SINGLE SIDEBAND ENCODER FOR MUSIC CODING IN COCHLEAR IMPLANTS

Kaibao Nie[14], Les Atlas[3] and Jay Rubinstein[124]

Departments of [1]Otolaryngology, [2]Bioengineering, and [3]Electrical Engineering
[4]Virginia Merrill Bloedel Hearing Research Center, University of Washington, Seattle, USA

ABSTRACT

The restoration of melody perception is a key remaining challenge in cochlear implants. We propose a new sound coding strategy that converts an audio signal into time-varying electrically stimulating pulse trains. A sound is first split into several frequency subbands and each subband signal is coherently downward shifted to a low-frequency base band, similar to demodulation used in single sideband (SSB) radios. These resulting coherent envelope signals have Hermitian symmetric frequency spectrums and are thus real-valued. A peak detector in each subband further converts the coherent envelopes into rate-varying and interleaved pulse trains. Acoustic simulations of cochlear implants with normal hearing listeners showed significant improvement in melody recognition over the most common stimulation approach used in cochlear implants.

Index Terms— hearing, auditory system, demodulation, ears, cochlear implants

1. BACKGROUND

More than 100,000 patients worldwide with profound hearing loss have received cochlear implants as a clinical treatment to regain partial hearing. In current cochlear implants, most speech coding strategies extract and deliver a small number of temporal envelope cues via pulsatile electrical stimulation. For instance, in the widely used continuous interleaved sampling (CIS) [1] coding scheme, sounds are split into a few subbands and the slowly-varying envelopes are extracted with a half- or full-wave rectifier followed by a low-pass filter in each subband. Alternatively, envelopes can be calculated from the magnitude of the Fast Fourier Transform (FFT) or the Hilbert transform.

For this paper, we instead assume a more general sum-of-products signal model

$$x(t) = \sum_{k=1}^{N} x_k(t) = \sum_{k=1}^{N} a_k(t) \cdot c_k(t) \qquad (1)$$

where k is a subband index, $x_k(t)$ is the output for each of N subbands, and some detection rule, incoherent or coherent, is used to determine the product decomposition of each subband output $x_k(t) = a_k(t) \cdot c_k(t)$ into slowly varying amplitude and higher frequency carrier signals, respectively.

The detection rule used previously in cochlear implants has been to decompose each subband signal $x_k(t)$ into a Hilbert envelope and associated carrier. This approach begins with the determination of the analytic signal

$$\tilde{x}_k(t) = x_k(t) + j\,\mathrm{H}\{x_k(t)\} \qquad (2)$$

where $\mathrm{H}\{x_k(t)\}$ is the Hilbert transform of $x_k(t)$. The amplitude part is then the non-negative and real magnitude of the analytic signal

$$a_k(t) = |\tilde{x}_k(t)| \qquad (3)$$

which is commonly called the "Hilbert envelope." The carrier part is the remaining unimodular phase of the analytic signal

$$c_k(t) = \cos\left\{\tan^{-1}\frac{\operatorname{Im}\tilde{x}_k(t)}{\operatorname{Re}\tilde{x}_k(t)}\right\} = \cos\varphi_k(t) \qquad (4)$$

In cochlear implants, only the non-negative and real envelope $a_k(t)$ is delivered to the selected stimulating electrode at a fixed stimulation rate. It is evident that this conventional envelope extraction process eliminates the temporal fine structure cues ($\cos\varphi_k(t)$) in each subband, yielding a coarse spectral and temporal representation of speech and music sounds. Psychoacoustic experiments have shown that, with limited number of envelopes, most patients are still able to understand speech to a certain high level [2] and they can even converse over the phone. However, the lack of temporal fine structure has led to poor speech recognition in noisy environment, near-chance level of melody recognition, poor Mandarin tone recognition and production, and inability to use ITD (Inter-aural Timing Difference) cues to localize sounds among the majority of cochlear implant users [3-6].

Smith *et al* [4] have demonstrated that the temporal fine structure is crucial to sound localization and music perception. Inspired by the study, a number of signal processing strategies have been proposed to encode the temporal fine structure cue in cochlear implant. Nie *et al* [3] proposed a frequency and amplitude-modulation encoding (FAME) strategy to incorporate continuous frequency modulation (FM) cues, which are extracted from the temporal fine structure, into speech processing for cochlear implants. Throckmorton *et al* [5] further investigated the possibility of delivering discrete FM to cochlear implants. Sit *et al* [6] recently proposed the AIS

(Asynchronous Interleaved Sampling) to encode the temporal fine structure with a race-to-spike algorithm. In these strategies, the stimulation rate is co-varied with either FM or some other features and they are still under investigation.

