Enhancing speech envelope by integrating hair-cell adaptation into cochlear implant processing

Mahan Azadpour*, 1, Robert L. Smith

Institute for Sensory Research, Department of Biomedical and Chemical Engineering, Syracuse University, 621 Skytop Road, Syracuse, NY 13244, United States

Abstract

Cochlear implants (CIs) bypass some of the mechanisms that underlie normal neural behavior as occurs in acoustic hearing. One such neural mechanism is short-term adaptation, which has been proposed to have a significant role in speech perception. Acoustically-evoked neural adaptation has been mainly attributed to the depletion of neurotransmitter in the hair-cell to auditory-nerve synapse and is therefore not fully present in CI stimulation. This study evaluated a signal processing method that integrated a physiological model of hair-cell adaptation into CI speech processing. The linear high-pass adaptation process expanded the range of rapid variations of the electrical signal generated by the clinical processing strategy. Speech perception performance with the adaptation-based processing was compared to that of the clinical strategy in seven CI users. While there was large variability across subjects, the new processing improved sentence recognition and consonant identification scores in quiet in all the tested subjects with an average improvement of 8% and 6% respectively. Consonant recognition scores in babble noise were improved at the higher signal-to-noise ratios tested (10 and 6 dB) only. Information transfer analysis of consonant features showed significant improvements for manner and place of articulation features, but not for voicing. Enhancement of within-channel envelope cues was confirmed by consonant recognition results obtained with single-channel strategies that presented the overall amplitude envelope of the signal on a single active electrode. Adaptation-inspired envelope enhancement techniques can potentially improve perception of important speech features by CI users.

Keywords:
Cochlear implant
Envelope enhancement
Envelope expansion
Adaptation

1. Introduction

Cochlear implants (CIs) have provided profoundly deaf individuals with substantial hearing for speech communication, but further improvement in CI performance remains an elusive goal. Speech understanding of CI users is generally poorer than that of normally hearing people and varies considerably among the users of the same device.

By directly stimulating auditory-nerve fibers, CIs bypass the inner ear and some of the mechanisms that underlie normal patterns of neural activity as occur during sound exposure in acoustic hearing. Therefore normal neural behavior does not fully occur in CI stimulation, which may result in inefficient coding of some stimulus features. Restoring missing neural mechanisms in CI stimulation may enhance transmission of stimulus information in the auditory system and provide device users with more usable cues for hearing and communication.

One of the potentially important characteristics of normal auditory neural response that may not be fully present in electrical stimulation is adaptation. Adaptation refers to the reduction in firing rate of a neuron in response to stimuli of constant intensity (e.g. Smith, 1977; Smith and Zwislocki, 1975, Smith et al., 1983). Auditory-nerve neuron’s firing rate is maximum at the onset of sounds and gradually decays to a steady-state level during sound presentation. After the sound stops, firing rate drops below the spontaneous rate and sensitivity to new sounds is reduced. Sensitivity to new stimulation increases gradually as the system recovers from adaptation (Smith, 1979). Adaptation can be described by a mechanism that emphasizes the increments and decrements in neural firing rates produced by rapid variations in sound intensity. Intensity variations provide highly-informative cues for perception
of dynamic speech features such as consonants (Chen and Loizou, 2012; Stilp and Kluender, 2010) and for segmenting continuous speech signals (Li and Loizou, 2008). In this regard, it has been hypothesized that adaptation plays a significant role in speech perception by highlighting dynamic time-varying speech cues (Carney and Geisler, 1986; Delgutte, 1980; Delgutte and Kiang, 1984).

The enhancing properties of acoustic adaptation, however, may not be fully replicated in CI stimulation. Adaptation in acoustically evoked auditory-nerve responses has been mainly attributed to the depletion of neurotransmitter in the hair-cell to auditory-nerve synapse (Goutman and Glowatzki, 2007; Palmer and Russell, 1986; Smith, 1998), which is bypassed in electrical stimulation. While some evidence of adaptation-like decay in firing rates has been observed in the electrically stimulated auditory nerve (Litvak et al., 2001; Miller et al., 2008; Zhang et al., 2007), the characteristics of this adaptation are somewhat different than those applicable to hair-cell adaptation. Electrical adaptation is thought to arise from the membrane potential dynamics of the neurons that respond to a sequence of electrical pulses (Negm and Bruce, 2014), and is highly dependent on the site and parameters of electrical stimulation (Miller et al., 2008; Wu et al., 2010; Zhang et al., 2007). Firing-rate adaptation appears to be relatively weak at the relatively low levels and rates of electrical stimulation that are used in most of the current cochlear implant devices (stimulation pulse rates are around 1000 pulses-per-second in most devices), with no adaptation occurring in some cases and some stimulated fibers (Miller et al., 2008; Zhang et al., 2007). In contrast the extent of auditory-nerve adaptation in acoustic hearing, i.e. the ratio on stimulus to steady-state response of a fiber to a constant stimulus, is approximately invariant across stimulus frequency and intensity level (Smith and Zwischenlock, 1975). This suggests that despite the existing mechanisms, reproducing hair-cell adaptation behavior in electrically-stimulated auditory nerve may restore some of the speech enhancing mechanisms of normal auditory-nerve responses and improve CI outcome.

