

Neural Adaptation and Perceptual Learning using a Portable Real-Time Cochlear Implant Simulator in Natural Environments

Christopher J. Smalt, Thomas M. Talavage, David B. Pisoni, Mario A. Svirsky

Abstract—A portable real-time speech processor that implements an acoustic simulation model of a cochlear implant (CI) has been developed on the Apple iPhone / iPod Touch to permit testing and experimentation under extended exposure in real-world environments. This simulator allows for both a variable number of noise band channels and electrode insertion depth. Utilizing this portable CI simulator, we tested perceptual learning in normal hearing listeners by measuring word and sentence comprehension behaviorally before and after 2 weeks of exposure. To evaluate changes in neural activation related to adaptation to transformed speech, fMRI was also conducted. Differences in brain activation after training occurred in the inferior frontal gyrus and areas related to language processing. A 15-20% improvement in word and sentence comprehension of cochlear implant simulated speech was also observed. These results demonstrate the effectiveness of a portable CI simulator as a research tool and provide new information about the physiological changes that accompany perceptual learning of degraded auditory input.

Index Terms—cochlear implants, fMRI, portable media players

I. INTRODUCTION

The cochlear implant (CI) is the first example in history where a human sense has been partially restored by a neural prosthesis. Such devices can restore hearing to profoundly deaf patients by stimulating auditory neurons electrically. In other words, these patients do not hear through the normal transduction of sound, but rather, through signal picked up by a microphone, processed by electronic equipment, and then delivered directly to the listener's auditory nerve.

The use of acoustic simulation models of CIs has become quite common over the last several years. These models manipulate and transform the acoustic signal in the same way as a typical CI, emulating for a normal hearing listener what a real CI user might hear. Since it is widely recognized that there are many advantages to studying speech processing and perceptual learning in normal hearing populations, experiments with vocoded speech are often conducted in the laboratory. However, it is much more desirable and realistic

This research was supported, in part, by NIH Research Grants to Mario A Svirsky and David B Pisoni from NIDCD and funds from Indiana University School of Medicine.

C. Smalt is with the School of Electrical and Computer Engineering, Purdue University, West Lafayette, Indiana, USA csmalt@purdue.edu

T. Talavage is with the faculty of the School of Electrical and Computer Engineering, Purdue University, West Lafayette, Indiana, USA tmt@purdue.edu

D. Pisoni is with the faculty of the Department of Psychological and Brain Sciences, Indiana University, Bloomington, Indiana, USA pisoni@indiana.edu

M. Svirsky is with faculty of the New York University School of Medicine, New York, New York, USA Mario.Svirsky@NYUMC.ORG

to conduct tests that better model the real life perceptual learning in natural environments experienced by users of cochlear implants and hearing aids.

With increasing demand for portable electronic devices and smartphones, it is possible to utilize this technology to explore and implement many forms of degradation and compensation algorithms, including a simulation of a CI. One such device, developed for a Dell PDA, allows for software flexibility by providing a direct portable interface to a cochlear implant [1]. In this paper, we present a CI simulator for the iPod Touch, and consider its effectiveness as a portable platform for auditory research in natural environments.

To study the effect of chronic exposure to the portable CI simulator, word and sentence comprehension tasks were conducted to determine the change in speech recognition performance. In addition, the neurobiological correlates of the adaptation process were measured using fMRI to compare changes in neural activation before and after exposure. We hypothesized that perception of speech in post-lingually deaf CI users may be limited by the ability for speech processing pathways to adapt to a distorted input. Our aim is to study the brain's ability to adapt to transformed speech in order to aid in the optimization of CI users' speech recognition [2].

II. OVERVIEW OF CI SIMULATION

A cochlear implant device comprises of a microphone which picks up external sounds, a speech processor that processes the acoustic signal, and a surgically implanted electrode array that stimulates the auditory nerve. The process underlying simulation of CI electrical stimulation is illustrated in Figure 1, where the stages of processing up to the envelope detection step are identical to what occurs in a real cochlear implant speech processor [3], [4]. The speech processor uses a series of non-overlapping filters, which divide the acoustic signal up into a number of frequency bands corresponding to the number of electrodes in the implant. The information in each band (i.e. channel) is used to determine the amplitude and rate at which that electrode discharges electrical current. This current directly stimulates nerves in the inner ear, producing the percept of sound for a CI patient. In simulation, the sound that reaches the listener becomes degraded due to use of the same frequency band extraction techniques as in actual CI processors.