The encoding of temporal fine structure in cochlear implants is ultimately restricted by the ability of temporal pitch perception in electrical stimulation. Zeng [7] and other studies have shown that cochlear implant patients can only perceive stimulate rate variation up to 1000 Hz. However, the frequency content of the temporal fine structure ($\cos\varphi_k(t)$) in speech and music can be up to 10,000 Hz at higher spectral subbands and it is not a bandlimited signal. To overcome this limitation, we proposed to use a single sideband demodulation approach to coherently shift a subband signal to its base band, generating a low-frequency, real coherent envelope signal. Theoretically, it carries both temporal envelope and fine structure cues yet in a slowly-varying manner. Single sideband demodulation makes it feasible to deliver perceivable temporal cues to cochlear implants.

In Section 2, the proposed coherent demodulation with a fixed carrier approach will be described. Section 3 presents the design of the single sideband encoder for cochlear implants. Acoustic simulations and listening experiment results will be given in Sections 4 and 5. Section 6 briefly summarizes conclusions and suggested further investigations.

2. FIXED-CARRIER COHERENT DEMODULATION

For the new product model (Equation 1), Schimmel and Atlas [8] have proposed a time-varying carrier estimate chosen as a center of gravity. The estimation of a timing-varying carrier is computationally expensive and complex for real-time cochlear implant applications. Also, the resulting coherent envelopes are complex quantities, which are not suitable for generating real electrical stimulation signals.

In this study, a fixed carrier was utilized to coherently demodulate each sub-band signal.

$$x(t) = \sum_{k=1}^{N} x_k(t) = \sum_{k=1}^{N} a_k(t)\cdot c_k(t) = \sum_{k=1}^{N} a_k(t)\cdot \cos\omega_k t \qquad (5)$$

where ω_k is a fixed carrier frequency at the lower edge of each subband. The envelope $a_k(t)$ is now real (positive- and negative-going) signal, yet intentionally is not a positive-only Hilbert envelope. This fixed-carrier demodulation is similar to that used in early single sideband receivers.

Figure 1 explains how the fixed carrier demodulation is performed. We assumed that the fixed carrier resides on the lower edge of the subband. Each subband was considered to be an upper sideband signal generated from an envelope signal by single sideband modulation. To perform coherent demodulation, an analytical signal was formed for each subband by the Hilbert transform. The analytic signal had one-sided spectrum and it was then multiplied with a complex carrier at the lower edge of that subband, producing a spectrum replica at the base band. The demodulated complex signal was conjugated and summed to construct a real-valued signal.

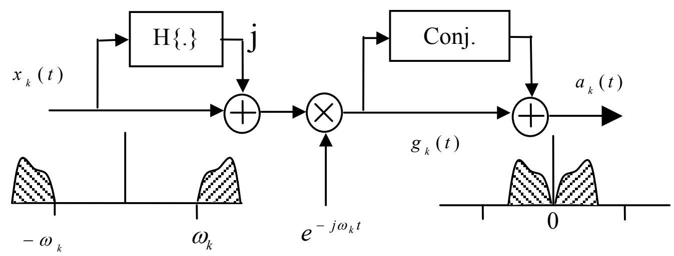

Fig. 1 Illustration of the fix-carrier demodulation in the single sideband encoder. Conj: conjugate; H{.}: Hilbert transform.

For each sub-band,

$$g_k(t)=\{x_k(t)+jH[x_k(t)]\}\cdot e^{-j\omega_k t} \qquad (6)$$

$$a_k(t) = g_k(t) + g_k^*(t) \qquad (7)$$

where * signifies complex conjugate.

Alternatively, the demodulation can be performed with a product detector, which mixes a subband signal with a carrier and then low-pass the mixture.

The coherent envelope signal has a maximum frequency equivalent to its bandwidth. Normally, when a sufficient number of subbands is used, the maximum signal bandwidth can be lower than 1000 Hz. Such low bandwidth signals would be within the perceivable range of rate pitch elicited by electrical stimulation.