One indirect approach to reproducing adaptation behavior in auditory-nerve responses is by introducing decays in the level of stimulation at the neural level. This is similar to the depletion of hair-cell transduction that apparently reduces the postsynaptic input to the auditory neurons (Safieddine et al., 2012). In cochlear implants the adaptation-like decay can be applied in speech processors to the level of electrical pulses delivered to implant electrodes. Adaptation-inspired modifications have been previously implemented in the processing strategies of CIs and have shown benefits of adding adaptation-like behavior to electrical stimulation (Geurts and Wouters, 1999; Holden et al., 2005; Koning and Wouters, 2012; Vanden). The previous methods generally detected and amplified rapid increments (onsets) in the envelope of each spectral channel of the signal. Channel envelopes contain important cues for speech perception (Shannon et al., 1995), and are used in most CI processing strategies to modulate level of electrical pulses delivered to implant electrodes. While onset enhancement techniques implement an important aspect of adaptation, they disregard decrement (offset) enhancement aspects of hair-cell adaptation (or the after effects of adaptation) and therefore do not fully realize a mechanism of acoustic auditory-nerve adaptation. The present study introduces and evaluates a signal processing method that applies a physiological model of hair-cell transduction adaptation to the time-varying level of electrical signal delivered to each implant electrode. The adaptation model expands the range of all rapid variations (including both increments and decrements) in the level of electrical signal. Therefore it can potentially facilitate detection and discrimination of envelope patterns by CI users and may aid, in particular, those CI users who are less sensitive to envelope variations and cannot easily use these cues for speech perception (Fraser and McKay, 2012; Fu, 2002; Garadat et al., 2012).

Section 2, next, describes the details of the adaptation-based signal processing strategy and the experimental procedures that were used to evaluate the new strategy. Results comparing consonant and sentence recognition with the adaptation-based and the non-processed clinical strategies are presented in Sections 3 and 4 is the overall discussion.

2. Methods

2.1. Adaptation-based signal processing

The adaptation-based processing strategy was implemented by adding a model of auditory-nerve short-term adaptation (Smith, 1977; Smith and Zwischenlock, 1975, 1983) to the CI processing strategy. Adaptation of auditory-nerve responses to acoustic stimuli has been described by a linear additive process, and its “step response”, i.e. adapted response of an auditory nerve fiber to a constant intensity acoustic tone, has been modeled by a combination of several decaying exponentials. The exponential decay time constants can vary from a few milliseconds up to tens of milliseconds spanning a range of rapid to short-term adaptation. The effective time constant of the short-term auditory-nerve adaptation (approximately 40 ms averaged over stimulated fibers and stimulation levels) has been proposed to be more related to perception of speech features than the short time constant (a few millisecond) of the rapid adaptation (Delgutte, 1980). In this exploratory study we used a simple model of short-term adaptation for investigating the feasibility and effects of implementing adaptation in the CI processing strategies. The adaptation model can be described by a single exponentially decaying response to a constant intensity stimulus, analogous to a linear first-order high-pass filter. The transfer function (the ratio of output and input) of the modeled adaptation process is as follows:

\[ H(S) = \frac{1}{\tau S + 1} \]

The transfer function \( H \) is characterized by \( \tau \) and \( \rho \) parameters, where \( \tau \) is the decay time constant, and \( \rho \) is the onset enhancement gain of the response to a step input. The high-pass process of (1) emphasizes rapid signal variations while preserving the energy of slower signal variations. The cut-off frequency of the high-emphasis filter (fc) is related to \( \tau \):

\[ fc = \frac{1}{2\pi \tau} \]

Fig. 1a shows the output of the adaptation model with
parameters $\tau = 25 \text{ ms} (fc = 6.4 \text{ Hz})$ and $\rho = 2$ to a pulse-shaped input signal. It can be observed that the adapted output signal has a larger onset and offset change compared to the input signal and its decayed steady-state level asymptotes the level of the original signal. The negative output during the post-stimulus recovery period resembles the post-stimulus reduction in spontaneous firing rates of adapted auditory-nerve neurons (Smith, 1977), and reduces the response to an immediately following stimulation. In this simple model the time constant of the post-stimulus recovery from adaptation is the same as the time constant of the adaptation decay ($\tau$). The response of the same adaptation model to a 10 Hz sinusoid (Fig. 1b) demonstrates that adaptation expands the range of signal variations for frequencies higher than cut-off frequency ($fc$).

