Prosody perception by postlingually-deafened cochlear implant recipients: a cross-language investigation.

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

Published in:
Proceedings of the International Symposium on Auditory and Audiological Research. August 2013, Nyborg, Denmark

Publication date:
2013

Document Version
Early version, also known as pre-print

Citation for published version (APA):
Morris, D. J., Faulkner, A., Juul, H., Magnusson, L., & Jönsson, R. (2013). Prosody perception by postlingually-deafened cochlear implant recipients: a cross-language investigation. In Proceedings of the International Symposium on Auditory and Audiological Research. August 2013, Nyborg, Denmark

Prosody perception by postlingually-deafened cochlear implant recipients: a cross-language investigation

DAVID MORRIS[1], ANDREW FAULKNER[2], HOLGER JUUL[1], LENNART MAGNUSSON[3], AND RADOSLAVA JÖNSSON[3]

[1] *Department of Scandinavian Studies and Linguistics-Speech Pathology and Audiology, University of Copenhagen, Denmark*

[2] *Department of Speech, Hearing and Phonetic Sciences, University College London, London, United Kingdom*

[3] *Department of Audiology, Sahlgrenska University Hospital, Gothenburg, Sweden*

Due to the inherent device limitations of cochlear implants (CI) and of auditory perception via an electrical-neural interface, the ability of CI listeners to perceive prosody is often reported as being worse than that of normal-hearing listeners. We tested the perceptual ability of postlingually-deafened adult CI listeners with stimuli where prosodic features signalled distinctive semantic contrasts. These contrasts were tested with a Danish (n=18) and a Swedish (n=21) cohort in quiet and in noise. We also tested other speech perceptual abilities that could be linked to prosody perception. These included word recognition, sentence perception in noise, and vowel identification. Results of this study show that speech-in-noise ability by CI listeners is related to abilities that underlie vowel identification, while word recognition is related to the identification of compound words and phrases. Comparison of the mean identification rates of the prosodic tasks showed that there was a disparity in the performance of Danish and Swedish CI listeners over tasks that are similar in both languages.

INTRODUCTION

Prosody is an integral part of spoken communication that imbues speech with dynamic variation. This variation is perceived as a result of variations in the acoustic cues of pitch, intensity, and rhythm during the course of an utterance. Cochlear implant (CI) listeners can have difficulty in perceiving changes in these cues. The purpose of the present investigation was to examine the prosodic ability of CI listeners where prosody provided a distinctive semantic contrast, the perception of which is necessary for accurate identification of natural utterances in both Swedish and Danish. In testing the Swedish participants we included a quiet and a noise condition in order to involve a situation relevant to everyday listening.

Noise generally has a deleterious effect on most measures of speech intelligibility by CI listeners. This is due to device factors 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, improvements in speech intelligibility in noise plateau when the number of channels is 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 a stimulating 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 (Zeng, 2002).

To examine the extent to which naturally-occurring prosodic detail is available to CI listeners we compared their performance to normal-hearing (NH) listeners in tasks which required the identification of a word or phrase from two known alternatives. Together these two answer choices constitute a prosodically contrastive minimal pair, that is, the two words are segmentally identical, or close to identical, but their prosodic characteristics differ. The prosodic features that were investigated were word stress, vowel length, stød, and compounds and phrases.

Vowel duration and word stress are critical features in many Scandinavian languages as they can denote semantic distinctions between words. For instance, in Swedish, ‘väg’ /vɛːg/, with a long vowel means ‘road’ and ‘vägg’ /vɛg/, with a short vowel, means ‘wall’. An example of word stress arises in the English word ‘convict’ which can be either a noun or a verb according to the position of the stress: for example, ‘Because of crimes committed in prison, the judge will ˌconˈvict (verb) the ˈconˌvict (noun)’. Stød is a prosodic feature that is peculiar to the Danish language. It is a syllabic feature characterized by irregular laryngeal vibrations that affect a long vowel or consonant (Grønnum, 2001). In the Swedish study, we also tested the ability of CI and NH listeners to identify compound words as opposed to phrases. 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.’