3. SINGLE SIDEBAND ENCODER (SSE)

To reduce channel interactions, analog signals should be transformed into interleaved pulse trains in cochlear implants, meaning different stimulating electrodes are sequentially activated within one stimulation cycle.

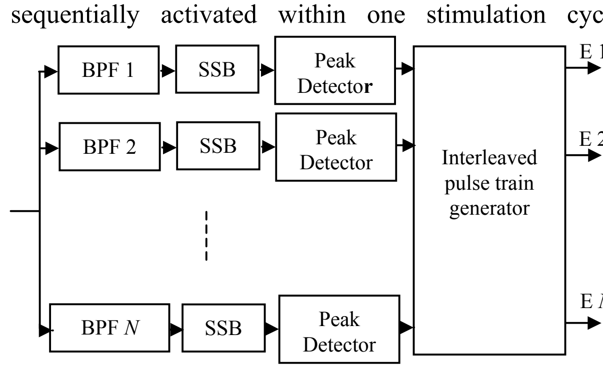

Fig. 2 Block diagram of the single-sideband encoder. BPF: Band-pass filter; SSB: single-sideband demodulator.

A sound signal is first filtered into *N* bands with equal bandwidth on a logarithmic scale. For instance, the cutoff frequencies are 300, 462, 687, 996, 1423, 2013, 2827, 3950 and 5500 Hz when the number of bands is 8. Each band-passed signal is coherently demodulated by single sideband demodulation as discussed above. Fig. 3 shows the waveforms (A) of sub-band signals of a short segment of music sound and its coherent envelopes (B) demodulated with fixed carriers. The coherent envelopes convey both temporal envelope and temporal fine structure cues embedded in each sub-band. It is also possible that delivering these analog signals directly to current cochlear implants like the compressed analog (CA) strategy or SAS (simultaneous analog stimulation) strategy in analog stimulation mode. One of the advantages of using coherent envelopes is that the demodulated analog signals would significantly reduce channel interactions due to the nature of low rates.

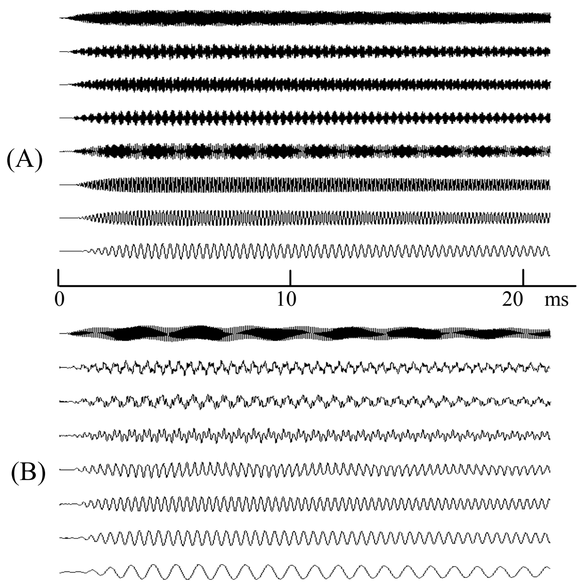

Fig. 3 Waveforms of original subband signals (A) and single-sideband demodulated real envelopes (B) in response to a 22.7 ms-long melody signal. The bottom trace on each panel corresponds to the lowest frequency subband.

For pulsatile stimulation, each analog waveform should be converted into a pulse train. To perform this conversion, pulses are generated in synchrony with the positive peaks in the analog waveforms. The inter-pulse interval then carries zero-crossing cues or phase information. The pulse heights are equal to the amplitudes at the peaks. It has been found that auditory neurons phase lock to temporal peaks only up to 4-5 kHz [9]. Here we use the similar mechanism in generating pulses.

Finally, these pulses might overlap each other in time. An interleaved pulse train generator is used to detect the overlapping pulses and force them to be interleaved. Within one stimulating cycle, all bands are sequentially scanned to find where a peak is present. If a peak is found, a flag will be raised, indicating that the corresponding electrode should be activated subsequently. A biphasic pulse will be generated to stimulate that electrode and the flag is cleared afterwards. This pulse selection procedure ensures only one pulse at a particular period.

The SSE strategy has been implemented using a research interface provided by the Cochlear Corporation. This system provides the needed flexibility to generate arbitrary interleaved electrical pulse stimulation patterns, as is required for patients with an implanted cochlear stimulator.