A block diagram illustrating the different stages of the adaptation-based processing strategy that was implemented in this study is shown in Fig. 2. Adaptation-based strategy was constructed by adding an adaptation stage (blocks within the dashed box) to the levels of the current pulses of each electrode at the output of the manufacturer’s clinical processing strategy, before electrical pulses were delivered to the implant electrodes. Adaptation stage followed the filtering, envelope extraction and amplitude compression stages that were performed in the clinical strategy. This was inspired by hair-cell transduction adaptation that follows the filtering and compression in the cochlea and presumably results in reduced postsynaptic input to the auditory neurons (Safieddine et al., 2012). The time sequence of pulse levels for each electrode was first extracted from the overall stimulation pattern at the output of the ACE strategy, and its sampling frequency was set equal to the channel stimulation rate in the subject’s clinical processor. Electrical levels were converted from clinical units to linear micro-amps before applying the adaptation process. This was done because the adaptation process applied a multiplicative gain factor to rapid signal variations, and the gain applied to the linear current level (in micro-amps) of an electrode is expected to cause a homogenous increase in the stimulation levels across the stimulated neural fibers (McKay, 2012). For instance, doubling the micro-amps level of an electrode would double the current that each stimulated neuron experiences assuming that the impedances in the neural tissue remain unchanged. To represent onset increments in terms of increments in perceptual loudness, the $\rho$ parameter was represented as a proportion of each electrode’s dynamic range (DR), which was the dB difference between electrode’s threshold and maximum comfort levels in the patient’s clinical map. This way, a fixed onset increase parameter $\rho$ produced an approximately equal perceptual loudness increase across electrodes of the implant array.

The $\rho$ (as percentage of electrode’s dynamic range) and $\tau$ parameters were equal across the electrode array. The output of the adaptation process was half-wave rectified to remove the negative values before it was used to modulate the level of electrical current pulses delivered to an implant electrode. Finally, the arrays of processed levels of pulses for different electrodes were recombined to produce the stimulation pattern that was streamed to the subject’s implant.

### 2.2. Experiments

Performance of the adaptation-based strategy was assessed in a series of sentence recognition in quiet and consonant identification in quiet and in background noise tests. Adaptation process is expected to affect mostly the dynamic time-varying features of speech rather than the time-invariant static features. Vowel recognition was not evaluated, because it relies more heavily on static formant cues and should be less affected by adaptation than consonants.

#### 2.2.1. Subjects

Seven native English speaking adult users of Nucleus cochlear implants participated in this study. Each subject’s implant type and the strategy in their clinical processor are shown in Table 1. The subjects had been using the ACE (Advanced Combination Encoder) strategy in their device for at least one year at the time of testing. Subject recruitment and testing was approved by the Syracuse University Institutional Review Board (IRB) and consent was obtained for each participant.

#### 2.2.2. Stimuli

Stimuli for consonant recognition tests were recordings of aCa syllables (from 16 consonants: /b/,/d/,/g/,/t/,/k/,/p/,/m/,/n/,/v/,/j/,/f/,/z/,/ch/,/sh/,/s/,/l/) by 6 different speakers (3 female). We used the

---

**Table 1**

<table>
<thead>
<tr>
<th>Implant type</th>
<th>Strategy</th>
<th>Channel rate (pps)</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1 CI24</td>
<td>ACE</td>
<td>900</td>
</tr>
<tr>
<td>S2 Freedom</td>
<td>ACE</td>
<td>900</td>
</tr>
<tr>
<td>S3 CI24</td>
<td>ACE</td>
<td>900</td>
</tr>
<tr>
<td>S4 CI24</td>
<td>ACE</td>
<td>1200</td>
</tr>
<tr>
<td>S5 Freedom</td>
<td>ACE</td>
<td>900</td>
</tr>
<tr>
<td>S6 Freedom</td>
<td>ACE</td>
<td>900</td>
</tr>
<tr>
<td>S7 CI24</td>
<td>ACE</td>
<td>900</td>
</tr>
</tbody>
</table>

---

**Fig. 2.** A block diagram of the adaptation-based processing strategy. The blocks within the dashed box are the adaptation stage that is added to the clinical strategy in the CI processor.
adaptation-based strategy with a pulse by pulse basis. The upper right panel (b) is the output of a single-channel adaptation-based strategies. The upper left panel (a) shows the output of electrode 20 stimuli to an intensity of the input sound is increased. We normalized speech current levels. Electrode levels approach the comfort level as the comfort level occurs when NMT transforms envelope levels to mapped to the maximum comfort level. Therefore no clipping at level all speech sounds remained within the range between maximum possible envelope level (which is +1 in Matlab) is mapped to the maximum comfort level. Therefore no clipping at comfort level occurs when NMT transforms envelope levels to current levels. Electrode levels approach the comfort level as the intensity of the input sound is increased. We normalized speech stimuli to a fixed RMS (Root Mean Square) level of 0.11. At this RMS level all speech sounds remained within the range between −1 and +1 and were not clipped. The RMS equalized stimuli were reported as comfortably loud by all the tested subjects. The stimuli were expected to be perceived equally loud by all the subjects assuming that stimulations at maximum comfort levels were perceived equally loud across subjects and across electrode arrays. Before normalizing levels, the original speech signals were downsampled to 16 KHz using Praat software (Boersma, 2001). The processing in the default NMT requires a sampling rate of 16 KHz for the input signal, which is analogous to the A/D sampling rate in the clinical speech processors.

