Proceedings of the 1st International IEEE EMBS
Conference on Neural Engineering
Capri Island, Italy • March 20-22, 2003

A Novel Acoustic Simulation of Cochlear Implant Hearing: Effects of Temporal Fine Structure

J.T. Rubinstein, C. Turner
Departments of Otolaryngology, Biomedical Engineering, Speech Pathology & Audiology, University of Iowa, IA, USA

Abstract-**Acoustic simulations of cochlear implant signal processing have been widely used to determine the upper bounds on speech perception using these devices. These simulations assume that the only temporal information delivered to the inner ear is the envelope of the signal and that spectral cues, determined by the number of channels present, are the primary determinant of speech perception ability. While highly informative, such simulations fail to explain open-set speech perception obtained by some patients who received single channel implants as young children. In addition, some implant recipients enjoy and appreciate music to an extent not possible if envelope is the sole source of temporal information. In this paper, a novel acoustic simulation algorithm is described which allows varying the degree to which temporal fine structure is delivered to the cochlea. Use of this algorithm demonstrates that while high levels of speech perception in quiet can be provided with only six channels of envelope, inclusion of even small amounts of temporal fine structure dramatically improves perception of speech in noise, timbre and melody.**

Keywords – **Cochlear implant, speech perception, music perception, simulation**

I. INTRODUCTION

Acoustic simulations of cochlear implants were devised to present to the normal hearing ear, a similarly degraded signal as that presented to the central nervous system by a cochlear implant. They were initially introduced by Shannon et al. [1] and by Dorman et al. [2] using two related approaches. The latter approach involved bandpass filtering the input signal into *N* bands and modulating *N* sinusoids, the center frequencies of the filters, by the envelope of each band. The sum of the modulated sinusoids was then presented to a loudspeaker. The former approach is identical, but instead of modulating sinusoids, bands of noise, spectrally corresponding to the filters, are modulated. Speech perception scores of normal hearing listeners using both approaches are similar and the former approach has become more widely adopted. Indeed cochlear implant acoustic simulations are now commonly referred to as "noise-band simulations."

As originally proposed, noise-band simulations were intended to simulate speech perception using the continuous interleaved sampling (CIS) strategy [3] because the processing is so similar. It was noted that speech perception scores of normal hearing listeners using the simulations were comparable to the "best" speech perception scores of cochlear implant patients and therefore represent an upper bound on possible performance using CIS or other strategies involving envelope extraction such as the spectral peak or advanced combination encoder strategies (SPEAK, ACE). The simulations may also well represent strategies which do not extract envelope, such as analog, since in some subjects it provides comparable levels of speech perception in quiet. This is likely due to ensembles of auditory neurons being relatively incapable of following the rapid fluctuations of temporal fine structure in analog electrical stimulation due to their refractory characteristics [4,5]. These neurons are more likely to faithfully represent the more slowly varying envelope and as such the electro-neural interface may be performing its own envelope extraction through lowpass filtering by the membrane and rectification by the nonlinear excitation process. Thus not only does the noise-band simulation algorithm resemble the signal processing strategies of the implant hardware, it may well represent the biophysical interface. This may help explain the success obtained predicting performance of "short electrode" cochlear implants in patients with residual low frequency hearing using an appropriate modification of the noise-band algorithm [6].

While noise-band simulations have been very effective at predicting the upper bound of cochlear implant speech perception scores in quiet, and in some cases in noise [7], they fail to account for at least three perceptual phenomena observed in implant patients.

1) *Single-channel devices:* While in general, speech perception with single-channel implants is much poorer than multi-channel devices and significant levels of monosyllabic word recognition are rare [8], there are some notable exceptions. The first author has personal experience with three profoundly deaf adolescents who received single-channel implants as young children and now have monosyllabic word scores higher than 20%. Single-channel noise-band simulations accurately predict poor speech perception in most adult-implanted single-channel patients but not these or other reported subjects implanted in early childhood in whom the central nervous system must have learned to extract some temporal fine structure from the signal.

2) *Speech perception in noise:* Noise-band simulations predict that while high levels of speech perception in quiet can be achieved with only six channels, more channels are needed for speech recognition in noise [7]. Experiments

---

Supported by NIH program project DC00242 and Neural Prosthesis Program contract DC92107.

with implant patients bear this out [9] but the simulations fail to account for the occasional subject who performs exceptionally well in noise [10]. Again, such subjects must be able to extract some fine structure from the electrically encoded signal in order to achieve this.