A channel selection scheme is incorporated to prevent over-stimulation of a particular electrode or channel [5]. In the simulator, instead of generating an electrical pulse as in a real CI, the low-pass envelope of the signal in each band

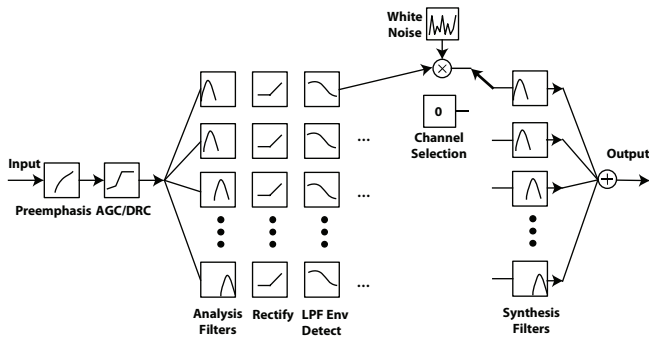


Fig. 1. Signal processing stages for a cochlear implant simulation. The input audio signal is converted to single precision floating point, followed by pre-emphasis and AGC/DRC (Automatic Gain Control / Dynamic Range Compression). The remaining stages include a frequency-specific filtering and envelope extraction followed by a noiseband simulation based on modulating white noise with the low-pass output of each channel. The output of each band is summed to produce the final acoustic output. Adapted from [3], [4]

is computed and modulated with white noise. The resulting output is again band-pass filtered and summed across all channels and delivered acoustically to the listener via insert earphones.

To simulate the mismatch between the analysis frequency bands assigned to each electrode and the characteristic frequencies of the neurons that are stimulated by each electrode, the synthesis band-pass filters can be identical to the input filters or shifted in frequency. This mismatch arises because the electrode array cannot be fully inserted into the cochlea, and as a consequence the neurons sensitive to the lowest frequencies (occurring at the apex of the cochlea) are not generally stimulated, forcing information about these lowest frequencies to be delivered to neurons that are normally responsive (in a healthy cochlea) to higher frequencies. This “basalward frequency shift” can be specified in mm and is implemented using Greenwood’s equation: $F = A(10^{ax} - k)$, with conventionally used constants for humans $A = 165.4$, $a = 0.06$, and $k = 0.88$ [6], [7].

III. METHODS

A. Portable Implementation

The implementation of the CI simulator was based on previously developed C code, intended for laboratory use [3]. Before porting the simulator to the iPhone, it was first implemented on a floating point DSP embedded processing platform. Although an embedded device is optimized for signal processing algorithms due to its specialized instruction set, the demand for powerful portable phones has produced capable, inexpensive devices. The Apple iPod Touch and iPhone both use the ARM processor with clock speeds ranging from 400 MHz up to 1 GHz in the most recent version [8]. Furthermore, these devices are all equipped with a vector floating point processor, allowing for ease of programming while ensuring accuracy of computations.

When considering experiments carried out with freely moving participants in real-time, it is crucial that the speech



Fig. 2. Graphical user interface for a portable cochlear implant simulator. The screen on the left shows the default view when starting the program. The screen on the right allows for specification of the channel and basalward shift.

signal be processed with minimal delay from input to output. In the case of audiovisual perceptual experiments where the speech information is recognized by the subject based on hearing speech and observing the talker’s lip movements, an asynchrony between the heard sound and the lip movement of the speaker may lead to confusions. We utilized the RemoteIO unit of Apple’s Core Audio libraries, an API that allows for simultaneous recording and playback of audio streams. This library allowed our portable CI simulator to run at a latency of 10 ms at sampling rate of 22 KHz on a 2nd Generation iPod Touch. Depending on the capabilities of the specific device in use, there is a tradeoff between the latency (buffer size), sample rate, and the number of channels. By increasing the latency, one may allocate a larger buffer for the audio data, allowing for more parallel execution with vectorized CPU instructions. The signal is separated into its frequency components by a series of Butterworth infinite impulse response (IIR) filters of order $N=4$, implemented in direct form II. An objective C graphical user interface shown in Figure 2 was used to control the system. The interface allows for virtually any number of pre-defined processing settings, most notably the number of channels and amount of basalward shift in mm. The device currently supports up to 12 channels and a basalward shift up to 6.5mm. Two channel section strategies used in commercial available devices, CIS and n-of-m, have been implemented but the selection cannot be alternated via the graphical interface at this time.

