UNIVERSITY OF COPENHAGEN

Identification of vowel length, word stress and compound words and phrases by postlingually-deafened cochlear implant listeners.

Morris, David Jackson; Magnusson, Lennart; Faulkner, Andrew; Jönsson, Radoslava; Juul, Holger

Published in:
Journal of the American Academy of Audiology

DOI:
10.3766/jaaa.24.9.11

Publication date:
2013

Document version
Early version, also known as pre-print

Citation for published version (APA):
Morris, D. J., Magnusson, L., Faulkner, A., Jönsson, R., & Juul, H. (2013). Identification of vowel length, word stress and compound words and phrases by postlingually-deafened cochlear implant listeners. *Journal of the American Academy of Audiology*, *24*(9), 879-890. https://doi.org/10.3766/jaaa.24.9.11

J Am Acad Audiol 24:879–890 (2013)

Identification of Vowel Length, Word Stress, and Compound Words and Phrases by Postlingually Deafened Cochlear Implant Listeners

DOI: 10.3766/jaaa.24.9.11

David Morris*
Lennart Magnusson†
Andrew Faulkner‡
Radoslava Jönsson†
Holger Juul*

Abstract

Background: The accurate perception of prosody assists a listener in deriving meaning from natural speech. Few studies have addressed the ability of cochlear implant (CI) listeners to perceive the brief duration prosodic cues involved in contrastive vowel length, word stress, and compound word and phrase identification.

Purpose: To compare performance in the perception of brief duration prosodic contrasts by CI participants and a control group of normal hearing participants. This study investigated the ability to perceive these cues in quiet and noise conditions, and to identify auditory perceptual factors that might predict prosodic perception in the CI group. Prosodic perception was studied both in noise and quiet because noise is a pervasive feature of everyday environments.

Research Design: A quasi-experimental correlation design was employed.

Study Sample: Twenty-one CI recipients participated along with a control group of 10 normal hearing participants. All CI participants were unilaterally implanted adults who had considerable experience with oral language prior to implantation.

Data Collection and Analysis: Speech identification testing measured the participants' ability to identify word stress, vowel length, and compound words or phrases all of which were presented with minimal-pair response choices. Tests were performed in quiet and in speech-spectrum shaped noise at a 10 dB signal-to-noise ratio. Also, discrimination thresholds for four acoustic properties of a synthetic vowel were measured as possible predictors of prosodic perception. Testing was carried out during one session, and participants used their clinically assigned speech processors.

Results: The CI group could not identify brief prosodic cues as well as the control group, and their performance decreased significantly in the noise condition. Regression analysis showed that the discrimination of intensity predicted performance on the prosodic tasks. The performance decline measured with the older participants meant that age also emerged as a predictor.

Conclusions: This study provides a portrayal of CI recipients' ability to perceive brief prosodic cues. This is of interest in the preparation of rehabilitation materials used in training and in developing realistic expectations for potential CI candidates.

Key Words: Cochlear implants, speech acoustics, speech intelligibility

Abbreviations: CI = cochlear implant; DL = difference limen; F0 = fundamental frequency; F1 = first formant; NH = normal hearing; RMS = root mean square

*Speech Pathology and Audiology, Department of Scandinavian Studies and Linguistics, University of Copenhagen, Denmark; †Department of Audiology, Sahlgrenska University Hospital, Gothenburg, Sweden; ‡Speech, Hearing and Phonetic Sciences, University College London

David Morris, Speech Pathology and Audiology, Department of Scandinavian Studies and Linguistics, University of Copenhagen, Copenhagen S, Denmark; E-mail: dmorris@hum.ku.dk

Prosody is an integral and vital part of spoken communication that imbues utterances with dynamic variation. This variation is perceived as a result of acoustic changes in pitch, intensity, and rhythm during the course of an utterance. These acoustic changes are secondary or overlaid functions of speech, yet they can have semantic implications (Lehiste, 1970). The purpose of the present study was to assess the prosodic ability of cochlear implant (CI) listeners where prosody provided a distinctive semantic contrast, the perception of which is necessary for competent real-world listening. We tested this competence in quiet and in a noise condition in order to involve a situation relevant to everyday listening. The listening ability of CI participants was also characterized by measuring discrimination thresholds for pitch, intensity, duration, and vowel quality, which were used to derive associative links to performance in the natural utterance prosody tasks.

The ability of CI listeners to perceive changes in voice pitch, which is an essential component of prosody, has often been reported to be limited. Research has been directed toward prosodic features that occur over longer time courses, for instance, those that occur at the sentence level. Peng et al (2008) found that accuracy in identifying sentences as questions or statements was 70% for prelingually implanted older children and young adults while normal hearing (NH) participants performed at ceiling level. Studies of postlingually implanted adults using question and statement speech material recorded from both male and female speakers have reported similar mean identification rates. For example, Green et al (2005) reported mean identification rates of around 70%, and Meister et al (2009) found mean rates of approximately 80% while also observing near perfect rates in a NH group. Considerable attention has also been given to tasks that require voice pitch judgments including speaker and speaker gender identification (McKay and McDermott, 1993; Fu et al, 2005; Meister et al, 2009; Kovačić and Balaban, 2010; Schvartz and Chatterjee, 2012). These studies report deficiencies in CI listeners' ability to identify voice pitch changes and ascribe gender to a speaker, which are both prosodic-like tasks.

Prosodic cues can overlay sentences and phrases, but they can also affect, and may be features of, phonemes, syllables, and words. These units of speech have been used as test material in CI listening experiments that have examined prosodic changes over relatively brief time courses. A phoneme-based intonation contour identification task was included in the assessment of two different processing strategies in McKay and McDermott (1993). The aggregate scores from the four CI participants in this study showed that they could identify contour direction (rising, steady, and falling) at levels above chance when listening to both a male and a female speaker. A recent study by Most et al (2011) of prosody occurring over brief durations tested the ability of 23 CI participants in a task where they were required to distinguish between 12 pairs of bisyllabic Hebrew words where each pair differed only in word stress. Participants were tested with their CI alone and also with their CI and a hearing aid on the contralateral ear, that is, in the "bimodal" condition. The mean results of this study showed that in the CI alone condition, participants performed at around 78% (or 55% after the scores were corrected for guessing) and, in the bimodal condition, at approximately 83% (or 65% after the scores were corrected for guessing). Along with word stress, this study also investigated the listening ability of CI participants in a question and statement identification task and also a word emphasis identification task. They found no significant differences between the intragroup results on these three tasks in either the CI alone or the CI and hearing aid groups.