Our interest in examining naturally uttered minimal pairs was also to consider performance on these tasks with testing that is carried out as a part of clinical routine. Cullington and Zeng (2011) reported that neither performance on the HINT sentence test in quiet nor in babble noise correlated with prosodic scores from tasks where adult CI listeners were required to identify affective emotion. In contrast, Rogers *et al.* (2006) reported that there was a link between performance on prosodic discrimination and sentence perception in quiet. They found that the ability of adult CI listeners to discriminate changes in prosodic stress signalled by concurrent F0 and intensity modification correlated with speech perception as measured with the CNC words. To further investigate the speech perceptual ability of CI listeners as measured with standard clinical tests in relation to their ability to perceive prosody, we used a series of minimal-pair identification tasks where prosodic features of words provided a distinctive semantic contrast.

METHOD

Participants

Swedish 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 years (range 40-82). All participants had been unilaterally implanted within the last nine years and, with the exception of two participants, all had word recognition scores in quiet above 50% at their most recent post-implant testing. The processing strategies used by the Swedish listeners included ACE (18), FSP (2), and HDCIS (1).

The Danish CI participants (n=18) were recruited by advertisements displayed online and in social media. There were no selection criteria based on prior knowledge of hearing ability. These participants had a mean age of 53.3 years (range 41-70). Eleven of the CI participants had been bilaterally implanted. However, all were tested with only one implant. The Danish participants used ACE (12), HiRes fidelity 120 (3), SPEAK (1), MP3000 (1), and CIS (1).

The Swedish CI participants had a mean of 4.8 yrs of experience listening via their CI while the Danish CI participants had 4.7 yrs. No participant suffered from a cognitive handicap or reading impairment and all were native speakers of their respective language. All participants were reliant on oral communication in their everyday lives. In instances where a participant wore a contralateral hearing aid, this was removed prior to testing. In cases where a participant used bilateral implants, they were instructed to select the side from which they believed they derived the most benefit. An earplug was used to block the contralateral ear of participants that reported residual hearing in that ear.

The Swedish control group (n=10) consisted of NH adults with a mean age of 51 years (range 36-70). The Danish control group consisted of NH (n=16) participants with a mean age of 30.4 years (range 23-61).

Minimal-pair testing

The minimal-pair tasks were chosen on the basis of their phonetic characteristics by the first author. A panel of native speakers reviewed the Swedish items and two expert phoneticians reviewed the Danish stimuli. The words in each pair were spelt differently so that participants could distinguish between the orthographically presented response alternatives. Table 1 provides an overview of the test stimuli in both languages.

The vowel identification stimuli from the Danish study consisted of both short and long vowels, which were tested separately. Short- and long-vowel stimuli were presented in factorial combination such that each stimulus item was presented with all other items in that series as response alternatives. For example, the word ‘mit’ was presented as a response alternative for the stimuli ‘mit,’ ‘midt,’ ‘mæt’, and ‘mat’.

Language	*Vowel length*	*Word stress*	*Compound word/phrases*	*Stød*	*Vowel quality - short*	*Vowel quality - long*
Swedish	väg [vɛːg] vägg [vɛg]* (34)	korset [ˈkɔrɕɛt] korsett [kɔrˈɕɛt]* (18)	rökfritt rök fritt* (82)	-	-	-
Danish	læser [lɛːsʌ] læsser [lɛsə] (44)	August [ˈɑwgɔsd] august [ɑwˈgɔsd] (14)	-	Brugsen [bʁuˀsən] brusen [bʁuːsən] (22)	mit [mɪd] midt [med] mæt [mɛd] mat [mad] (12)	mile [miːlə] mele [meːlə] mæle [mɛːlə] male [mæːlə] (12)

Table 1: Examples of the minimal-pair stimuli used in these experiments from both languages. Values in parentheses are the number of stimulus items per test. * indicates contrasts that were tested in quiet and in a noise condition where ICRA unmodulated speech-shaped noise was combined with the stimuli at an SNR of 10 dB.

Two male adult non-professional speakers recorded the stimuli. Both Swedish and Danish speakers could be considered as having representative dialects with which all participants would be very familiar. Recordings of all stimuli were made in a sound-treated environment with a quality microphone. The individual members of each minimal pair were edited so that they were preceded by the carrier phrase 'Ordet är ___' (Swedish) or 'Ordet er ___' (Danish) [The word is ___]. The root mean square (RMS) level of each stimulus item and the carrier phrase were adjusted to a uniform value. To create the noise conditions used in the Swedish study, an unmodulated speech-spectrum-shaped random noise from the International Collegium of Rehabilitative Audiology (ICRA) collection was used. This noise was added at an SNR of 10 dB. 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

Stimuli were presented via a single loudspeaker placed at a distance of one meter 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. Although no training round was provided, 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).