3) *Melody, pitch perception and sound quality:* It is clear that cochlear implants produce very poor melody and pitch perception when compared with normal hearing [11]. At first glance, noise-band simulations of music are consistent with these data and in fact produce sound quality and music perception that is so poor, it is difficult to imagine anyone voluntarily listening to it. Studies of implant patients demonstrate a range of musical enjoyment [11] including some who never listen to music, but it is clear that a significant percentage enjoy and have levels of musical ability well beyond those consistent with noise-band simulations. Studies of "auditory chimeras" show that greater levels of musical ability are possible with increasing numbers of channels, but it requires between 64 and 128 channels to produce high levels of music perception and even then, sound quality is poor [12]. Since current implant devices effectively present only six to eight channels of information even with as many as twenty two electrodes [9], some level of temporal fine structure must be coded in the auditory nerve of the implant patients with higher levels of complex pitch and melody perception.

A signal processing strategy has recently been described which attempts to improve representation of temporal fine structure in the electrically stimulated ear [4]. Studies with implant subjects reveal substantial increases in dynamic range with this strategy as expected from the underlying theory and physiological data [13,14]. Development of a signal processor implementing this strategy in such a way as to permit speech and music testing is underway, but it is informative to understand what the expected perceptual effects of improved temporal resolution would be in advance of such testing. A novel acoustic simulation strategy, a generalization of the noise-band approach, has been developed to assess the impact of improved temporal resolution on signal perception with a cochlear implant. In addition, it permits better understanding of the three perceptual phenomena described earlier, which demonstrate the shortcomings of the otherwise theoretically and empirically satisfying noise-band algorithm.

II. METHODOLOGY

On one extreme, noise-band simulations assume that no temporal fine structure is presented within each band. On the opposite extreme, complete representation of temporal fine structure with a large number of channels should reproduce the original signal. Both of these extremes can be approximated with a noise-band simulation where the "noise" signal variably incorporates the fine structure of the original signal. If we define the input signal as x(t), then this "fine structure noise" is generated by taking a bandlimited noise signal n(t) and extracting its Hilbert envelope *abs(H(n(t)))* where *H* represents the Hilbert transform [11]. This envelope then modulates the fine structure of *x(t)* which has been fractionally randomized:

$$X_{FS}(t) = abs(H(n(t))) * \cos(angle(H(f(t))) + 2\pi l\tilde{r}) \quad (1)$$

where $\tilde{r}$ is a uniformly distributed random number between -0.5 and +0.5 and l is the fractional randomization which can be varied between 0 and 1. The signal generated by (1) includes high frequency components due to the randomization that are removed by bandpass filtering. The bandpass filtered signal has a random Hilbert envelope and a fine structure which when l=0 is identical to the fine structure of *f(t)*. When l=1, this signal is identical to bandlimited noise.

Equation (1) can then be used in a noise band simulator instead of the bandlimited noise that has been previously described [1,2,3]. With a large number of channels when l=0, the original signal is reproduced. With any number of channels when l=1, the output is identical to a noise-band simulation with an equal number of channels. In practice, quadrature filters are used rather than Hilbert transforms to improve computational efficiency [12]. The program is written in MATLAB and takes WAV audio files as input and generates WAV files as output. Fig. 1 demonstrates a gated 131 Hz sinusoid used as an input signal to a single-band simulation and below it are three outputs representing 0%, 30% and 60% randomization of the fine structure.