In the present study we tested the ability of CI and NH listeners to use brief prosodic cues to identify naturally occurring changes in vowel length and word stress and also to distinguish between compound words and phrases. The acoustic correlates that signal changes in word stress and compound words/phrases are mediated by pitch, intensity, and duration, while changes in vowel length are largely mediated by duration cues. Vowel length and word stress are critical features in many Scandinavian languages as they can denote semantic distinctions between words. For instance, in Swedish, *ful* means ugly, and *full* means complete or drunk. A translated example of word stress is the word *convict.* When the first syllable is stressed (*con*vict) it is a noun, and when the second syllable is stressed (*convict*) it is a verb. Generally, the Swedish language exhibits prominent stress patterns that can be transcribed by listeners even when the language is hummed (Svensson, 1974; Bruce, 1998). We also tested the ability of CI and NH listeners to identify compound words as opposed to phrases. For a listener, the disambiguation of compound words and phrases can be made by the use of the stress pattern that is contained within a word (word stress) and that which is contained within a phrase (phrasal stress). This might be supplemented with segmental duration information and prior knowledge of language specific distributional characteristics (Smith et al, 1989). The ability to identify compound words and phrases involves accurate placement of word boundaries. This perceptual ability is critical to parsing the speech stream, because there are semantic distinctions between compound words and phrases. An example of a compound word and a phrase pair is the word *blackbird* and the phrase *black bird.* The distinction between the two is exemplified in the sentence, "the raven is a black bird, but not a blackbird."

The effect of noise on the identification of brief prosodic cues in natural utterances by CI listeners has not been investigated. Noise generally has a deleterious effect on most measures of speech intelligibility by CI listeners (Hochberg et al, 1992; Fu et al, 1998; Nelson et al, 2003; Fu and Nogaki, 2005; Tobey and Geers, 2011). The magnitude of this effect is dependent on the test material and the signal-to-noise ratio (SNR) that is used. Data from Alkaf and Firszt (2007) shows that for a large group of postlingually implanted CI participants (n = 68) there was a reduction of approximately 13% in performance scores on the Hearing In Noise Test (HINT) when noise was introduced at an SNR of 8 dB. Similarly, Shannon et al (2011) found a decrease in consonant-nucleus-consonant word scores of approximately 22% when noise was added at 10 dB SNR.

The deleterious effects of noise on CI listening are attributable to factors that are associated with the operation of CIs and to limitations in the electrical-neural interface. Device factors include limited transmission of both spectral and temporal information. Electrical-neural interface limits also appear to restrict the number of effective spectral channels, so that, for example, performance in speech intelligibility in noise plateaus when the number of channels are increased above eight (Friesen et al, 2001). This plateau may be attributable to a broad spread of in-vivo excitation caused by the distance between the electrode and the receptor site. The net effect of these factors is that spectral detail is poorly represented to CI listeners, and listening in background noise is problematic. The encoding of more rapid temporal information is also limited by neural factors beyond the limits imposed by the speech processing (e.g., Zeng, 2002).

The noise condition in our study was included primarily to gain a better understanding of the real-world ability of CI listeners. We expected to observe a performance decrease with the introduction of noise in the NH group and a more pronounced decrease in the CI group. With the CI group we also suspected that noise would have a greater adverse effect on some tasks than others. It was anticipated that the reduced vowel and the finer intensity variation of unstressed syllables involved in the word stress task would be harder for CI listeners to use as identification markers in the presence of noise. It was also anticipated that the type of noise that was employed in this study, with spectral shaping based on male speech, would effectively interfere with the fundamental frequency (F0) of the male talker and thus render pitch cues less perceptible.

Vowel length tasks and other word based prosodic tasks have been incorporated into prosodic test batteries. For instance, the vowel length portion of the German prosodic battery introduced in Meister et al (2008) tests short and long vowel identification. Also, the second part of section two of the speech pattern contrast (SPAC) test provides stimuli for vowel quality perception tasks. These are ten series of CVC stimuli featuring different vowel quality, for instance, *Sue's, sees, saws,* and *says* (Boothroyd, 1984). The perception of brief prosodic cues by CI listeners warrants further study as CI recipients often undergo auditory training where these skills are drilled (for example, Plant, 1994). As many CI recipients undergo such training, they might be expected to be adept at performing these tasks. Given the focus that prosodic dimensions have received in rehabilitation and in testing, and the limited performance observed by CI listeners on these tasks (Meister et al, 2011; Most et al, 2011), there are grounds to suspect that CI listeners may not be able to develop the ability to identify some aspects of prosodic detail. The study of prosodic ability in natural utterances that exhibit different prosodic features is therefore relevant to rehabilitation, as it can guide clinical progress to material that is suitably challenging yet not perceptually unavailable to CI listeners.

To examine the extent to which naturally occurring brief duration prosodic detail is available to CI listeners, we compared their performance to NH listeners in tasks where participants were required to identify a word from an answer choice that included the target word or phrase and a similar word or phrase. Together these two answer choices constitute a prosodically contrastive minimal pair; that is, the two words are segmentally identical, or close to identical, but they differ prosodically when phonologically realized. In this study we used minimal pairs of naturally uttered stimuli to test CI and NH participants' ability to identify the target word in the absence of contextual information that would usually distinguish that word from other probable alternatives. We hypothesized that in minimal pair testing where only naturally occurring prosodic cues were available, the CI listeners would not be able to perform as well as the NH listeners.

In the analysis of this study the scores from the CI listeners in both quiet and noise conditions were used as the dependent variable in a regression model. We used this to examine some individual participant parameters including age and duration of implant listening and also discrimination thresholds, called here difference limens (DLs), for F0, intensity, duration, and vowel quality made with synthetic vowel stimuli. The regression analysis was included to identify which parameters and measurements were associated with better performance on the tasks of brief prosodic duration identification.

METHOD

Participants

CI participants (n = 21) were recruited from the Sahlgrenska University Hospital register. These were 8 males and 13 females with a mean age of 64.3 yr (range 40–82). All participants had been unilaterally implanted within the last nine years and did not have any cognitive handicap. They were all native Swedish speakers and had all had considerable auditory experience prior to the onset of their hearing loss. In instances where a participant wore a contralateral hearing aid, this was removed prior to testing. Details of this group are given in Table 1.

Table 1 also shows word recognition scores for all participants obtained prior to CI implantation with conventional hearing aids, and their most recent results with their CI, recorded during routine follow-up. These word scores were determined with results from the open set Swedish PB word identification test (for a detailed description of this test, see Magnusson, 1995). One list containing 50 words was used and was presented without the addition of noise, from a loudspeaker positioned 1 m in front of the subject in a sound treated room. With the exception of two participants, all scored above 50% with their CI at postimplant testing. An attempt was made to include participants who had high word recognition scores and those who showed a marked improvement between their pre- and post-CI scores, as this was taken as grounds for the assumption that these participants were deriving substantial communication benefit from their implant.

