Perception and interpretation of low-onset rising tunes by prelingually deaf cochlear implant users

Colleen Holt[1] & Janet Fletcher[2]

[1]Audiology and Speech Pathology, The University of Melbourne, Australia
[2]Linguistics and Applied Linguistics, The University of Melbourne, Australia
cholt@unimelb.edu.au; jfletcher@unimelb.edu.au

ABSTRACT

Production studies have shown that the intonational tunes L*H-H% and L*L-H% may be used for different discourse functions in Australian English [1,2,3,4]. This study investigates whether prelingually deaf cochlear implant (CI) users can discriminate between changes in the starting point of a rise, as distinguishes these tunes phonetically. We also assessed their interpretation of these tunes.

10 CI users and 19 NH listeners completed an AXB task where the alignment of the rise startpoint was varied in 50 milliscond steps, and the rise endpoint was either 10 or 19 semitones. The CI users could only discriminate between rises with at least 200 ms difference in rise startpoints, while the NH listeners only required 100 ms difference. Both groups most frequently interpreted the L*H-H% tunes as questions and the L*L-H% tunes as statements. For NH listeners, earlier rise startpoints were associated with more 'question' responses.

Keywords: intonation, perception, cochlear implants

1. INTRODUCTION

Analyses of discourse interaction in Australian English have reported the use of rising intonational tunes for a range of different discourse functions. High-onset high rises (H*H-H% in the Tone and Break Indices (ToBI) annotation scheme [5]) are often associated with information requests [1-4]. Rises occurring on statements are also frequently reported and characterise the 'uptalk' for which Australian speakers are known [2,6].

Statement rises are reported to differ both in phonetic realisation and in discourse function, with low-onset expanded range rises (L*H-H%) used for directives and low-onset narrow range rises (L*L-H%) used for acknowledgements and backchannels [2]. In contrast, McGregor and Palethorpe [2] found L*H-H% used in both questions and statements in their production study.

In terms of listener interpretation of these low-onset rising tunes, Fletcher and Loakes [3] found L*H-H% tunes were more likely to be interpreted as questions as the pitch span increased (i.e. as the height of the boundary tone increased). Tunes with a low-onset and low rise (L*L-H%) were more frequently interpreted as statements in their study.

For cochlear implant (CI) users, for whom the accurate perception of pitch changes is difficult [8], this relationship between tune realisation and tune use may pose particular challenges. At particular risk are those prelingually deaf CI users who have had only the input from their CI from which to develop all of their phonological representations of language. These individuals tend to perform well on measures of segmental perception [9] but relatively little attention has been directed to their perception, production or interpretation of intonation.

In terms of the perception of rises, Peng et al. [10] asked a group of prelingually deaf CI users to classify a series of recorded utterances as being statements or questions. No detail is provided as to the intonation of the utterances, so it is assumed that the 'question' utterances included a final rise. While 97% of the normally-hearing (NH) group correctly classified the stimuli, only 70% of the CI users' classifications matched the target discourse function.

Holt and McDermott [11] investigated the perception of final rises and of rising pitch accents. Their prelingually deaf CI group could discriminate perceptually between a flat intonation contour and a linear rise of 11.5 st with 92% accuracy. They also explored the discrimination of pitch accents with the same rise but differing alignment within a syllable. They found that the CI users could not discriminate alignment changes, although their NH peers performed significantly above chance levels once the F0 peak alignment reached 80 ms. These results clearly indicate that the perception of rise startpoints is at issue for CI users.

Given that L*H-H% and L*L-H% are reported to be associated with potentially different discourse functions in Australian English, it is important for speech understanding that CI users can 1) perceive the difference between these two tunes, and 2) interpret them as do their NH peers. In this study, we use stimuli with manipulated F0 to explore both these questions.

2. METHOD

2.1. Participants

Ten CI users aged between 12.3 – 18.8 years completed the testing. The mean age at implantation was 3.0 years (*SD* 1.3). All used Nucleus devices from Cochlear Ltd, Freedom processors and the Advanced Combinational Encoder (ACE) or SPeak processing strategies. All have congenital profound deafness, with three participants receiving a second CI during early adolescence. A control group (NH group) comprised 19 adolescents with typical acoustic hearing. The mean age at time of testing for the CI group was 14.4 years (*SD* 2.2), and for the NH group, 14.7 years (*SD* 1.6).

2.2. Stimuli

An adult male speaker of Australian English recorded the utterance *I think it was Melanie*. The original tune for the utterance was H* L* L-H%, with L* L-H% occurring on the word *melanie*. Using Praat [12], the word *melanie* was excised from the utterance, stylised using a Pitch-Synchronous-Overlap-Add (PSOLA) technique, and its F0 manipulated so as to create two series of seven stimuli. We created two linear rises beginning at the onset of the first syllable of *melanie*: one with a pitch span of 19 st ('expanded rise') and one of 19 st ('narrow rise'). The startpoint (elbow) of each rise was then moved rightwards through the word *melanie* in 50 millisecond (ms) increments to create 12 further stimuli (see Figure 1). This resulted in two sets of stimuli: one with an expanded rise and seven different alignments of the rise startpoint, and; one with a narrow rise also having seven different rise startpoints.