The stimuli processed by the adaptation-based processing were generated by applying the adaptation processing stage (Fig. 2) to the current level output of the ACE strategy for each of the active electrodes as described in section 2.1 above. An example of the output of the ACE and adaptation based processing for syllable ‘aka’ is shown in Fig. 3. The time sequence of the amplitude of the actual pulse outputs of electrode 20 (which is typically assigned to the center frequency of 500 Hz in Nucleus implants) is shown in the left panel. The right panel shows the amplitude of the pulse outputs for syllable ‘aka’ generated by a single channel strategy (which uses the overall broad-band amplitude envelope of the speech signal to modulate pulses on a single electrode). The upper rows are the non-processed outputs (the outputs of the clinical strategy), and the lower rows are after applying the adaptation process with parameters \(\tau = 25\) ms and \(\rho = 30\% DR\). It can be observed that compared to the output of standard ACE strategy, the range of moderate to rapid signal variations (including onsets) are expanded at the output of the adaptation-based strategy.

2.2.3. Estimating optimized adaptation parameters

Consonant recognition tests were initially performed to obtain a rough estimation of the optimized \(\rho\) and \(\tau\) parameters of the adaptation-based strategy, individualized for each subject. Combinations of a number of \(\tau\) (10 ms, 25 ms, 50 ms, and 75 ms) and \(\rho\) (30% and 50% of each electrode’s dynamic range) parameters were used to generate several adaption-based strategies. The \(\tau\) s were on the order of 10’s of millisecond, consistent with the range of short-term adaptation in acoustically stimulated auditory neurons (Smith et al., 1983). The values for the \(\rho\) parameter were chosen based on preliminary tests that showed that smaller values were barely noticeable and larger values could generate uncomfortably loud stimuli. Closed set identification of two blocks of aCa syllables presented in quiet was obtained for different adaptation-based strategies (generated from different combinations of \(\tau\) and \(\rho\) parameters) as well as the non-processed ACE strategy. Each block consisted of 96 trials that contained 6 productions of each of the 16 consonant by different speakers (3 female). No feedback was provided during these tests. Presentation of blocks of different strategies was in random order. To prevent learning effects, the second block for each strategy condition was presented after the first block of all other conditions was presented. Thus the stimuli processed by each strategy could be presented at the beginning, middle and end of a testing session. Custom software presented stimuli and collected subject responses. The combination of \(\tau\) and \(\rho\) parameters that produced the highest consonant identification score was used to create a personalized adaptation-based envelope enhancement (Adapt-EE) strategy for each subject.

2.2.4. Speech perception with Adapt-EE and ACE strategies

To obtain acute comparisons between the Adapt-EE and the clinical ACE strategies, sentence and consonant recognition tests were performed with these two strategies. Speech recognition tests were performed in the order presented in this section.

2.2.4.1. Consonant recognition in quiet. Consonant recognition patterns in quiet were obtained for each of the ACE and Adapt-EE strategies by presenting three blocks of the consonants tests that were used in the initial tests of section 2.2.3. In order to prevent learning effects, presentation of consonant blocks alternated between ACE and Adapt-EE strategies such that the stimuli processed by each strategy could be presented at the beginning, middle and end of a testing session. The starting strategy was randomized across subjects.

2.2.4.2. Sentence recognition in quiet. Sentence recognition scores were obtained with each of the ACE and Adapt-EE strategies by measuring percentage of correctly identified keywords in two lists of 10 HINT sentences for each strategy (four lists in total). The four lists were randomly assigned to the two strategies and list presentation alternated between the ACE and Adapt-EE strategies. The starting strategy was randomized across subjects. Before starting the test, sentences of a different sentence list processed with the clinical ACE strategy were presented to the subjects to familiarize them with the task.

Fig. 3. An example of the implant signal for syllable ‘aka’ generated by ACE and adaptation-based strategies. The upper left panel (a) shows the output of electrode 20 on a pulse by pulse basis. The upper right panel (b) is the output of a single-channel strategy. The lower panels are the output levels of pulse sequences after applying an adaptation-based strategy with \(\tau = 25\) ms and \(\rho = 30\% DR\).
2.2.4.3. Consonant recognition in background noise.

Identification scores of consonants in background noise were obtained by presenting three blocks of consonant tests with each strategy in the presence of eight-talker babble noise (4 female talkers at 10 dB, 6 dB and 2 dB SNRs). Testing of each SNR was completed before testing a new SNR and the presentation order of different SNRs was randomized across subjects. For each SNR, the three blocks were alternated between Adapt-EE and ACE strategies to prevent learning effect. The starting strategy was randomized.

2.2.5. Training with the adaptation-based strategy

Speech signals processed by the Adapt-EE strategy were novel to the subjects who used ACE strategy in their clinical processors. The users may have required a period of exposure or training to be able to adapt to the novel signal and use the potential cues provided by the adaptation process. To evaluate the effects of a short training period on speech perception performance with the Adapt-EE strategy, consonant recognition was assessed in five of the subjects after they were provided with a short training session in the lab. Training was performed after the pre-training consonant and sentence recognition tests were completed. During training, 8 to 10 blocks of consonant identification tests were administered while feedback was provided on the correct answer. Only the stimuli processed with the Adapt-EE strategy were used in these feedback-provided consonant tests. After the training blocks, post-training consonant scores were obtained for both ACE and Adapt-EE strategies in three blocks of consonant tests for each strategy without feedback. The post-training test was identical to the pre-training consonant tests of section 2.2.4 and presented three blocks of consonants processed with each strategy in alternating order. The post-training performance of ACE strategy was obtained as a control condition.