Figure 4 shows the pulse train pattern created with the CIS and SSE strategy respectively. All parameters are identical except that the stimulation rate of the CIS strategy was set at the typically constant 800 pulses per channel. The pulse pattern of the SSE strategy intentionally contains detailed temporal cues via time-variable yet independent stimulation rates on each electrode.

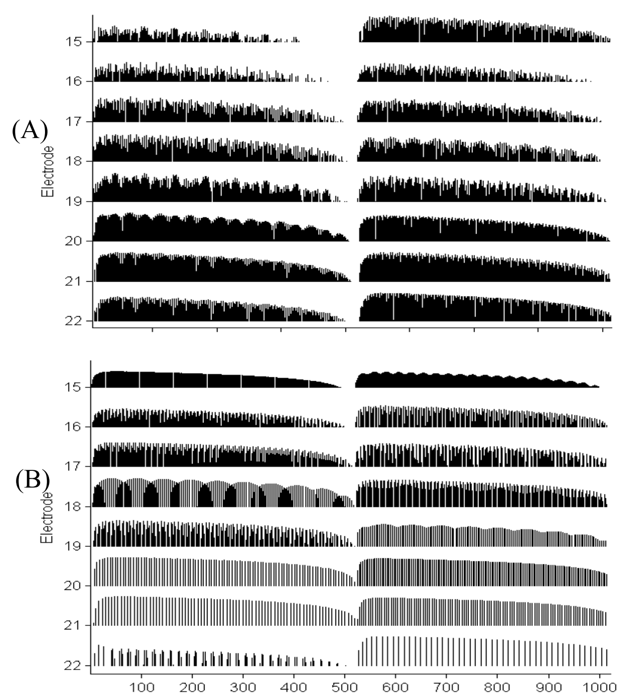

Fig. 4 Pulse train patterns produced by the CIS (A) and the SSE (B) strategy processing two music notes. This figure was generated using the Nucleus MATLAB™ toolbox [10]. Electrode 22 corresponds to a low frequency subband.

4. ACOUSTIC SIMULATIONS

As a preliminary test of the relative efficacy of the conventional CIS and proposed SSE strategies, acoustic sounds were reconstructed from the above pulse trains. Pulse trains were convolved with the impulse response of a 2^{nd}-order Butterworth low-pass filter at 300 Hz (see Fig

5), mimicking the deteriorating temporal pitch discrimination ability in electrically-induced hearing. Each filtered signal was modulated by a carrier identical to the demodulation carrier. All signals were summed together to re-synthesize an acoustic sound.

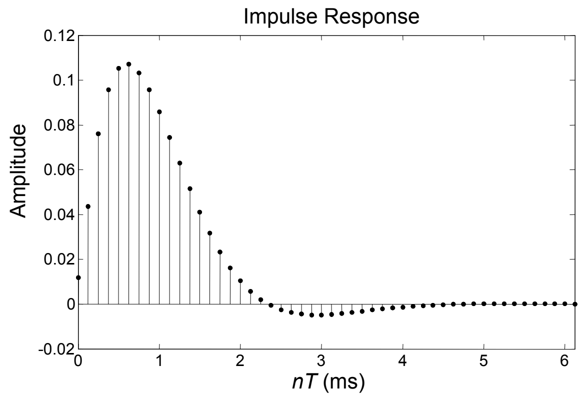

Fig. 5 Impulse response of the Butterworth low-pass filter

Melody recognition was performed to assess whether the SSE coding strategy could improve music recognition performance. Subjects were asked to identify a closed set of twelve common melodies, e.g. “Happy Birthday,” “Frere Jacques,” “Jingle Bells,” etc. Rhythmic cues were removed and all 12 melodies were isochronous. During listening experiments, the subject was allowed to practice twice prior to test. Each melody was presented twice and no feedback was given. The stimuli and user interface is part of a test battery for assessing cochlear implant users’ melody recognition performance in our center [11].