The control group (n = 10) was an ad hoc group who all reported normal hearing and had Swedish as their mother tongue. These were four males and six females with a mean age of 51 yr (range 36–70). This study was granted ethical approval by the ethics committee of the region Gothenburg (reference number 083-12).

Discrimination Measures for Prosodic Cues

Discrimination thresholds for the prosodic cues of F0, duration, intensity, and first formant (F1) were measured using the synthetic syllable [ba] as stimulus. An adaptively controlled three-interval three-alternative forced choice (AFC) task with a one-up one-down staircase tracked the 50% performance level. For the F0 discrimination, a rising/falling F0 contour, from an onset value of 100 Hz, was varied in 48 steps to have F0 peaks between 100 and 219 Hz, while all other acoustic parameters were constant. For duration discrimination the vowel portion of the [ba] alone was varied, being lengthened in 48 steps from 505 to 805 msec. For intensity discrimination, intensity alone was varied in steps of 1 dB up to a 20 dB increase. Finally, the vowel formant discrimination task examined differences in the steady-state F1 of the vowel portion of the syllable [ba] from a standard of 524 Hz in 48 steps up to 982 Hz. Perceptually, this is similar to a shift in vowel quality from /ə/ as in *Burberry,* to /ɑ/ as in *Bartholomew.* In this stimulus set, F0 and duration were constant, but the stimulus level was roved by a random value between plus and minus 4 dB. Stimuli for the discrimination tasks were synthesized with the Klatt synthesiser.

Table 1. Demographic Data of the CI Participants, Their CI Experience, and Implant Characteristics

Participant ID	Gender	Age	Implant Experience	Processing Strategy	Implant Processor	Implant Type	Etiology	PB Word Pre	PB Word Post
CI01	F	58	3	ACE	Freedom	Freedom contour	Hereditary	40	92
CI02	F	82	2	FSP	Opus2	Sonata	Unknown	0	46
CI03	M	49	6	ACE	CP810	Freedom contour	Unknown	16	88
CI04	F	41	0.25	ACE	CP810	CI422	Unknown	8	Unavailable
CI05	F	72	5	ACE	Freedom	Freedom contour	Unknown	0	72
CI06	M	67	5	ACE	Freedom	Freedom contour	Unknown	4	72
CI07	F	74	5	ACE	Freedom	Freedom contour	Postinfection	4	90
CI08	F	40	2	ACE	CP810	CI512	EVA syndrome	0	88
CI09	M	71	5	ACE	Freedom	Freedom contour	Hereditary	0	72
CI10	F	47	4	ACE	Freedom	Freedom contour	Perinatal	12	70
CI11	M	76	6	ACE	Freedom	Freedom contour	Otosclerosis	16	52
CI12	M	82	6	ACE	Freedom	Freedom contour	Ménière's disease	8	68
CI13	F	75	5	ACE	Freedom	Freedom contour	Otosclerosis	0	68
CI14	F	81	5	FSP	Opus2	Sonata	Ménière's disease	8	68
CI15	M	79	5	ACE	Freedom	Freedom contour	Hereditary	28	94
CI16	F	77	9	HDCIS	Opus2	C40+	Unknown	0	46
CI17	M	56	6	ACE	Freedom	Freedom contour	Ménière's disease	0	94
CI18	M	69	5	ACE	Freedom	Freedom contour	Ototoxic drugs	24	76
CI19	F	62	1.5	ACE	CP810	CI512	Hereditary	0	92
CI20	F	42	6	ACE	Freedom	Freedom contour	Unknown	0	84
CI21	F	51	5	ACE	Freedom	Freedom contour	Unknown	10	74

Note: ACE = Advanced Combination Encoder.

Ten reversals were recorded, of which the initial two featured a larger stepsize. These first two reversals were discarded from the analysis. The starting value of the tracking variable was set to the eleventh stimuli from the series of 48 (F0, duration, and F1) and the seventh stimuli from the intensity stimulus set. This corresponded to an initial difference relative to the standard stimulus of 18 Hz for F0, 55 msec for duration, 6 dB for intensity, and 70 Hz for F1. The standard stimuli were uniformly lower in pitch for F0, shorter for duration, lower for intensity, and lower in F1 value for vowel quality than the deviating stimuli.

Minimal Pair Testing

The word stress stimuli comprised nine minimal pairs of 18 items; the vowel length stimuli were 17 pairs of 34 items; and the compound word/phrase stimuli were 41 pairs of 82 items (see Appendix). Examples of a pair of word stress stimulus in noise (Audio 1–2), a pair of vowel length stimulus in noise (Audio 3–4) and a pair of compound word/phrase stimulus in quiet (Audio 5–6) are available supplemental to the online version of this article. Two pairs from the vowel length set were drawn from Hadding-Koch and Abramson (1964) and three pairs from the compound word/phrase stimuli were drawn from the Web site www.skrivihop.nu (accessed June 30, 2011). All remaining stimuli from the minimal pair tasks were chosen on the basis of their phonetic characteristics by the first author and were reviewed by a panel of five native Swedish speakers. A secondary, but nonetheless critical, criterion was that the words in the pair were spelled differently so that participants could distinguish between the orthographically presented response alternatives. The compound words and phrase stimuli consisted only of semantically possible phrases, although some phrases were highly unlikely but not syntactically impossible.

A 40-yr-old male nonprofessional speaker recorded the stimuli. This speaker had Swedish as his mother tongue and grew up in the vicinity where testing took place. He can be considered as having a representative dialect with which all participants would be very familiar. The speaker had a mean F0 of 113 Hz and spoke slowly. The mean speech rates (syllables per second) were 3.9 for the word stress stimuli, 4.8 for the compound word/phrase stimuli, and 8.5 for the vowel length stimuli. The speech rate of the vowel length stimuli seems fast, but this is due to the inclusion of the short vowel items and the fact that all but three of the pairs are monosyllables while the word stress and compound word/phrase sets are comprised of polysyllabic items.

Recordings of all stimuli were made in a sound-treated environment with a microphone (Sennheiser MKH40 P48) and a 32-bit digital audio recorder (Sound Devices 722). Stimuli were recorded at a sampling rate of 32 kHz. The recorded stimuli were edited so that all stimuli were paired with a carrier phrase, "Ordet är ___" ["The word is ___"]. The root mean square (RMS) level of each stimulus item and the carrier phrase were adjusted to a uniform value (either 70 or 75 dB). For the noise conditions, an unmodulated speech-spectrum shaped random noise from the International Collegium of Rehabilitative Audiology (ICRA) noise collection was used. The selected noise had a spectrum based on the speech spectrum of a male speaker speaking with a normal vocal effort, and this was deemed desirable in order to ensure adequate interference with the recorded material used in our study. This noise was added at an SNR of 10 dB measured relative to the long-term RMS levels. This SNR was considered to be challenging but not impossible for CI listeners based on the mean CI perceptual results for the steady noise reported in Fu and Nogaki (2005) and Nelson et al (2003).