B. Behavioral Testing

To assess perceptual learning to the real-time CI simulator, the device was worn by 6 normal-hearing participants over a two week chronic exposure period for 2 hours per day. Participants were asked to actively listen to speech or music in some form during the 2 hour period. The apparatus consisted of a lapel microphone and insert earphones, rated at 35-42

dB noise isolation [9]. Eight un-shifted (i.e. ideally inserted, no basalward shift) noise band channels [10] were used, with the filter bank cutoff frequencies ranging from 100 Hz to 8 kHz. A word and sentence recognition task was carried out in a sound treated room before and after exposure under free field conditions; participants wore their portable simulator during testing. Three different tests to measure behavioral performance were used: phonetically-balanced (PB) words, PRESTO multi-talker sentences, and semantically anomalous sentences [11]. The number of presentations of each stimulus type was 50 (words), 90, and 50 (sentences) respectively, where no stimuli were repeated across sessions. Subjects were asked to repeat out loud the word or sentence they heard while wearing the portable CI simulator. Scoring was done for each participant based on the percent correct of key words, as obtained from a recording of their verbal responses. For each subject, the overall score for a given test was taken as the mean score of all trials.

C. fMRI Paradigm

An fMRI protocol that has been demonstrated to be highly reliable across imaging sessions at the group level was used to evaluate neural activation patterns before and after the chronic exposure to the portable CI simulator. This auditory sentence comprehension paradigm with normal sentences was run on 16 subjects across 5 sessions and showed that 84% of the acquired volume were active/inactive at group level and 59% at the single-subject level [12]. In this study, two runs of both normal and CI simulated sentences were presented to the 6 participants using this blocked paradigm, where functional runs were obtained using a multi-slice echo planar imaging sequence (TR=2.5s, TE=22ms, 34 slices, slice thickness=3.8mm, spacing=0mm, matrix size=64x64, FOV=24cm, flip angle=77). Participants were asked to listen to 6 sentences from the CUNY (City University of New York) sentences test [13], then respond to 3 fill-in-the-blank multiple choice questions regarding the presented sentences. Comparing across sessions (pre and post exposure), changes in brain activation exhibited by the CI sentences that differ from the normal condition would likely be a result of changes in processing strategies due to adaptation to transformed inputs.

IV. RESULTS

A. Word / Sentence Comprehension

After chronic exposure, participants improved in word and sentence recognition, demonstrating that it is possible to adapt in a short period of time to a spectrally degraded auditory signal (Figure 3). Both absolute and post-pre scores are shown with error bars indicating standard error calculated across subjects. Two sided *t*-tests performed across session for each test type revealed that both the PRESTO and anomalous sentences showed statistically significant improvements in performance ($p < 0.05$). Previous studies have shown approximately a 90 percent correct on a similar sentence recognition task to the PRESTO sentences [14]. In our study, previously untrained subjects performed at a level

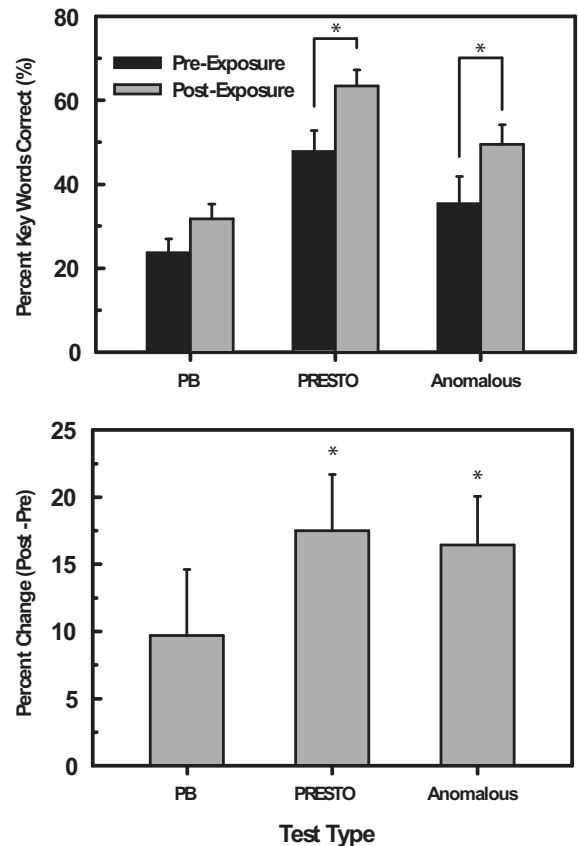


Fig. 3. Behavioral test results before (pre) and after (post) chronic exposure to a portable cochlear implant simulator. Both absolute (top panel) and relative (bottom panel) results are shown. Scores are expressed in percent of key words correct for open set word identification using phonetically-balanced (PB) words, PRESTO multi-talker sentences, and semantically anomalous sentences [11]. The error bars indicate the standard error across the subjects, and a * indicates $p < 0.05$ for a 2-sided *t*-test across session.