2.3. Procedure

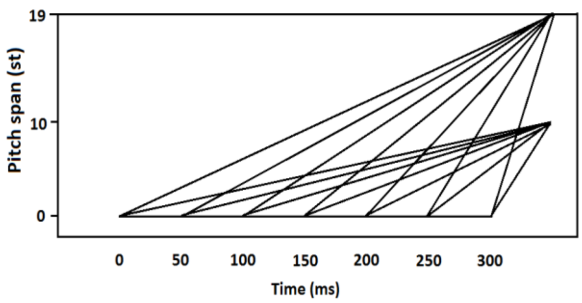
2.3.1. Discrimination task

All participants were tested individually in a sound-treated booth. All CI users were tested using only their first (or sole) CI. Stimuli were presented at 65 dB SPL from a loudspeaker located 1 metre directly in front and responses were automatically recorded via a touchscreen.

An AXB design was used to present the stimuli in each experiment. Participants heard three stimuli in total, and had to choose whether the second stimulus presented was the same as the first or third presented. For both the expanded and narrow rise series, 6 stimulus pairings were created, with the stimulus with the earliest rise startpoint being paired with each of the other stimuli (i.e. 1-2, 1-3, 1-4, 1-5, 1-6, 1-7). Each of the pairs were presented 3 times, in each configuration of the AXB design (e.g. 112, 211, 221, 122 etc). The stimuli were presented in two blocks of 72 trials, with the order of presentation randomised.

The NH participants completed only the narrow rise series (72 trials), as the pitch spans used were large and so discrimination of the rises were not considered to be at issue for those with typical hearing.

Figure 1: Stylised illustration of the two stimuli series. The rise elbow (i.e. onset of the rise) was moved rightwards through the word *melanie* in 50 ms increments to create a series of seven stimuli with two pitch spans (19 st – expanded rise; 10 st – narrow rise).



2.3.2. Identification task

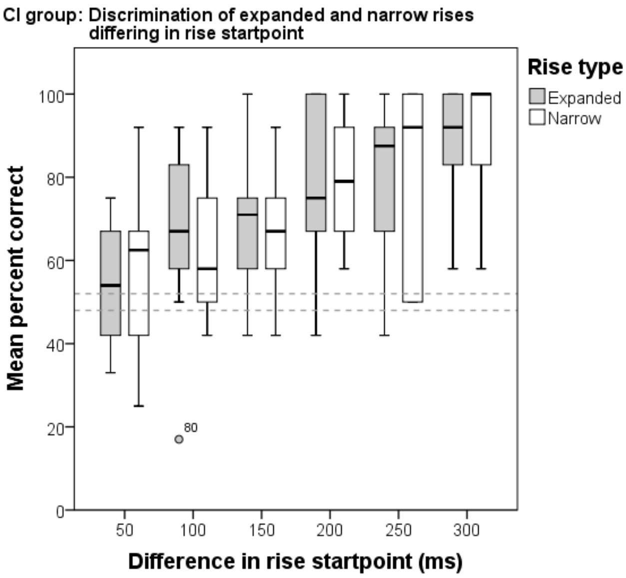
The expanded and narrow rise stimuli formed the basis of an identification task, which all participants completed. Using the same equipment as for the discrimination task, participants heard each manipulated token of *melanie* presented in the original carrier phrase *I think it was.* They were asked to nominate whether it sounded more like a question or a statement. The trials were presented as two blocks of 84, with the order of presentation randomised.

3. RESULTS

3.1. Discrimination task

The CI group completed the discrimination task for both the expanded and narrow rise stimuli, while the NH group completed the narrow rise discrimination task only. Figure 2 presents CI group results comparing discrimination of changes in rise elbow alignment for both the expanded and narrow rise stimuli. The boxplot was derived using the mean percent correct for each participant over 12 trials for each stimulus pair.

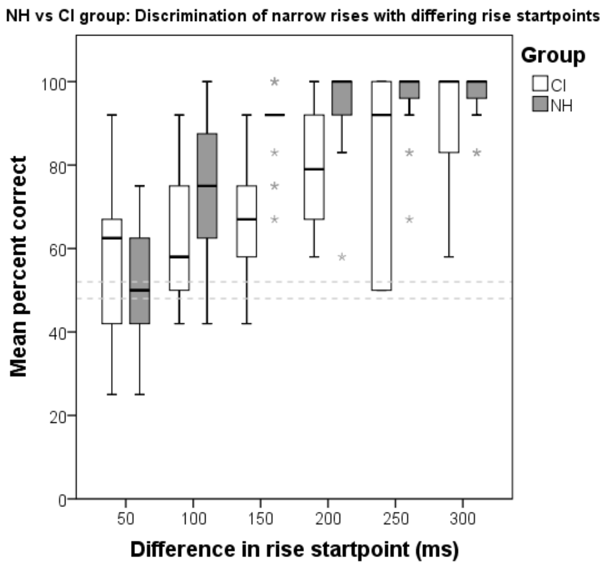
Figure 2: Mean percent correct scores for the CI users for narrow rise (clear bars, left) and expanded rise series stimuli (dark grey bars, right) for each stimulus pair. The rise elbow increases by 50 ms between each stimulus pair. The black line in each box represents the median score. The dashed lines represent scores significantly above and below chance.