2.2.6. Perception of overall envelope cues

Adaptation expanded rapid envelope variations and was expected to enhance both across-channel spectral and within-channel temporal envelope cues. To investigate the effects of adding adaptation on the perception of purely temporal consonant features, consonant recognition was obtained with single-channel experimental programs that activated a single electrode of the array, similar to Azadpour and McKay (2014). In the single-channel strategies, current level on a single implant electrode was modulated by the overall temporal envelope of the speech signal. The overall amplitude envelope of the signal carries information of some consonant features, particularly manner of articulation and voicing features (Azadpour and McKay 2014; Fishman et al., 1997; Van Tasell et al., 1987). Because only one electrode is activated, no spectral cue is available in a single channel strategy. An electrode located in the middle region of the array was chosen for the single-channel strategy. Adaptation-based single-channel strategy was created by applying the adaptation model with the best \( \tau \) and \( \rho \) parameters obtained in the preliminary consonant tests (Table 2) to the output of the single-channel strategy. The active electrode’s stimulation rate was set at the channel stimulation rate in the subjects’ clinical processor. The threshold and maximum levels of this electrode were measured using psychophysical procedures. The levels in the clinical maps were not used for creating single-electrode strategies because these levels are usually set for multiple-electrode stimuli and may be too quiet when only a single electrode is activated. Maximum comfortable level of the active electrode was obtained in a method of adjustment in which the subject adjusted the level of the pulse train on the target electrode until the perceived loudness reached the maximum comfort level. Threshold was obtained in an adaptive three-interval three-alternative forced choice paradigm: in one of the intervals, randomly chosen, a pulse train was presented on the target electrode and the other two intervals contained silence. The subject’s task was to point to the stimulus interval. The level of the pulse train was adapted in a 2 down-1 up procedure with a step size of 4 CL (clinical level) initially and 2 CL after two reversals (One CL increase is equivalent to 0.157 and 0.176 dB current increase in Nucleus Freedom and CI24 implants respectively). The procedure continued for a total of 10 reversals of which the average of the last 6 was set as the threshold level. The stimuli created by the single-channel experimental maps were reported to be comfortably loud by all subjects and were expected to be equally loud across subjects.

3. Results

Initial consonant scores for the stimuli processed by several parameters of the adaptation-based strategy are shown in Table 2. Some of the subjects were only evaluated with a subset of the parameters. The scores corresponding to the sets of \( \tau \) and \( \rho \) parameters that produced maximum consonant scores are shown in underlined bold. These parameter sets were used to create individualized Adapt-EE (adaptation-based envelope enhancement) strategies and were used to assess performance with the adaptation-based strategy. Note that the overall loudness of stimuli increased as \( \tau \) and \( \rho \) parameters were increased: increasing \( \rho \) made the range of envelope variations of the stimuli larger, and increasing \( \tau \) made the cut-off frequency of the high-pass adaptation filter smaller (a wider range of signal frequencies was expanded). Expanded envelope variations could make the overall stimulus louder even if the average envelope remained constant, perhaps because envelope peaks contribute more to the overall loudness than the average envelope (McKay and Henshall, 2010). However, the observed higher performance with the adaptation-based strategy cannot be exclusively attributed to the overall loudness; in some cases louder stimuli produced by larger \( \tau \) and \( \rho \) parameters resulted in poorer performance (Table 2). Presumably the enhanced cues by the adaptation process contributed to better speech perception performance and not the loudness of the stimuli. Since adaptation expanded the range of rapid signal variations, electrode

**Table 2**

Initial percent consonant scores with different \( \rho \) (onset enhancement factor) and \( \tau \) (decay time constant) parameters. Empty cells refer to non-tested parameters.

<table>
<thead>
<tr>
<th>Clinical ACE strategy (no adaptation)</th>
<th>( \rho = 30% ) DR</th>
<th>( \rho = 50% ) DR</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \tau = 10 \text{ ms} )</td>
<td>( \tau = 25 \text{ ms} )</td>
<td>( \tau = 50 \text{ ms} )</td>
</tr>
<tr>
<td>S1</td>
<td>50</td>
<td>42.7</td>
</tr>
<tr>
<td>S2</td>
<td>71.3</td>
<td>—</td>
</tr>
<tr>
<td>S3</td>
<td>59.9</td>
<td>59.3</td>
</tr>
<tr>
<td>S4</td>
<td>55</td>
<td>51</td>
</tr>
<tr>
<td>S5</td>
<td>71.9</td>
<td>—</td>
</tr>
<tr>
<td>S6</td>
<td>53.8</td>
<td>—</td>
</tr>
<tr>
<td>S7</td>
<td>72.4</td>
<td>—</td>
</tr>
</tbody>
</table>
stimulation levels could sometimes be greater than the maximum comfort levels in the clinical map. However, these greater levels usually occurred at short time intervals, because of relatively short duration of speech signal variations and relatively short adaptation time constants. Subjects reported that the stimuli remained comfortable in all the conditions tested here.