Procedure

Prior to testing, the electrode impedance values of the implants were checked as well as the microphone function of the processor. Stimuli were presented via a single loudspeaker placed at a distance of 1 m from the listener. CI participants used their clinically assigned speech processor in the preferred setting. Participants were instructed to identify the interval that was different and respond on a keyboard. Repetitions of stimuli were not permitted.

The minimal pair tests were performed in the following order: word stress (randomized presentation of stimuli with and without noise), vowel length (noise), compound word/phrase, vowel length, compound word/phrase (noise). The carrier sentence and stimulus were presented to the participants prior to display of the response choices on the screen. Participants selected with the computer mouse the word or phrase they thought they had heard. Phrase response alternatives, in the compound word/phrase task, were displayed with a double space between the constituent words to visually contrast their division relative to the compound word choice. Participants were warned that some of the phrase response alternatives might appear strange but that they should select the alternative that they hear. Prior to the minimal pair tests, participants were asked if they had ever been diagnosed with a reading impairment, and none had. Although no training round was permitted, it was noted that no participant was consistently hesitant in identifying answer choices. The presentation level of the minimal pair stimuli was 70 or 75 dB (A) (long term). For the discrimination tasks, the sound field was calibrated so that all the standard stimuli were at 65 dB (A) before the application of level roving.

An observer was present during testing. With the discrimination task, in instances where the reversal pattern exceeded five consecutive stimuli steps during the last eight reversals, the task was repeated. Together both the discrimination and minimal pair tasks took approximately 1 hr and 15 min to complete.

RESULTS

Discrimination Tasks

Results on the discrimination tasks for both groups are displayed in Figure 1. The mean F0 DLs for the CI and NH groups were 15.14 Hz (CI) and 2.7 Hz (NH); for duration means were 62.47 msec (CI) and 84 msec (NH); for intensity means were 3.69 dB (CI) and 1.28 dB (NH); and for F1 means were 59.04 Hz (CI) and 35 Hz (NH) for F1. A Mann-Whitney *U*-test revealed significantly higher DLs in the CI group for F0 ($W = 8.5$, $p < 0.001$) and intensity ($W = 54$, $p < 0.05$). DLs for F1 did not differ between groups ($W = 68$, $p = 0.12$), while the NH group had a significantly higher DL for duration ($W = 152$, $p = 0.049$).

Minimal Pair Tasks

The individual and mean results for the vowel length, word stress, and compound word/phrase tasks can be seen in Figures 2, 3, and 4. It can be noted that the NH participants showed ceiling levels of performance on all three tasks, and the addition of noise did not markedly affect this. The CI participants performed with varying degrees of success. In general they did quite well on the vowel length task, not so well on the word stress task, and worse still on the compound word and phrase task. All participants except for CI11 scored consistently above chance level (50%). This subject's results were retained in the mean scores (as shown in Figures 2, 3, and 4) but were removed from all subsequent analysis including the regression model due to the fact that this participant spoke Finland Swedish, a dialect that has substantial prosodic differences from standard Swedish. It can also be

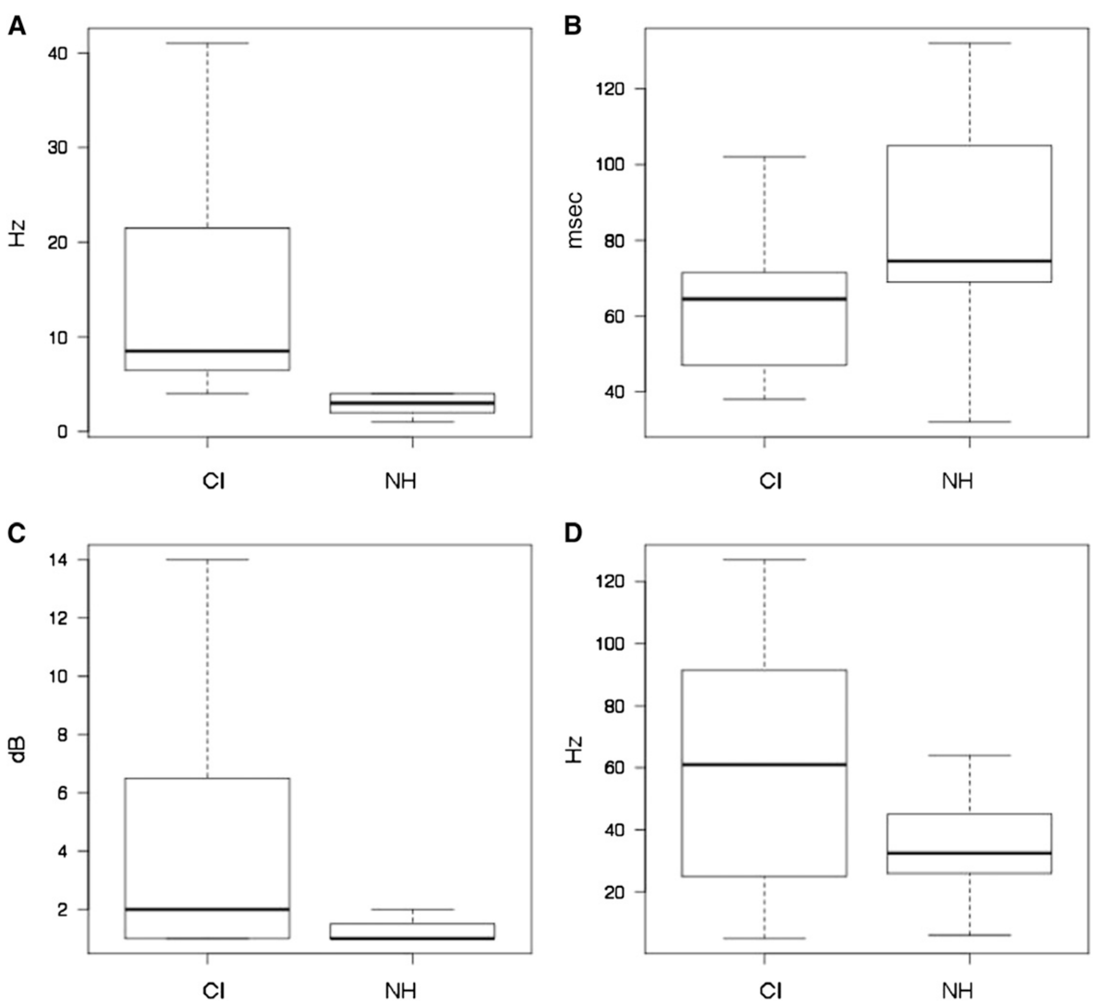

Figure 1. Boxplots of the AFC results for *(A)* F0, *(B)* duration, *(C)* intensity, and *(D)* F1. The line in the center of the box is the median value. The box shows the interquartile range, and the whiskers are the highest and lowest values that are not outliers or extreme values.