5. RESULTS

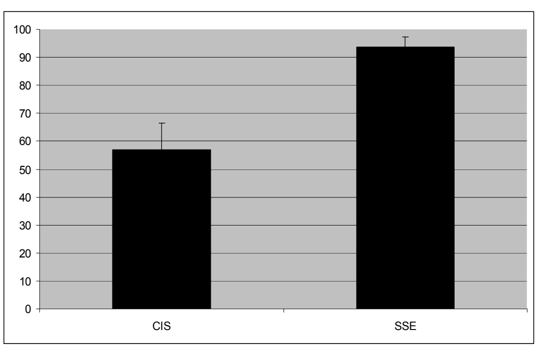

Fig. 6 Melody recognition with acoustic simulations of the CIS and SSE strategies. Error bars indicate standard errors.

The melody recognition scores from 4 normal hearing subjects are presented in Fig. 6. The mean score of the SSE is 92%, whereas the mean for CIS is only 57%. A paired-*t* test suggested that the recognition performance of the SSE strategy is significantly different from the CIS strategy ($p < .001$) [12].

6. DISCUSSION AND CONCLUSIONS

The SSE strategy presents a faithful representation of a high-frequency signal at the lower rates required for electrical stimulation. (Higher rates saturate the patient’s perception[7].) The demodulation and pulse train conversion process can be easily implemented in real-time. Most importantly, the coherent envelopes possibly provide usable temporal cues to implant users. As lower-frequency channels normally have sparse pulse trains, implementing interleaved pulse train stimulation is also feasible. The analog version of the SSE strategy provides low-rate analog stimulation comparable to other cochlear implant strategies, with less simultaneous channel interaction.

The fixed-carrier demodulation also provides a potentially useful tool for extracting slowly-varying features from speech and music. Further investigations are needed on cochlear implant patients. Also, these results suggest a new model on how the normal auditory system encodes dynamic temporal and spectral cues.

ACKNOWLEDGEMENT

We thank Ward Drennan and Bishnu Atal for their helpful comments. This work was supported by the AFOSR Grant FA95500610191, NIH R01 DC007525 and P30 DC004661.

REFERENCES

[1] B.S. Wilson, C.C. Finley, D.T. Lawson, R.D. Wolford, D.K. Eddington, W.M. Rabinowitz. “Better speech recognition with cochlear implants,” *Nature*, 352: 236-238, 1991.

[2] R.V. Shannon, F.G. Zeng, V. Kamath, J. Wygonski, and M. Ekelid. “Speech recognition with primarily temporal cues,” *Science*, 270: 303–304, 1995.

[3] K. Nie, G.S. Stickney, and F.G. Zeng. “Encoding frequency modulation to improve cochlear implant performance in noise,” *IEEE Transactions on Biomedical Engineering*, 52: 64–73, 2004.

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

[5] C.S. Throckmorton, M.S. Kucukoglu, J.J. Remus, L.M. Collins. “Acoustic model investigation of a multiple carrier frequency algorithm for encoding fine frequency structure: implications for cochlear implants,” *Hearing Research,* 218(1-2): 30-42. 2006.

[6] J. Sit, A. M. Simonson, A. J. Oxenham, M. A. Faltys, and R. Sarpeshkar, “A low-power asynchronous interleaved sampling algorithm for cochlear implants that encodes envelope and phase information,” *IEEE Trans. Biomedical Engineering*, 54: 138-149, 2007.

[7] F.G Zeng, “Temporal pitch in electric hearing,” *Hearing Research*, 174: 101-106, 2002.

[8] S. M. Schimmel and L. E. Atlas, “Coherent envelope detection for modulation filtering of speech,” *Proc. IEEE ICASSP*, 221–224, 2005.

[9] S.B. Dynes, B. Delgutte. “Phase-locking of auditory-nerve discharges to sinusoidal electric stimulation of the cochlea,” *Hearing Research,* 58(1): 79-90, 1992.

[10] B. Swanson and H. Mauch. *Nucleus MATLAB Toolbox 4.20*. Cochlear Corporation, 2006.

[11] G.L. Nimmons, R.S. Kang, W.R. Drennan, J. Longnion, C. Ruffin, T. Worman, B. Yueh, and J.T. Rubinstein. “Clinical assessment of music perception in cochlear implant listeners,” *Otol Neurotol*, 2007.(in press)

[12] K Nie, B. King, L.E. Atlas, B.S. Atal, W.R. Drennan and J.T. Rubinstein. “Speech Recognition with coherent envelopes,” *Conference on Implantable Auditory Prostheses*, Lake Tahoe, CA, July 15-20, 2007.