3.1. Sentence recognition

Fig. 4 shows sentence comprehension scores in quiet with both Adapt-EE and clinical strategies. Performance improved with the Adapt-EE strategy in all the tested subjects and the improvement varied between 2% and 17% correct keywords. A paired t-test analysis confirmed that scores obtained with the Adapt-EE strategy were significantly higher than the scores obtained with the clinical strategies ($t(6) = 4.58, p = 0.004$). Adding a simple model of acoustic short-term adaptation improved speech understanding of the tested CI users, and the improvement varied substantially across users.

3.2. Consonant recognition in quiet and effects of training

Correct consonant recognition scores in quiet with the ACE and the Adapt-EE strategies are shown in Fig. 5. Individual subject scores and the subject average scores are presented. Pre- and Post-training scores, respectively, refer to scores obtained before and after training with feedback was performed in the lab. Error bars are the 95% binomial confidence intervals on the individual subject graphs, and standard errors on the subject average graph. To examine the effects of both training and strategy type on consonant identification scores, two-way repeated-measures ANOVA was used. The analysis showed a main effect of processing strategy ($F(1,19) = 23.47, p = 0.008$), no main effect of training ($F(1,19) = 5.68, p = 0.076$), and no interaction of training and processing strategy ($F(1,19) = 3.73, p = 0.126$). This confirms that the adaptation-based strategy improved consonant perception over the clinical ACE strategy, despite the fact that it was novel to the subjects. Although the observed improvement might be an underestimation of the real improvement after long-term use of the new strategy, a short lab-based training did not show a significant effect on consonant scores. A longer practice period might have been required for the subjects to significantly adapt to the consonant cues provided by the Adapt-EE strategy.

3.2.1. Information transfer analysis of consonant features

In order to obtain a deeper insight on how adding adaptation affected perception of cues related to different consonant features, pre-training consonant confusion pattern results were analyzed for information transfer analysis of consonant features (Miller and Nicely, 1955). Information transfer analysis provides a more global vision of the patterns of confusions than the percent correct score, which only uses the diagonal elements of the confusion patterns. Information transfer analysis was applied to consonant feature matrices to evaluate the extent to which the consonants that shared certain features were perceived similar to each other and distinct from the other consonants. To construct feature confusion matrices, consonant confusion matrices were reorganized by grouping together the consonants that belonged to certain feature categories. Manner feature matrices were obtained by categorizing consonants in three categories: plosives and affricates (/b/,/p/,/d/,/t/,/k/,/g/,/j/,/ch/), fricatives (/s/,/sh/,/z/,/f/,/v/,/l/) and nasals (/m/,/n/). Place feature matrices were obtained by categorizing the consonants into front (/b/,/p/,/f/,/v/,/m/), middle (/t/,/d/,/n/,/s/,/z/,/l/) and back (/k/,/g/,/ch/,/sh/,/j/) categories. Voicing feature matrices were obtained by categorizing the consonants as voiced (/b/,/d/,/g/,/m/,/n/,/z/,/j/,/v/,/l/) and unvoiced (/k/,/p/,/t/,/ch/,/s/,/sh/,/f/). Only three categories were used for manner and place features to minimize the possible bias in information transmission analysis (Azadpour et al., 2014; Sagi and Svirsky, 2008). 

Fig. 4. Sentence recognition scores with the clinical and the Adapt-EE strategies. Error bars on the subject average scores are standard errors.

Fig. 5. Consonant identification scores with the clinical and the Adapt-EE strategies for individual subjects and subject averages. Pre- and post-refer to scores obtained before and after training with feedback, respectively. Error bars on the individual subjects plots are binomial confidence intervals, and on the average plot are standard errors.
affricate consonants (/ch/, /j/) were included in the plosive category because of their short rise times when used in aCa contexts (Howell and Rosen, 1983). Fig. 6 shows information transfer of each feature for each subject and the subject average. The error bars on the individual subject graphs are 95% confidence intervals of information transfer analysis obtained using a recently introduced bootstrapping method (Azadpour et al., 2014). According to Azadpour et al. (2014), these confidence intervals were relatively bias-free given the relatively large number of stimuli presented in three blocks of consonant tests (18 repetitions for each of the 16 consonant). The error bars on the subject average graph are standard errors. Within subject comparisons of confidence intervals in Fig. 6 suggests that in most of the subjects the Adapt-EE strategy significantly improved information transmission of manner and place features, but not information transmission of voicing feature. Information transmission of each feature was compared between the clinical and the Adapt-EE strategies using paired t-test analysis, which revealed that Adapt-EE strategy significantly improved information transmission of manner (t(6) = 4.76, p = 0.003) and place features (t(6) = 4.85, p = 0.003) but not information transmission of voicing feature. Information transmission of each feature was compared between the clinical and the Adapt-EE strategies using paired t-test analysis, which revealed that Adapt-EE strategy significantly improved information transmission of manner (t(6) = 4.76, p = 0.003) and place features (t(6) = 4.85, p = 0.003) but not information transmission of voicing feature (t = 0.98, p = 0.37). The analysis remained significant after Bonferroni correction for multiple comparisons. Manner feature is related to both within and across channel dynamic cues, while place feature is more dependent on channel cues and requires comparing energy variations across multiple frequency channels (Azadpour and McKay 2014; Fishman et al., 1997). These results suggest that enhancing channel envelope variations using the adaptation-based strategy may improve access to both within and across channel consonant cues. Interestingly, however, transmission of voicing feature was not improved by the adaptation-based strategy. Voicing feature is related to periodicity, duration and overall energy of channel envelopes. Envelope periodicity was presumably enhanced by the adaptation process that expanded the range of sinusoidal envelope variations (see Fig. 1b), thus voicing perception based solely on periodicity cue was expected to be improved by Adapt-EE strategy. These results suggest that envelope periodicity may not be the only cue to voicing for CI users. Overall energy and energy variation of consonant onsets, which can be affected by the adaptation-based envelope enhancement strategy, may also provide significant cues to perception of voicing feature in some CI users.