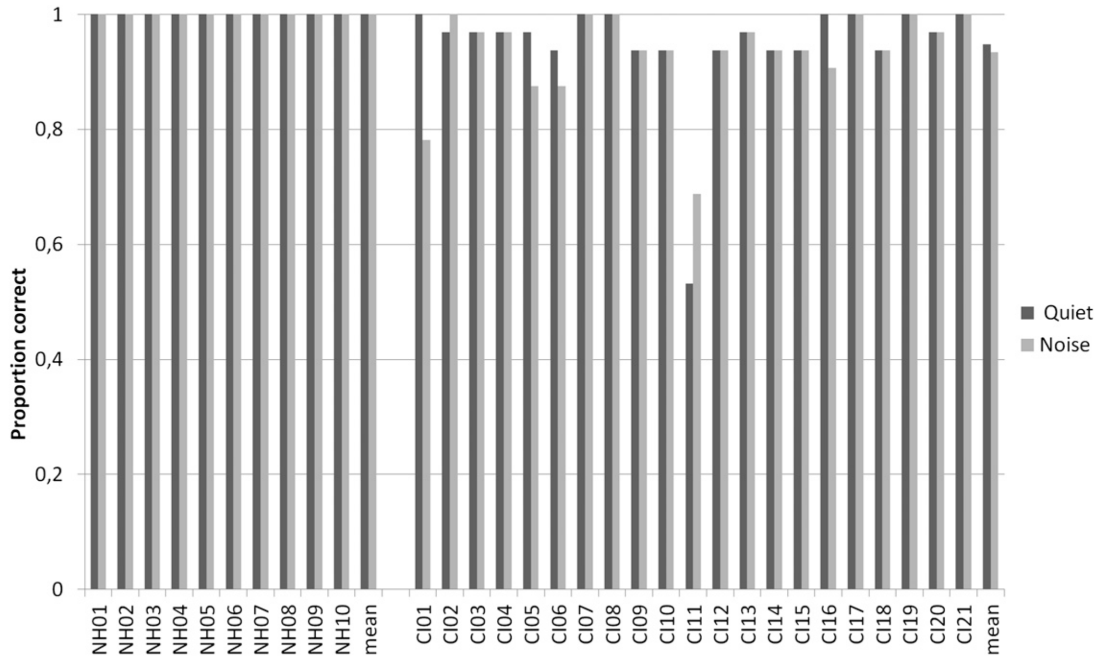

Figure 2. Proportion correct identification of vowel length in quiet and in noise. Individual data and mean scores are shown for the NH and the CI participants. Chance level is 50%.

noted that participants CI07 and CI08 achieved ≥95% in the quiet condition of all three tests.

Mixed Effects Regression Modeling

Correct and incorrect responses from the vowel length, word stress, and compound word/phrase tests from both CI and NH groups were used as the dependent variable in a generalized linear mixed regression model. This form of regression examines the extent to which a dependent variable is related to a combination of predictors via a link function, for instance, the logit function. A mixed effect regression model is appropriate when the data includes factors with repeatable levels,

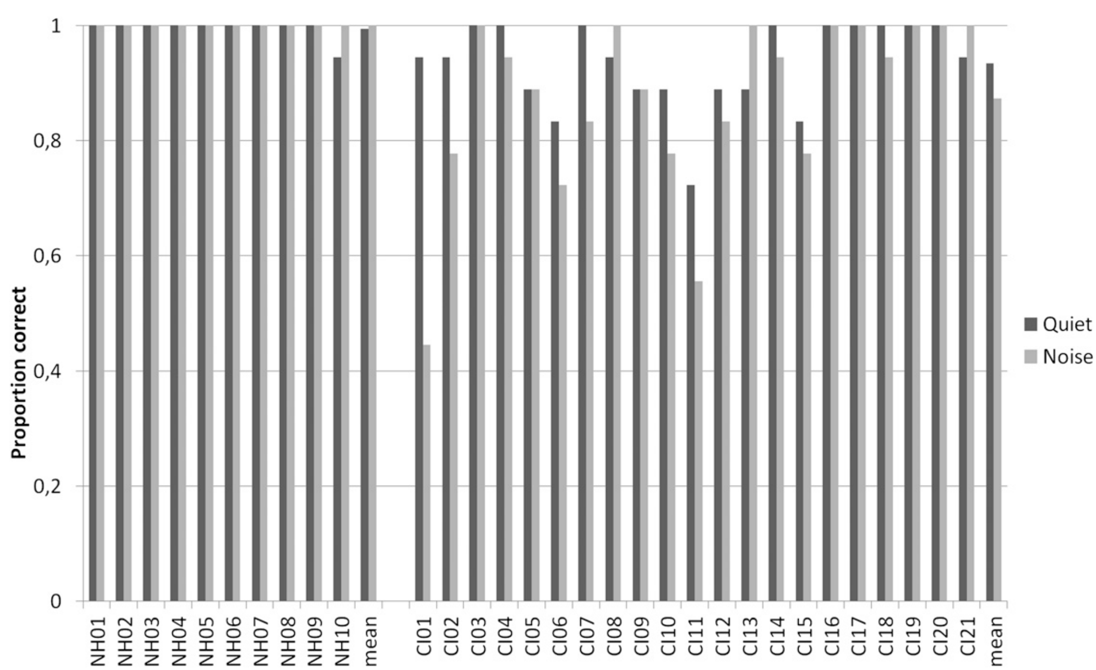

Figure 3. Proportion correct identification of word stress in quiet and in noise. Individual data and mean scores are shown for the NH and the CI participants. Chance level is 50%.