not inconsistent with scores obtained by subjects directly trained on the task. The performance for the open-set word identification task is noticeably less than that of the sentence tasks, due to the fact that there is no contextual clues to aid in identification.

B. Pre / Post Exposure fMRI activation

Figure 4 compares pre-/post-training fMRI activations for degraded and normal speech averaged across the participants, contrasted with a rest condition. For normal speech (used here as a control that is not expected to change), there are few activation differences across sessions, almost all of which are localized to primary auditory cortex. For CI simulated speech, an increase in activation was observed near the right insula and inferior frontal gyrus, which may indicate greater processing of prosodic structure after training [15]. In the left hemisphere, a decrease in activity post-training in left prefrontal cortex (BA 9-10) was observed, consistent with a previous study that showed this region to be active in semantic and syllabic tasks using degraded speech stimuli [16].

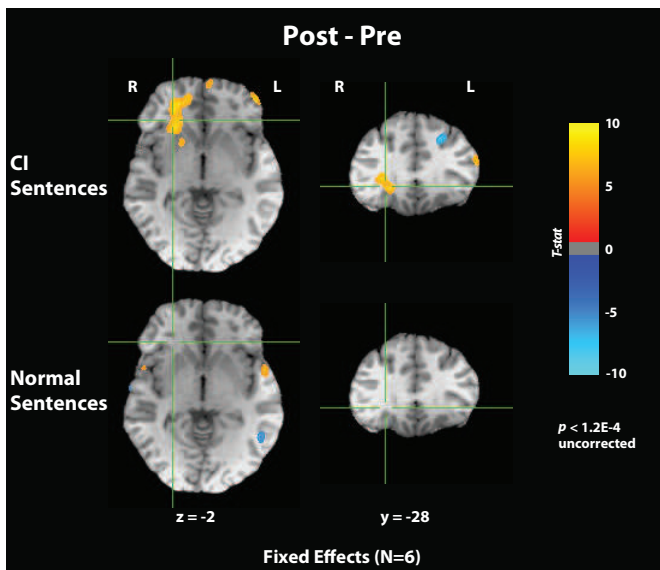


Fig. 4. Mean pre-/post-training fMRI activations for degraded and normal speech when contrasted with rest. Functional runs were acquired using a multi-slice echo planar imaging sequence. Standard processing was performed on fMRI data using AFNI [17], including slice timing alignment, volume registration, normalization, and spatial smoothing (6mm).

V. CONCLUSIONS

A cochlear implant simulation for normal hearing listeners has important research and clinical applications. Our results indicate that a portable electronic device, when used to present a transformed acoustic signal, can induce neural and behavioral changes in speech perception through freely moving active learning in everyday environments. Furthermore, adaptation to spectrally degraded speech is accompanied by changes in working memory and language processing, indicating that knowledge is integrated over the exposure period. Through the understanding of physiological changes revealed from fMRI, we may better understand and model speech networks in the brain. This could also lead to new signal processing and targeted training strategies can be developed to optimize patient rehabilitation and speech recognition.

Other possible applications of portable devices exist for the purpose of testing new signal processing algorithms, or modeling behavioral performance under auditory conditions similar to those associated with a CI. Using a standalone wearable device like an iPod Touch is an easy way for researchers to conduct real-world experiments in natural habitats on vocoded speech and CI performance without having to deal with a smaller subject pool or the potential confounds introduced by individual patient variations in implantation, etiology of hearing loss, or device factors (e.g., number of channels/electrodes, etc.). Furthermore, it is realistic for participants to experience talkers and environmental sounds outside of the lab or highly controlled soundproof room. Given the relatively low cost of these portable multimedia devices (specifically the iPod Touch),

many could be deployed at once, facilitating larger scale experimentation.