III. RESULTS

No quantitative psychoacoustic evaluation of this algorithm in normal hearing listeners has yet been performed. A number of experiments are underway or planned for the near future. Pitch perception for pure and complex tones are being examined using single channel

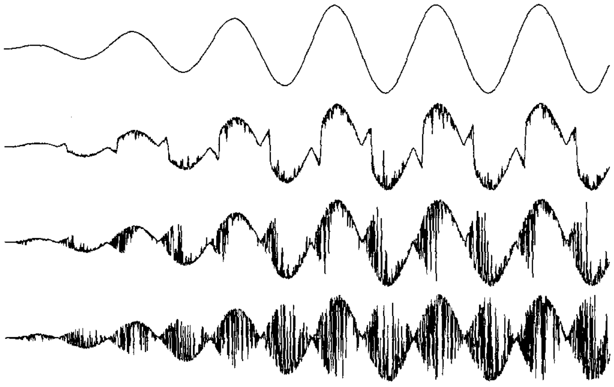

Fig.1: Example of single-band simulation. Input of a 131 Hz gated sinusoid in the top row produces outputs for 0%, 30% and 60% randomization in the subsequent three rows. The "glitches" at the zero-crossings of the 0% simulation are a signal processing artifact.

simulations, speech perception studies are underway in noise using six channel simulations, and music perception studies using six to twelve channels are planned. In all cases, the fine structure may be randomized anywhere between zero and one hundred percent. Despite the lack of quantitative data a number of qualitative observations are clear.

1) Speech presented through a six-band algorithm with 100% fine structure randomization sounds identical to that presented through a standard six-band noise-band simulation.

2) With only one spectral band, speech has normal intelligibility even if the fine structure is randomized by 70-80%. In addition, characteristics of the speaker's voice, such as nasality, become prominent even at these levels of randomization. These features are entirely absent from standard noise-band simulations and perhaps provide an informative simulation of the successful single-channel implant users mentioned previously.

3) With no randomization of the fine structure, speech sounds normal, though noisy, with only one band. Melody and pitch perception are also grossly normal with a single band. To reproduce high-fidelity music however requires more than thirty bands to eliminate the effects of the random envelope introduced with the "fine structure noise."

4) With a single channel, pitch perception of low frequency tones (~150 Hz) is present even with 98% randomization of the fine structure. For high frequency tones (~7 kHz), this percept vanishes above approximately 70% randomization.

5) With 6-12 bands, even small amounts of fine structure representation significantly improve intelligibility of speech in noise. The more bands that are included, the less fine structure is needed (and vice versa) to maintain intelligibility in noise.

6) Even small amounts of fine structure coding, ie. large amounts of randomization, produce levels of music perception comparable to noise-band simulations with more than 60 bands.

IV. DISCUSSION

The limited qualitative results noted so far have significant implications for cochlear implant design but must be interpreted with caution. One of the bases for positing a similarity between noise-band simulations presented to a normal ear and cochlear implant hearing is the lack of temporal information, other than envelope, presented within each band by the simulation. Because of this absence, it is not possible for the normal ear to extract spectral cues from temporal ones using auditory filters not present in implant recipients. When fine structure cues are introduced using (1), extraction of spectral cues from temporal information for simultaneous sounds becomes possible if the width of the bandpass filters exceeds the critical bandwidth. It should also be noted that cochlear implant patients may possess widely divergent pathologies both at the periphery and in the central auditory pathways that could produce differential abnormalities of either fine structure or envelope processing. Mindful of these caveats, it is apparent that even small amounts of fine structure coding can dramatically improve pitch and melody perception as well as speech perception in noise. The improvements in music perception and sound quality are substantially greater than those produced by using an increased number of spectral bands until this number is well over fifty. Since current implant devices, even those with twenty two electrodes, deliver only approximately six to eight channels of information [9], it seems unlikely that electrode modifications in the near future will deliver the number of channels necessary to produce high levels of music perception using current signal processing. Efforts to improve fine structure coding, however, are feasible [4,13,14] and more likely to soon result in such improvements.

It is tempting to speculate that the differences seen between perception of low and high frequency tones are due to differences in cochlear mechanisms at these different frequencies. At low frequencies, very small amounts of fine structure information produce a clear pitch percept consistent with temporal coding via phase-locking at this frequency [15,16]. At frequencies above 5 kHz, much larger amounts of fine-structure information are necessary perhaps because the auditory nerve is incapable of phase-locking to the fine structure at these frequencies requiring instead that spectral analysis be performed. This intriguing speculation is currently undergoing more detailed study.

A number of other experiments are planned to assess the similarity of acoustic simulations to hearing with a cochlear implant. As part of a separate study, a patient has received a cochlear implant in a deaf ear with a nearly normal hearing contralateral ear. This subject provides the rare opportunity to present acoustic simulations to the hearing ear and compare the resulting percepts to electrical stimulation of the deaf ear. By adjusting the number of spectral bands and the degree of fine structure representation so as to equalize speech and melody perception on the two sides, the subject's subjective comparison may allow determination of how cochlear implants actually sound to a normal hearing listener.