3.3. Consonant identification in background noise

Consonant identification scores with the ACE and Adapt-EE strategies at three different SNR levels of multi-talker babble noise are shown in Fig. 7. Error bars are 95% confidence intervals on the individual subject graphs, and standard errors on the subject average graph. To evaluate the effects of both noise level and the processing strategy, a two-way repeated-measures ANOVA was used. This analysis showed significant main effects of processing strategy (F(1,29) = 7.93, p = 0.048) and noise level (F(2,29) = 18.95, p < 0.001) and a significant interaction of the two (F(2,29) = 6.52, p = 0.021). Post-hoc analysis showed that the Adapt-EE strategy improved consonant scores at the SNRs of 10 dB (p = 0.003) and 6 dB (p = 0.047), but not at 2 dB (p = 0.66). Adaptation-based

---

Fig. 6. Information transfer analysis of consonant features with the clinical and Adapt-EE strategies for individual subjects and the subject averages. Error bars on the individual subjects plots are confidence intervals, and on the average plot are standard errors.

Fig. 7. Consonant identification scores at different levels of multi-talker babble noise for individual subjects and subject averages. Error bars on the individual subjects plots are binomial confidence intervals, and on the subject average plot are standard errors.
strategy enhanced consonant cues in speech babble noise but only at lower noise levels. Babble noise is dynamic and contains rapid envelope variations. The adaptation process presumably emphasized noise variations and those variations may have masked the enhanced time-varying cues of the target speech, particularly at higher noise levels. However, at the moderate SNRs of 10 and 6 dB the emphasized speech cues by the adaptation-based processing seem to dominate the enhanced background noise. The adaptation-based processing does not seem to fall behind the ACE strategy at higher noise levels, at least up to 2 dB SNR.

3.4. Consonant recognition with single-channel strategies

Consonant recognition patterns obtained with the non-processed and adaptation-based processed single-channel programs were analyzed for information transmission of manner of articulation, place of articulation and voicing features. Fig. 8 shows information transmission of consonant features obtained with the non-processed clinical and the adaptation-based processed single-channel strategies. The electrode that was activated is shown on the individual subject panels. S1 was tested twice with E12 and E8 as active electrodes. Information of place of articulation feature is known to be related to the across channel signal variations and was expected to be little transmitted by the single-channel strategies. The present study provides evidence that speech communication of cochlear implant users can be improved by incorporating hair-cell adaptation mechanism in CI signal processing. Short-term auditory-nerve adaptation, which may not be fully present in electrical stimulation, has been hypothesized to play an important role in speech perception. To restore the short-term adaptation mechanism in CI stimulation, a physiological model of the adaptation observed in the acoustically-evoked auditory nerve responses was applied to the current level output of the ACE strategy of the Nucleus processors. Adaptation was modelled by a linear first-order high-pass process and was characterized by its decaying response to a step stimulus: decay time constant (τ) and onset enhancement gain factor (ρ). The performance of the adaptation-based strategy was assessed in seven CI user subjects. For each subject, an approximation of the adaptation parameters that resulted in the highest consonant recognition scores was first estimated. Initial consonant scores were obtained with several adaptation-based strategies at different combinations of τ and ρ parameters. The parameter combination that resulted in the highest consonant scores in each subject was used to construct a personalized adaptation-based envelope enhancement (Adapt-EE) strategy. Further tests confirmed that the adaptation-based strategy outperformed the clinical strategy for sentence identification in quiet and consonant identification in quiet and in the presence of multi-talker babble noise. The average improvement was approximately 8% for sentence recognition and 6% for consonant identification in quiet. The improvement in sentence comprehension ranged between 2% and 17% in the tested subjects. Vowel identification was not evaluated in this study because adaptation was expected to minimally affect perception of vowels that are typically longer and contain slower energy variations than consonants. However, it is possible that the vowel cues provided by the transitions of vowels with adjacent consonants, particularly in the relatively short vowels produced in natural conversations, may be affected by the adaptation model. The improvement observed in the recognition of sentences suggests that vowels were not at least heavily distorted by the adaptation model. A detailed feature analysis of consonant recognition patterns in quiet showed that adaptation improved information transmission of manner and place of articulation features. This suggests that adding the adaptation process presumably enhanced across- and within-channel
envelope variations that provide important cues to manner and place features. Information transmission of voicing feature, however, was not significantly affected by adding adaptation to the ACE strategy, suggesting that perception of voicing feature relies on some other cues than those that are enhanced by the adaptation model. Particularly, envelope periodicity that is presumably enhanced by adding adaption (as shown in Fig. 1b) may not provide major cues to voicing. Identification scores of consonants presented in babble noise were significantly improved, but only at higher SNRs. Consonant recognition was similar between Adapt-EE and ACE strategies at the lowest tested SNR. At lower SNRs, the enhanced dynamic speech cues may be masked by the noise variations that are themselves enhanced by the added adaptation process. Analysis of consonant identification with single channel programs was consistent with the results obtained with multiple-channel programs and showed that adaptation improved within-channel cues to manner of articulation but not voicing.

