	0.00263	0.99693	0.00384
	K(6,m)	0.00312	0.00593	0.00553	-0.00160	-0.00647	0.97952

TABLES

Table 1. The confusion matrix (input-output response matrix) obtained by patient S5 on the vowel identification task. The numbers indicate percent correct scores rounded to the nearest integer.

	had	**hod**	**head**	**hayed**	**heard**	**hid**	**heed**	**hoed**	**hood**	**hud**	**who'd**
had	40	12	22	2	8	0	0	0	3	13	0
hod	70	15	5	0	5	0	0	0	0	5	0
head	5	13	28	7	18	3	0	0	3	22	0
hayed	0	0	0	72	0	0	27	0	0	0	2
heard	0	0	0	13	75	3	5	0	2	2	0
hid	0	0	3	7	7	72	2	0	8	2	0
heed	0	0	0	5	0	5	88	0	0	0	2
hoed	0	3	0	0	0	0	0	82	5	0	10
hood	0	2	0	0	13	10	0	0	50	18	7
hud	2	8	7	0	8	0	0	2	7	67	0
who'd	0	0	0	7	0	2	8	13	17	0	53

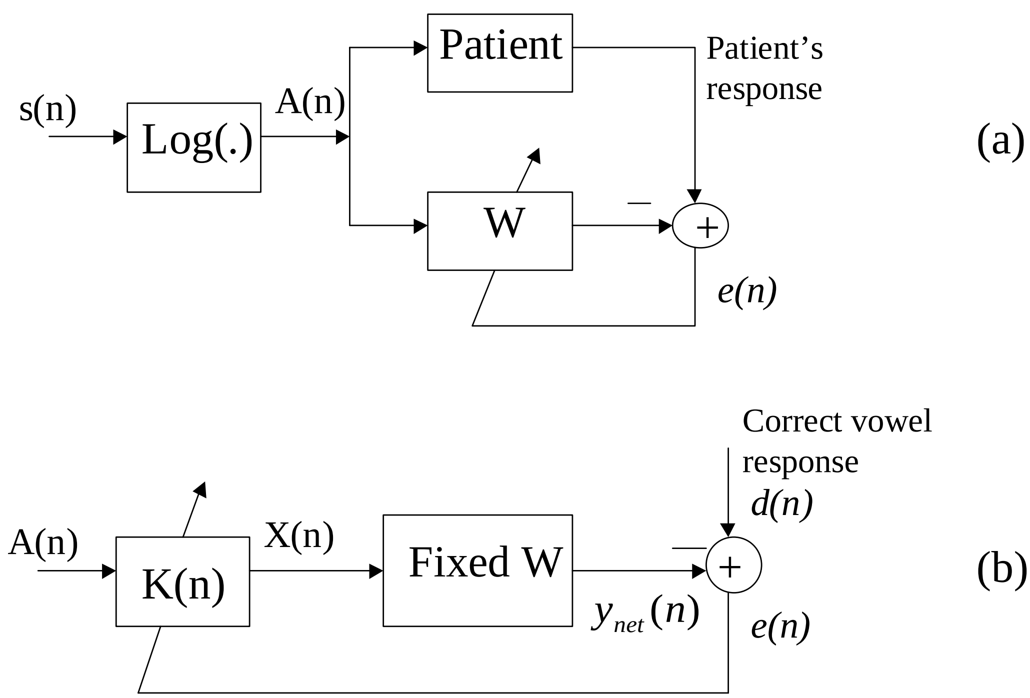
Table 2. The confusion matrix (input-output response matrix) obtained by the neural network trained to model the performance of patient S5. The numbers indicate percent correct scores rounded to the nearest integer.

	had	**hod**	**head**	**hayed**	**heard**	**hid**	**heed**	**hoed**	**hood**	**hud**	**who'd**
had	84	0	0	0	0	0	0	0	0	16	0
hod	54	0	8	7	8	0	0	0	0	23	0
head	7	0	0	17	17	0	0	0	17	42	0
hayed	0	0	0	64	18	0	5	0	0	5	8
heard	3	0	0	11	59	15	4	0	4	4	0
hid	0	0	0	5	0	60	15	0	10	10	0
heed	0	0	0	16	4	4	72	0	0	0	4
hoed	0	0	0	10	0	0	0	71	5	0	14
hood	0	0	0	5	21	11	0	5	42	5	11
hud	17	0	3	7	5	3	3	0	10	52	0
who'd	0	0	0	0	0	0	0	15	8	0	77

Table 3. Comparison of neural network performance, in terms of percent correct, before and after applying the weighting functions to the input channel amplitudes.