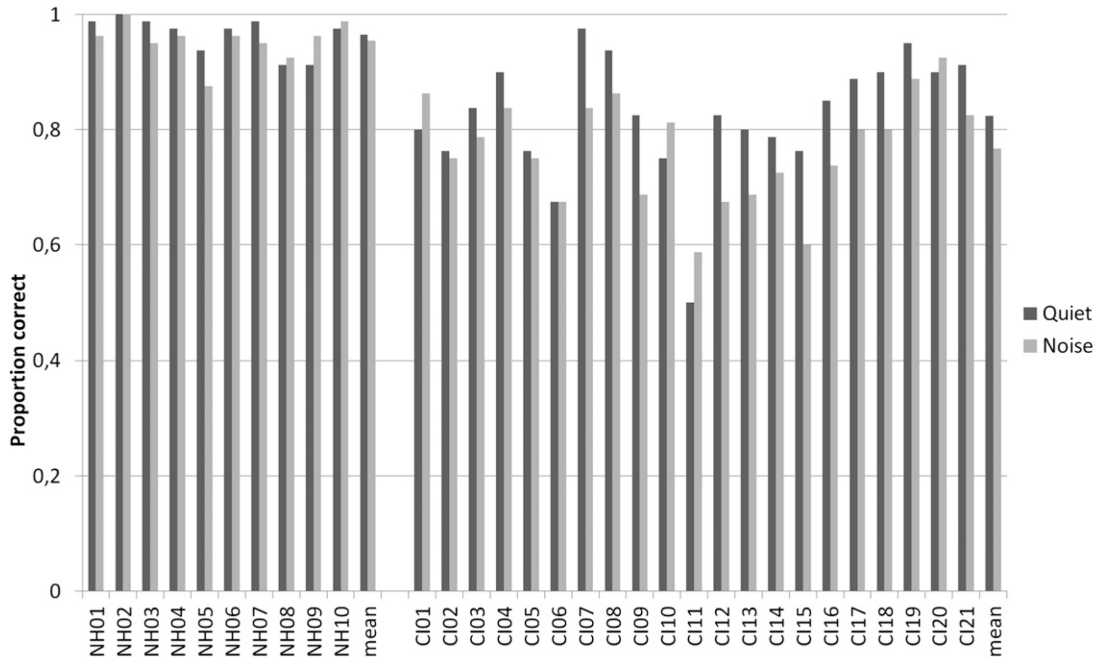

Figure 4. Proportion correct identification of compound words and phrases in quiet and noise. Individual data and mean scores are shown for the NH and the CI participants. Chance level is 50%.

for which fixed-effect terms are used, and also factors with levels that are randomly sampled, for which random-effect terms are used. The slope and intercept of random effects are adjusted for each level of that factor; for instance, if participant is included as a random effect, a mixed effect regression model will adjust the slope and intercept for each participant allowing us to take sampling variability into account and make formal inferences about the population from which the participant was drawn.

The initial model selected from the CI and NH group data included age, noise condition, DL for F0, DL for duration, DL for intensity, and DL for F1 as fixed factors, and participant as a random factor. Another model was created with the same fixed factors but this time including stimulus item as a random factor. An ANOVA performed on both models justified the inclusion of stimulus item as a random factor. The regression model revealed a significant difference between the CI and NH groups, $z(19) = 8.462$, $p < 0.001$. This showed that CI participants did not perform as well as the NH participants on the minimal pair tasks.

Since it was anticipated that analysis of the CI and NH groups would reveal different patterns of associative links, between the minimal pair scores, discrimination measurements, and individual factors, the results from the CI group alone were subjected to a separate logistic regression analysis. Stimulus item and participant were again included as random effects. Interactions between predictors were retained in the final model if they improved the fit of the model. This was determined by the value of the Akaike information criteria. Specifics of the final model are given in Table 2.

Age and noise condition emerged as significant predictors of CI user performance, while the interaction between these parameters was not significant and did not improve the fit of the model. The DL for intensity was found to be a significant predictor, and the interaction between the DL for intensity and the DL for F1 improved the fit of the model. The effect of years of implant experience was found to be close to significance, $z(10) = 1.807$, $p = 0.07$.

Table 2. Summary of Fixed Effects in the Mixed Logit Model (N = 5360 observations; Akaike information criterion = 3598; log-likelihood = −1753) with Noise Condition (reference quiet), Age, Implant Experience, DL for F0, DL for Duration, DL for Intensity, DL for F1, and Interaction between the DLs for Intensity and F1

Predictor	Coeff.	SE	Z	p
Intercept	4.175	0.700	5.965	<0.001
Noise	0.636	0.178	3.559	<0.001
Age	−0.021	0.008	−2.702	<0.05
Implant experience	−0.008	0.004	1.806	ns
DL F0	−0.003	0.012	−0.267	ns
DL duration	−0.003	0.006	−0.506	ns
DL intensity	−0.132	0.067	−1.981	<0.05
DL F1	−0.006	0.005	−1.092	ns
DL intensity x DL F1	0.001	0.001	1.407	ns

DISCUSSION

Data from this study shows that CI participants could not identify the brief prosodic cues that distinguish words in vowel length, word stress, or compound word/phrase minimal pairs, as well as NH participants. Mean test performance for the CI group was best on the vowel length task and worst on the compound word/phrase task. The performance of CI participants on all minimal pair tasks was significantly negatively affected by the introduction of noise. This noise condition had little or no effect on the performance of the NH participants, and the scores from this group were close to ceiling. The age of the CI participants and their DL for intensity were observed to be predictors of their overall performance in both quiet and noise conditions.

The mean deterioration on the minimal pair tasks seen in the CI group when noise was added to the stimuli was 1.6% (quiet, 95%; noise, 93.4%) for the vowel length task, 5.8% (quiet, 93.4%; noise, 87.6%) for the word stress task, and 5.6% (quiet, 82.2%; noise, 76.6%) for the compound word/phrase task. In comparison, the deteriorations observed in the NH group were negligible. The differences in mean scores between quiet and noise conditions observed in the CI group should be seen in light of the favorable SNR that was employed. This SNR was found to be virtually innocuous for the NH group. They should also be considered in terms of the 50% chance level dictated by the two-response alternative minimal pair closed set.

The identification of age as a predictor of the ability of CI participants to perceive brief prosodic cues is one noteworthy finding of this study. The direction of the slope and intercept indicate that older CI participants performed worse on the minimal pair tasks than younger participants. Souza et al (2011) found a similar age effect when NH participants attended to vocoded stimuli with different shaped voice pitch contour (rising, falling, and relatively flat) that was nested in a diphthong. The younger group had steeper psychometric functions in comparison to the older group. Hoyte et al (2009) found a similar difference with natural utterances in judgments of syntactical parsing of sentences by two groups of younger and older NH listeners. They posited that on the basis of response time measurements, their results were attributable to the younger listeners' ability to develop situation-specific strategies. When considering the hearing loss and device factors that contribute to poor speech in noise performance among CI listeners, it would seem to be of interest to better understand the location and cause of the age effect. In particular, it would be of interest to ascertain if the age effect observed in the present study stemmed from difficulties with frequency discrimination associated with aging (Schneider and Pichora-Fuller, 2000) or cognitive decline that imposes restriction on resourceful utilization of the relatively meager electrical signal provided by the CI (Heydebrand et al, 2007), or a combination of these and other factors. This is a direction for further research, and it is relevant to the consideration of bespoke processing strategies in CIs.