A portable implementation of a CI simulator could also have additional clinical applications as a teaching tool. Since the device was developed on a commonly available portable platform, the device is easy to use and requires no specialized hardware. Normal hearing parents and teachers of young CI patients could gain a greater understanding of the benefits and limitations of CIs with respect to acoustic communication with children in their care, particularly in multi-talker or high noise adverse listening environments.

REFERENCES

- [1] V. Peddigari, N. Kehtarnavaz, and P. Loizou, "Real-time LabVIEW implementation of cochlear implant signal processing on PDA platforms," in *Acoustics, Speech and Signal Processing, 2007. ICASSP 2007. IEEE International Conference on*, vol. 2. IEEE, 2007.
- [2] A. Hervais-Adelman, M. H. Davis, I. S. Johnsrude, and R. P. Carlyon, "Perceptual learning of noise vocoded words: Effects of feedback and lexicality," *Journal of Experimental Psychology: Human Perception and Performance*, vol. 34, no. 2, pp. 460–474, 2008.
- [3] A. R. Kaiser and M. A. Svirsky, "Using a personal computer to perform real-time signal processing in cochlear implant research," in *Proceedings of the IXth IEEE-DSP Workshop*, 2000.
- [4] S. Rosen, A. Faulkner, and L. Wilkinson, "Adaptation by normal listeners to upward spectral shifts of speech: Implications for cochlear implants," *The Journal of the Acoustical Society of America*, vol. 106, no. 6, pp. 3629–3636, 1999.
- [5] F. Wilson, B. D. C., Lawson, R. Wolford, D. Eddington, and W. Rabinowitz, "Better speech recognition with cochlear implants," *Nature*, vol. 352, no. 6332, pp. 236–238, 1991.
- [6] D. D. Greenwood, "A cochlear frequency-position function for several species—29 years later," *Journal of the Acoustical Society of America*, vol. 87, no. 6, 1990.
- [7] M. Dorman, P. Loizou, and D. Rainey, "Simulating the effect of cochlear-implant electrode insertion depth on speech understanding," *Journal of the Acoustical Society of America*, vol. 102, pp. 2993–2996, 1997.
- [8] J. Snell, "That iPod touch runs at 533 MHz," Available: http://www.pcworld.com/article/154518/html?tk=rss_news, 2008.
- [9] Etymotic, "Earphone comparison," Available: <http://www.etymotic.com/ephp/epcomp.html>, 2010.
- [10] A. Hervais-Adelman, M. Davis, I. Johnsrude, and R. Carlyon, "Perceptual learning of noise vocoded words: Effects of feedback and lexicality," *Journal of Experimental Psychology: Human Perception and Performance*, vol. 34, no. 2, p. 460, 2008.
- [11] H. Park, R. Feltz, K. Lormore, and D. B. Pisoni, "PRESTO: perceptually robust English sentence test: Open-set—design, philosophy, and preliminary findings," *The Journal of the Acoustical Society of America*, vol. 127, no. 3, pp. 1958–1958, 2010.
- [12] J. Gonzalez-Castillo and T. M. Talavage, "Reproducibility of fmri activations associated with auditory sentence comprehension," *Neuroimage*, vol. 54, pp. 2138–55., 2011.
- [13] A. Boothroyd, T. Hnath-Chisolm, L. Hanin, and L. Kishon-Rabin, "Voice fundamental frequency as an auditory supplement to the speechreading of sentences," *Ear and Hearing*, vol. 9, no. 6, pp. 306–12, 1988.
- [14] R. Shannon, Q. Fu, and J. Galvin, "The number of spectral channels required for speech recognition depends on the difficulty of the listening situation," *Acta Oto-Laryngologica*, vol. 124, pp. 50–54, 2004.
- [15] Y. Xu, J. Gandour, T. Talavage, D. Wong, M. Dzemidzic, Y. Tong, X. Li, and M. Lowe, "Activation of the left planum temporale in pitch processing is shaped by language experience," *Human Brain Mapping*, vol. 27, no. 2, pp. 173–83, 2006.
- [16] D. Sharp, S. Scott, and R. Wise, "Monitoring and the controlled processing of meaning: distinct prefrontal systems," *Cerebral Cortex*, vol. 14, no. 1, p. 1, 2004.
- [17] R. W. Cox, "Afni: software for analysis and visualization of functional magnetic resonance neuroimages," *Computers and Biomedical Research*, vol. 29, no. 3, pp. 162–73, 1996.