Weighting function	Patients				
	P1	**P2**	**P3**	**P4**	**P5**
No weighting	85.30%	85.91%	87.88%	69.24%	58.33%
Weighting vector	86.21%	87.27%	89.39%	70.00%	61.06%
Weighting matrix	86.52%	89.24%	91.52%	71.06%	62.58%

Figure 1. The proposed neural network structures for modeling vowel perception by CI patients (Figure a) and for optimizing patient's performance (Figure b). In Figure (a), *s(n)* are the envelope amplitudes estimated using the CIS strategy, and *A(n)* are the logarithmically compressed channel amplitudes. The neural network weights **W** are adjusted using the back-propagation algorithm to match the CI patient's response to a given input vowel stimulus. In Figure (b), *x(n)* are the channel amplitudes weighted by the function *k(n)* to optimize CI patient's performance. With the neural network weights **W** fixed (as estimated from Figure a), the weighting function *k(n)* is adjusted using a gradient-descent algorithm to minimize the error, *e(n)*, between the network output and the true vowel response.



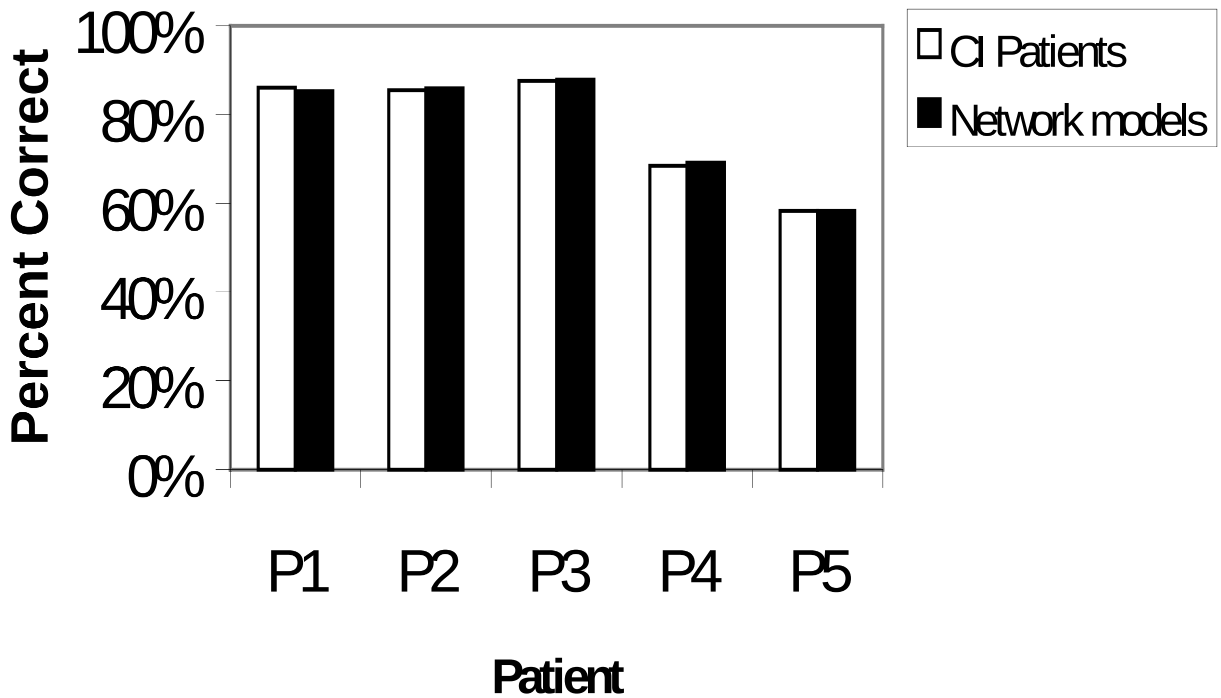


Figure 2. Comparison of performance on vowel identification by CI patients and neural network models.