Speech in noise and word identification testing

During the testing session the speech reception threshold (SRT) in noise of the Danish CI listeners was measured with the Danish HINT (Nielsen and Dau, 2011). The speech material in this test consists of five-word sentences read by male speakers presented in a speech-shaped noise with minor amplitude modulation. Participants were required to repeat the sentence that was presented and scoring was by words correct.

The Swedish PB word identification test was measured during a routine follow-up appointment with the Swedish CI listeners. For a 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 one meter in front of the subject in a sound treated room.

RESULTS

Minimal pair tasks

The individual and mean results for the minimal-pair identification tasks performed by the Swedish group can be seen in Table 2 (upper panel). The lower panel shows the mean results for the minimal-pair identification tasks performed by the Danish group. It can be noted that in both languages the NH participants showed ceiling levels of performance on all tasks, and the addition of noise in the Swedish study did not markedly affect this.

	vowel length quiet	*vowel length noise*	*word stress quiet*	*word stress noise*	*compounds/ phrases quiet*	*compounds/ phrases noise*
CI	0.95	0.93	0.93	0.87	0.82	0.77
NH	1	1	0.99	1	0.97	0.95

	vowel length	*word stress*	*stød*	*vowel quality - long*	*vowel quality - short*
CI	0.83	0.74	0.77	0.86	0.81
NH	1	0.97	0.96	0.99	0.98

Table 2: Upper panel shows mean proportions correct on the Swedish minimal-pair tasks and lower panel shows the Danish results.

Other measures of speech perceptual ability

The mean results from the Danish listeners in the vowel identification tasks are provided in Table 2 (lower panel). In both the long and short vowel sets, the open and closed vowels at the extremities of the production plane from which the stimuli were drawn were often successfully identified. The mean proportions of correct identifications for these stimuli were 0.93 for /iː/, 0.87 for /i/, 0.96 for /æː/, and 0.97 for /æ/. In contrast, mean identification rates for vowels from the middle of the frontal plane were poorer. For /eː/ they were 0.7, for /ɛː/ 0.84, and for /e/ 0.66.

Group mean results from the Swedish CI listeners on the PB word test were 73.3% (S.D. = 15). The mean Danish HINT results from the CI group were 17.8 dB SNR (S.D. = 9).

Correlations between minimal-pair results and standard measures

In comparing the results from the Swedish CI listeners to their most recent performance on the PB word test a correlation was observed between the compound word and phrase task in quiet ($r = 0.58$, $p < 0.01$) and in noise ($r = 0.64$, $p < 0.01$). Correlation coefficients for the word stress results in noise were high but were not significant ($r = 0.39$, $p = 0.09$). Neither the word stress in noise nor the vowel-length results in quiet nor in the noise condition were significantly correlated with the PB word-recognition scores. Significant negative correlations were found between the SRT from the Danish HINT and the individual mean scores on the tests of vowel length ($r = -0.78$, $p = 0.001$) and long vowel identification ($r = -0.73$, $p = 0.001$). The negative correlations indicate that participants with lower SRTs performed better on the identification tests. The word-stress identification ($r = -0.59$, $p = 0.01$) and short-vowel identification ($r = -0.51$, $p = 0.03$) were also found to correlate with the Danish HINT results, but these were not significant after Bonferroni-adjusted corrections.

DISCUSSION

This study shows that CI participants could not identify naturally-uttered prosodic cues that distinguish words as well as NH participants in both of the Scandinavian languages that were tested. The performance of Swedish CI participants on all minimal-pair tasks was significantly negatively affected by the introduction of noise. This noise condition had little effect on the performance of the NH participants, and the scores from this group were close to ceiling. The results from the identification tasks from the Danish and the Swedish listeners are of interest as they can provide a cross-language comparison of prosodic features that are similar. Although minimal-pair test items in both languages were not cognate words, the Swedish CI listeners generally performed better than the Danish CI listeners.