V. CONCLUSION

A novel acoustic simulation of cochlear implant hearing has been developed. It allows quantitative assessment of the interaction between the number of spectral bands and the degree of temporal fine structure represented within each band in determining speech and music perception. The results suggest that signal processing strategies which improve coding of temporal fine structure in the auditory nerve should yield substantial improvements in music perception as well as speech perception in noise. Although both improved fine structure coding and increased numbers of spectral bands enhance the fidelity and intelligibility of both speech and music, music perception is improved more by better fine structure coding than by increases in the number of spectral bands.

ACKNOWLEDGMENT

JTR would like to acknowledge multiple fruitful discussions on this topic with Drs. Kate Gfeller and Richard Tyler.

REFERENCES

[1] R.V. Shannon, F.G. Zeng, V. Kamath, J. Wygonski, M. Ekelid. "Speech recognition with primarily temporal cues," *Science*, **270**:303-304, 1995.

[2] M.F. Dorman, P.C. Loizou, D. Rainey. "Speech understanding as a function of the number of channels of stimulation for processors using sine-wave and noise-band outputs," *J Acoust Soc Am* **102**:2403-2411, 1997

[3] M.F. Dorman and P.C. Loizou PC. "The identification of consonants and vowels by cochlear implant patients using a 6-channel continuous interleaved sampling processor and by normal-hearing subjects using simulations of processors with two to nine channels," *Ear Hear* **19**:162-166, 1998

[4] J.T. Rubinstein, B.S. Wilson, C.C. Finley CC and P.J. Abbas. "Pseudospontaneous activity: stochastic independence with electrical stimulation of the auditory nerve." *Hear. Res.,* 127, 108-118, 1999.

[5] C.A. Miller, P.J. Abbas, B.K. Robinson. "Response properties of the refractory auditory nerve fiber." *JARO.* **2**(3):216-32, 2001.

[6] C. Turner and B.J. Gantz. "Combining acoustic and electric hearing for patients with high frequency hearing loss," *2001 Conf Implantable Auditory Prostheses,* Asilomar, CA, p.33.

[7] Q. Fu, R.V. Shannon, X. Wang. "Effects of noise and spectral resolution on vowel and consonant recognition: acoustic and electric hearing," *J Acoust Soc Am* **104**:1-11, 1998.

[8] J.T. Rubinstein, W.S. Parkinson, M.W. Lowder, B.J. Gantz, R.S. Tyler. "Single-channel to multichannel conversions in adult cochlear implant subjects," *Am J Otol,* 19 (4): 461-466,1998.

[9] K. Fishman, R.V. Shannon, W.H. Slattery. "Speech recognition as a function of the number of electrodes used in the SPEAK cochlear implant speech processor," *J Speech Hear Res* **40**:1201-1215, 1997.

[10] E. Overstreet. "Auto-conditioning with high resolution neurostimulation," Advanced Bionics Corp., unpublished.

[11] K. Gfeller, C. Turner, et al. "Recognition of familiar melodies by adult cochlear implant recipients and normal-hearing adults." *Coch. Impl. Intl.* **3**(1): 29-53, 2002.

[12] Z.M. Smith, B. Delgutte, and A.J.Oxenham. "Chimaeric sounds reveal dichotomies in auditory perception," *Nature* 416, 87-90, 2002.

[13] R.S. Hong, J.T. Rubinstein, D. Wehner and D. Horn. "Dynamic range enhancement for cochlear implants." *Otol. Neurotol.,* in press.

[14] L. Litvak, B. Delgutte, and D. Eddington. "Auditory nerve fiber responses to electric stimulation: Modulated and unmodulated pulse trains," *J. Acoust. Soc. Am.* 110, 368-379, 2001.

[15] B.C.J. Moore. "Frequency difference limens for short-duration tones," *J Acoust Soc Am* 54(3):610-9, 1973.

[16] D.H. Johnson. "The relationship between spike rate and synchrony in responses of auditory-nerve fibers to single tones," *J Acoust Soc Am* 68(4):1115-22, 1980.