The attempt made in the present study to compare the profile of the auditory ability of CI listeners, yielded by their discrimination thresholds, to their ability to perceive brief durational prosody revealed that the DL for intensity provided predictive value for performance on the minimal pair tasks. This finding supports those of Rogers et al (2006) who measured DLs with isolated and also concurrent F0 and intensity changes of pseudo-synthesized words featuring a stressed syllable in the middle, for example, *potato*. They found that CI listeners could discriminate smaller changes in F0 and intensity when these were varied simultaneously than when only one cue was present. Meister et al (2011) also reported significant correlation coefficients between sentence stress discrimination in natural utterances and results from stimuli that were modified with isolated and concurrent changes in intensity. They found that the identification of stress did not correlate with the discrimination of duration, indicating that this cue is of little importance to CI listeners in identifying sentence stress. Further research in this area could examine the interplay of pitch, intensity and duration so that they are presented in various combinations to CI listeners.

In preparing the lists of natural utterance minimal pairs used in this study it was necessary to control for a number of criteria including phonology and syntax. One of the critical criteria was that the stimuli pairs had to be heterographic, so that participants could distinguish between answer choices on the computer screen. This placed a considerable restriction on the word stress task where many pairs of candidate stimuli, like *'fasan* (horror) and *fa'san* (pheasant), had to be excluded. The exclusion of such stimuli meant that there was a tendency toward more inclusive phonological criteria in the derivation of these lists. This inclusivity may have been to the benefit of overall performance on the vowel length task. In some of the stimulus pairs the difference was not solely based on the duration of the vowel. There was also a degree of covariation of duration and vowel quality between words in the minimal pair. In some of the stimulus pairs there was negligible vowel quality difference (for instance, *mäta* and *mätta*), while for other stimuli pairs there were distinct vowel quality differences (for instance, *kål* and *koll*). The perception of vowel quality requires an ability to identify an alteration in formant structure. One measured indicator of this ability was the DL for F1. However, the regression model did not identify the DL for F1 as a significant predictor of overall performance.

Another explanation for the relatively good performance seen in the vowel length task is the acoustic prominence of vowels. They are greater in amplitude and are relatively longer than consonants. Also, the mean durational difference between the long and the short vowels (as measured at the offset of the initial consonant and the offset of the middle vowel), was 96 msec. This was considerably longer than the CI participants observed mean DL for duration (62 msec). This indicates that they were probably able to make use of the durational differences within the vowel segment of each pair.

A further reason that might explain the CI listeners' performance for vowel length is that intrasyllabic durational cues in Swedish are relatively symmetrical. This relationship is what Lehiste (1970) termed the "mutual complementation of vocalic and consonantal quantity," which means that a short vowel is followed by a long consonant and a long vowel by a short consonant. These relationships deliver essentially a doubling of durational cues to the listener as the durations of both the vowel and the postvocalic consonant signal the length of the vowel. It is also possible that the high CI performance scores observed on the vowel length test were promoted by Swedish orthography, in which the representation of the postvocalic consonant clearly marks vowel length oppositions with a double letter for the short vowel and a single letter for the long vowel; for instance, *söt* is a long vowel and *sött* is a short vowel.

The duration of deafness and the resultant period of auditory deprivation were considerations that were not addressed in this study. These proved hard to ascertain in this cohort due to the considerable number of participants with progressive onset of hearing loss and unknown etiologies. The variation observed in the CI group may have been related to the different processing strategies used by three participants of the CI participants, but this was not investigated. Another methodological limitation of this study was the omission of a training round prior to the minimal pair task. This was necessitated by the test time that was deemed to be demanding and at the upper limit of what participants would tolerate. Also, the one-up one-down adaptive procedure used in the discrimination task yielded a DL that was above chance level. However, measurement of discrimination thresholds at a convergence point higher on the psychometric function might have provided a better predictor for use in the regression model.

This study highlights the limitations of the everyday listening abilities of CI listeners and the problems they face in perceiving natural utterances. It has implications for health professionals, caregivers, and those who communicate regularly with CI listeners. CI listeners are not as good at identifying the prosodic cues that relate to words, segments, and word boundaries, as their NH counterparts. Clear accentuation of word level features is likely to result in more successful communication. Also, the age effect observed in this study supports the assertion that older CI listeners do not perform as well as younger CI listeners in brief prosodic identification tasks when a noise condition is included that simulates a real-life listening environment. This is relevant when considering CI candidacy and when providing CI candidates with information as a basis for developing realistic expectations of postimplant listening ability.

Acknowledgments. The authors are grateful to the cooperative CI and NH participants who helped in this study and Erik Lindeman, Preben Dømler, Brent Kirkwood, Jens Bo Nielsen, and Sebastien Santurette for their assistance in preparing the stimuli and test scripts. Special thanks also to Laura Winther Balling, Björn Johansson, Agneta Sandh, Hanna Kihlberg, Jenny Iwarsson, Maria Einer, Nicolai Pharao, John Tøndering, and Pontus Lindeman for their advice and helpful comments.

REFERENCES

Alkaf FM, Firszt JB. (2007) Speech recognition in quiet and noise in borderline cochlear implant candidates. *J Am Acad Audiol* 18(10):872–882.

Boothroyd A. (1984) Auditory perception of speech contrasts by subjects with sensorineural hearing loss. *J Speech Hear Res* 27(1):134–144.

Bruce G. (1998) Vol. 16 of *Allmän och svensk prosodi* [General and Swedish prosody]. Lund: Reprocentralen.

Friesen LM, Shannon RV, Baskent D, Wang X. (2001) Speech recognition in noise as a function of the number of spectral channels: comparison of acoustic hearing and cochlear implants. *J Acoust Soc Am* 110(2):1150–1163.

Fu QJ, Chinchilla S, Nogaki G, Galvin JJ, 3rd. (2005) Voice gender identification by cochlear implant users: the role of spectral and temporal resolution. *J Acoust Soc Am* 118(3, Pt. 1):1711–1718.

Fu QJ, Nogaki G. (2005) Noise susceptibility of cochlear implant users: the role of spectral resolution and smearing. *J Assoc Res Otolaryngol* 6(1):19–27.

Fu QJ, Shannon RV, Wang X. (1998) Effects of noise and spectral resolution on vowel and consonant recognition: acoustic and electric hearing. *J Acoust Soc Am* 104(6):3586–3596.

Green T, Faulkner A, Rosen S, Macherey O. (2005) Enhancement of temporal periodicity cues in cochlear implants: effects on prosodic perception and vowel identification. *J Acoust Soc Am* 118(1): 375–385.

Hadding-Koch K, Abramson AS. (1964) Duration versus spectrum in Swedish vowels: some perceptual experiments. *Stud Linguist* 18:94–107.

Heydebrand G, Hale S, Potts L, Gotter B, Skinner M. (2007) Cognitive predictors of improvements in adults' spoken word recognition six months after cochlear implant activation. *Audiol Neurootol* 12(4):254–264.

Hochberg I, Boothroyd A, Weiss M, Hellman S. (1992) Effects of noise and noise suppression on speech perception by cochlear implant users. *Ear Hear* 13(4):263–271.