Improvements in speech perception have been previously shown with signal processing methods that implemented some aspects of neural adaptation mechanism in the CI signal (Geurts and Wouters, 1999; Koning and Wouters, 2012; Vandali, 2001). Geurts and Wouters (1999) implemented an algorithm that detected and magnified rapid increments (onsets) in the envelope of each spectral band of the signal. The onset enhancement method improved identification of stop consonants, and perception of phonemes in consonant-vowel-consonant word contexts. Vandali (2001) developed a signal processing method that emphasized short-duration transients in envelope of each frequency band of the signal, and implemented the algorithm in the subjects’ clinical processors. Transient enhancement improved sentence comprehension and phoneme perception after a take-home experience with the new strategy. However, in a later study, less improvement was observed when the transient enhancement algorithm was added to the ACE processing strategy in the newer Nucleus processors (Holden et al., 2005). An important feature of the current study that makes it distinct from the previous studies is the use of a physiological model of neural adaptation that creates a “recovery” period at the offset of stimulation and expands the range of all rapid variations in the implant signal (both increments including onsets and decrements including offsets). Channel envelopes carry important dynamic time-varying speech cues, particularly for CI users who do not have access to temporal fine structure information. Expanding the range of rapid envelope variations can enhance CI users’ detection and discrimination of envelope patterns (Fraser and McKay, 2012; Fu, 2002) and can potentially improve their speech perception. Envelope expansion techniques have been applied to acoustic signals in normal hearing listeners and have shown moderate or no improvements in perception of distorted speech signals (van Buuren et al., 1999) or spectrally-degraded vocoded CI simulations (Fu and Shannon, 1999). This study shows that adding an adaptation-based envelope expansion to the CI signal can improve CI users’ perception of some speech cues such as the cues to consonant manner and place of articulation features. The method had less effect on the perception of some other speech cues such as consonant voicing cues. This may be one of the reasons that the overall speech recognition improvement was incremental at least in some subjects. Although the outcome of the adaptation-based envelope expansion method of this study seems to be comparable to the outcome of the mentioned envelope enhancement methods, the different studies cannot be easily compared because of differences in the subjects’ clinical processor strategies, speech testing procedures and the level of experience the subjects gained with the novel strategy.

The present study used a simple adaptation model as a first attempt to evaluate the effects of adding hair-cell adaptation to the CI signal. Larger benefits of adding adaptation may be obtained by making modifications to the adaptation model and the implementation of the strategy. One possible modification would be to use a more complex model of auditory-nerve adaptation that has multiple time constants for the onset decay and the post-stimulus recovery periods. Particularly, models with asymmetric time constants that are longer for the recovery period may be more physiologically valid (Smith et al., 1983). A second possible modification would to use electrode-specific values of r and p parameters instead of using the same parameters for all the electrodes. The optimized adaptation parameters for each electrode may depend on the fibers that are being stimulated and/or the importance of the speech envelope cues that are being transmitted by that electrode. Another possible modification is to add adaptation to an earlier stage of ACE strategy, for instance to the envelope of each spectral band of the acoustic signal. However in this case because the nonlinear acoustic-electric mapping will be applied to the output of the adaptation model, the expansion gain factor may be level dependent, which is not consistent with the adaptation observed in the auditory nerve. In the method of this study, adaptation was added to the output of the ACE strategy to better mimic the hair-cell transduction adaptation that occurs after cochlear compression and reduces input to auditory-nerve fibers. However, on the other hand, applying adaptation to an earlier stage of ACE strategy may increase the chance of the channels that convey more dynamic speech cues to be selected in each cycle of the ACE strategy. Overall, the present study suggests potential for improvements in CI processing strategies by applying adaptation-like envelope enhancement techniques and holds out the promise that a more refined model and/or better optimization of parameters and fittings could result in larger improvements.

Acknowledgements

We thank the CI user subjects for their participation in this study. We also thank Lorraine Pawson for technical assistance, Mark Robert and Karen Doherty for providing speech stimuli, and Cochlear Corporation for making research software and hardware tools available. This research was supported by the Jim and Arlene Gerber Fund for Auditory Research at the Institute for Sensory Research, LC Smith College of Engineering and Computer Science, Syracuse University, and a grant from the National Organization for Hearing Research (NOHR).

References