It is possible that the selection of the Swedish CI participants according to results on the PB words test contributed to the performance differences that were noted. This meant that only Swedish CI listeners with word-recognition scores above approximately 50% were admitted to the study, whereas the admittance of Danish participants to the study was not based on any speech-performance measures. Nonetheless, due to the magnitude of the performance differences between the CI listeners from both languages (0.12 for vowel length and 0.19 for word-stress identification), other factors may have been involved. One such factor that may explain the poorer performance observed in the Danish cohort is the lenition process to which the Danish language has been subject. The consonant reduction and schwa assimilation and deletion (Bleses *et al.*, 2008) that has been a consequence of this may have made the minimal-pair identification tasks harder for Danish CI listeners as segmental reductions would diminish the distinctive cues on which identification was based. A factor that may have contributed to the better performance observed in the Swedish cohort on the word-stress identification task is that the Swedish language exhibits prominent stress patterns that can be transcribed by listeners even when the language is hummed (Svensson, 1974; Bruce, 1998).

Another reason that may have benefited the Swedish CI listeners' performance on both vowel-length identification and word stress 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 higher identification score observed for the Swedish CI listeners on the vowel-duration test were promoted by Swedish orthography, in which the representation of the postvocalic consonant clearly marks vowel durational 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. However, neither of these explanations accounts for the difference in word-stress identification scores observed in the Swedish and Danish CI participants.

The correlation that was observed in the Swedish cohort between PB word results and compound word and phrase identification scores indicates that there is a link between the ability to identify monosyllables and the accurate placement of word boundaries. This correlation was found to be stronger in the noise condition than in quiet, suggesting that those listeners that are adept at placing word boundaries when the stimuli are marred are also better at word recognition in quiet. It has been suggested that sensitivity to the stress patterns of a language assists in the division of the speech stream during language acquisition (Jusczyk *et al.*, 1999). One of the perceptual abilities that are believed to underlie the placement of word boundaries is differentiation of stress patterns that are contained within words (word stress) and those which are contained within phrases (phrasal stress). The correlation that we found, albeit weak, between PB word scores and word-stress identification in noise supports this notion. It would be of interest to investigate how postlingually-deafened adult CI listeners use their knowledge of other language-specific characteristics, for instance, phonotactic properties and distributional characteristics, to perform speech segmentation.

This study highlights the limitations of the everyday listening abilities of CI listeners and the problems that they face in perceiving prosody as it occurs in natural utterances. The correlation results suggest that there is a link between sentence perception in noise and abilities that underlie vowel identification by CI listeners. Also, word recognition in quiet appears to be linked to the ability to assign word boundaries accurately. Furthermore, these findings indicate that results from some common speech audiometric routines can yield information about the prosodic perceptual abilities of CI recipients.

REFERENCES

Bleses, D., Vach, W., Slott, M., Wehberg, S., Thomsen, P., Madsen, T.O., and Basboll, H. (**2008**). "Early vocabulary development in Danish and other languages: a CDI-based comparison," J. Child. Lang., **35**, 619-650.

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

Cullington, H., and Zeng, F.-G. (**2011**). "Comparison of bimodal and bilateral cochlear implant users on speech recognition with competing talker, music perception, affective prosody discrimination and talker identification," Ear Hearing, **32**, 16-30.

Friesen, L.M., Shannon, R.V., Baskent D., and 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**, 1150-1163.

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

Grønnum, N. (**2001**). "Fonetik og fonologi – almen og Dansk" [Phonetics and phonology – general and Danish], Akademisk forlag.

Jusczyk, P.W., Houston, D.M., and Newsome, M. (**1999**). "The beginnings of word segmentation in English-learning infants," Cogn. Psychol., **39**, 159-207.

Lehiste, I. (**1970**). *Suprasegmentals* (Cambridge, Mass.: M.I.T. Press).

Magnusson, L. (**1995**). "Reliable clinical determination of speech recognition scores using swedish PB words in speech-weighted noise," Scand. Audiol., **24**, 217-223.

Nelson, P.B., Jin, S.H., Carney, A.E., and Nelson, D.A. (**2003**). "Understanding speech in modulated interference: cochlear implant users and normal-hearing listeners," J. Acoust. Soc. Am., **113**, 961-968.

Nielsen, J.B., and Dau, T. (**2011**). "The Danish hearing in noise test," Int. J. Audiol., **50**, 202-208.

Rogers, C.F., Healy, E.W., and Montgomery, A.A. (**2006**). "Sensitivity to isolated and concurrent intensity and fundamental frequency increments by cochlear implant users under natural listening conditions," J. Acoust. Soc. Am., **119**, 2276-2287.

Svensson, S.G. (**1974**). "Prosody and grammar in speech perception," Monographs from the Institute of Linguistics, **2**.

Zeng, F.G. (**2002**). "Temporal pitch in electric hearing," Hear. Res., **174**, 101-106.