Hoyte KJ, Brownell H, Wingfield A. (2009) Components of speech prosody and their use in detection of syntactic structure by older adults. *Exp Aging Res* 35(1):129–151.

Kovačić D, Balaban E. (2010) Hearing history influences voice gender perceptual performance in cochlear implant users. *Ear Hear* 31(6):806–814.

Lehiste I. (1970) *Suprasegmentals*. Cambridge, MA: MIT Press.

Magnusson L. (1995) Reliable clinical determination of speech recognition scores using Swedish PB words in speech-weighted noise. *Scand Audiol* 24(4):217–223.

McKay CM, McDermott HJ. (1993) Perceptual performance of subjects with cochlear implants using the Spectral Maxima Sound Processor (SMSP) and the Mini Speech Processor (MSP). *Ear Hear* 14(5):350–367.

Meister H, Landwehr M, Pyschny V, Wagner P, Walger M. (2011) The perception of sentence stress in cochlear implant recipients. *Ear Hear* 32(4):459–467.

Meister H, Landwehr M, Pyschny V, Walger M, von Wedel H. (2009) The perception of prosody and speaker gender in normal-hearing listeners and cochlear implant recipients. *Int J Audiol* 48(1):38–48.

Meister H, Pyschny V, Landwehr M, Wagner P, Walger M, von Wedel H. (2008) Konzeption und Realisierung einer Prosodie-Testbatterie [Conception and realisation of a prosody test battery]. *HNO* 56(3):340–348.

Most T, Harel T, Shpak T, Luntz M. (2011) Perception of suprasegmental speech features via bimodal stimulation: cochlear implant on one ear and hearing aid on the other. *J Speech Lang Hear Res* 54(2):668–678.

Nelson PB, Jin SH, Carney AE, Nelson DA. (2003) Understanding speech in modulated interference: cochlear implant users and normal-hearing listeners. *J Acoust Soc Am* 113(2):961–968.

Peng S-C, Tomblin JB, Turner CW. (2008) Production and perception of speech intonation in pediatric cochlear implant recipients and individuals with normal hearing. *Ear Hear* 29(3):336–351.

Plant G. (1994) *Analytika: Analytic Testing and Training Lists*. Somerville, MA: Audiological Engineering.

Rogers CF, Healy EW, Montgomery AA. (2006) Sensitivity to isolated and concurrent intensity and fundamental frequency increments by cochlear implant users under natural listening conditions. *J Acoust Soc Am* 119(4):2276–2287.

Schneider BA, Pichora-Fuller MK. (2000) Implications of perceptual deterioration for cognitive aging research. In: Salthouse TA, Craik FIM, eds. *Handbook of Aging and Cognition* 2nd ed. 2nd ed. Mahwah, NJ: Erlbaum 155–220.

Schvartz KC, Chatterjee M. (2012) Gender identification in younger and older adults: use of spectral and temporal cues in noise-vocoded speech. *Ear Hear* 33(3):411–420.

Shannon RV, Cruz RJ, Galvin JJ, 3rd. (2011) Effect of stimulation rate on cochlear implant users' phoneme, word and sentence recognition in quiet and in noise. *Audiol Neurootol* 16(2):113–123.

Smith MR, Cutler A, Butterfield S, Nimmo-Smith I. (1989) The perception of rhythm and word boundaries in noise-masked speech. *J Speech Hear Res* 32(4):912–920.

Souza P, Arehart K, Miller CW, Muralimanohar RK. (2011) Effects of age on F0 discrimination and intonation perception in simulated electric and electroacoustic hearing. *Ear Hear* 32(1): 75–83.

Svensson SG. (1974) *Prosody and Grammar in Speech Perception.* Monographs from the Institute of Linguistics, University of Stockholm, 2.

Tobey EA, Geers AE. (2011) Factors influencing speech intelligibility in adolescent users of cochlear implants. *Int J Pediatr Otorhinolaryngol* 75(Suppl. 1):51.

Zeng FG. (2002) Temporal pitch in electric hearing. *Hear Res* 174(1–2):101–106.

Appendix

Vowel Length		Word Stress		Compound Word	Phrase
Long	Short	Initial or medial	Medial or final		
ful [fʉ:l]	full [fɵl]	annonser [ɑ'nonsər]	annonsör [ɑnon'sœ:r]	sabbatsår	Sabbats sår
väg [vɛ:g]	vägg [vɛg]	Oliver ['olɪvər]	oliver [o'li:vər]	envis	en vis
mål [mɔ:l]	moll [mɔl]	korset ['kɔrʂet]	korsett [kɔr'ʂɛt]	entita	en tita
bus [bʉ:s]	buss [bɵs]	formel ['fɔr:mel]	formell [fɔr'mel:]	stödlinje	störd linje
hat [hɑ:t]	hatt [hat]	armen ['ar:men]	armén [ar'me:n]	syrgas	sur gas
kål [ko:l]	koll [kol]	kornet ['ku:ɽnet]	kornett [ku:ɽ'net]	vitmossa	vit mossa
glas [glɑ:s]	glass [glas]	kaffe ['kafə]	kafé [ka'fe:]	morgonrockar	morgon rockar
mat [mɑ:t]	matt [mat]	finnes ['fɪnes]	finess [fɪ'nes]	engelska	ängel ska
vit [vi:t]	vitt [vɪt]	kallas ['kalas]	kalas [ka'la:s]	uppmuntra	upp muntra
hut [hʉ:t]	hutt [hɵt]			reklamera	reklam mera
våt [vo:t]	vått [vot]			Andersson	Anders' son
tät [tɛ:t]	tätt [tɛt]			genombrott	genom brott
söt [sø:t]	sött [søt]			räknebok	räkna bok
stöt [stø:t]	stött [støt]			ungkarl	ung Karl
mäta [mɛ:ta]	mätta [mɛta]			fredstid	Fred's tid
skuta [sgʉ:ta]	skutta [sgʉta]			sjukgymnast	sjuk gymnast
pruta [prʉ:ta]	prutta [prɵta]			mittfält	mitt fält
				dåtid	då tid
				kycklinglever	kyckling lever
				snorkråka	snor kråka
				förrätt	för rätt
				rökfritt	rök fritt
				tekniker	teknik är
				felmeddelande	fel meddelande
				elementär	element är
				renkorv	ren korv
				majstång	maj stång
				kassamedarbetare	kassa medarbetare
				tennislag	tennis lag
				antikälskare	antik älskare
				sakförare	sak förare
				utanför	utan för
				slänggungan	släng gungan
				segertåg	seger tog
				gentemot	jänta mot
				badaren	bad Daren
				godsak	god sak
				efterhand	efter hand
				tältstång	tälts tång
				danskorna	dansskorna