Performance was only significantly above chance for *all* CI users once the difference in rise elbows was 300 ms. A repeated-measures ANOVA was performed on the CI group data with Rise (Expanded, Narrow) as a between-subjects factor and Rise Startpoint as a within-subjects factor. There was no significant effect of Rise on scores, and no significant interaction between Rise and Rise Startpoint. The effect of Rise Startpoint was significant (F [5,90] = 15.85, $p < 0.0001$). Post-hoc paired t-tests indicated that, for both expanded and narrow rise stimuli, discrimination scores were significantly larger once the difference in ms between the rise elbows was 200 ms or more.

The NH group completed discrimination of only the narrow rise stimuli. Figure 3 presents their mean correct percent scores (derived as per Figure 2) compared to the performance of the CI group. A repeated-measures ANOVA was performed with Group (NH, CI) as a between-subjects factor and Rise Startpoint as a within-subjects factor. The interaction of Group and Rise Startpoint was significant (F [5,27] = 4.93, $p < 0.0001$). Post-hoc independent samples t-tests indicated that the CI users performed significantly more poorly than the NH listeners when the stimuli differed by 150, 200 and 250 ms in rise elbow alignment. A post-hoc paired-sample t-test of the NH listener results indicated that scores increased significantly once the difference in rise startpoint was 100 ms or more.

Figure 3: Mean percent correct scores for the CI users (clear bars, left) and NH listeners (dark grey bars, right) for narrow rise stimuli. The black line in each box represents the median score. The dashed lines represent scores significantly above and below chance.

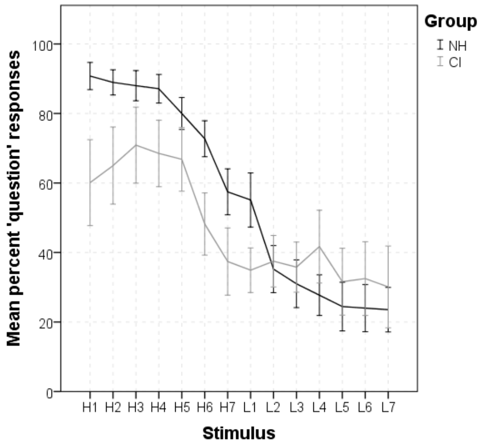


3.2. Identification task

One female NH listener did not complete the identification task due to time constraints on the testing day. The mean percentage of ‘question’ responses for both groups for each stimulus and rise type are presented in Figure 4. A repeated measures ANOVA with Group and Rise as between-subjects factors and Rise Startpoint as a within-subjects factor indicated a significant interaction of Group and Rise (F [1,52] = 4.82, $p < 0.03$). For the NH listeners, more ‘question’ responses were associated with the expanded rises and with earlier rise startpoints (i.e. when the rise was longer in duration). The narrow rises were more often identified as ‘statements’, with more ‘statement’ responses elicited for later rise startpoints (i.e. as the rise became shorter).

The CI user responses patterned differently, with more ‘question’ responses elicited for the expanded rise rather than narrow rise stimuli, but earlier rise onsets were *not* associated with more ‘question’ responses (see Figure 4). Although the CI users judged the expanded rises to be question-like more often than the narrow rises, they judged the narrow rises as questions more often than their NH peers.

Figure 4: Mean percent correct 'question' responses for the CI users (light grey lines) and NH listeners (black lines) for both expanded and narrow rise stimuli. On the x-axis, expanded rise stimuli are marked 'H' and narrow rise stimuli as 'L'.



4. DISCUSSION

We investigated whether differences in the alignment of the rise startpoints could be discriminated by listeners who use CIs and their peers with typical hearing. We found that compared to NH listeners, the CI users (as a group) could not successfully discriminate between rises with startpoints which differed by less than 200 ms. A rise endpoint of 19 st or 10 st did not affect discrimination between rise startpoints. There was however wide variation in performance between CI users, with some users not scoring above chance until the difference in rise startpoints reached 300 ms. This is contrast to the performance of the NH listeners, who could discriminate rise startpoint differences of 100 ms or more.

These results indicate that NH listeners are perceptually sensitive to changes in the startpoint of a rise (and the CI users less so). In interpretation, we found a consistent effect of rise startpoint for our NH listeners, similar to that reported by Warren [13] for New Zealand English. Early rise startpoints (i.e. *slower* rises) biased listeners towards 'question' responses, with more 'statement' responses returned for later rise startpoints (equivalent to *faster* rises). This pattern held for both expanded rises and narrow rises, indicating that either the stretch of low level F0 between the pitch accent and boundary tone, or the brief short rise (or both), produces a "statement" interpretation.

The expanded rises with early startpoints were consistently identified as questions by the NH listeners, in the same way as H*H-H% tunes were identified as questions in [7]. It may be the case that the low pitch elbow of these very early rises were instead interpreted as the starting point for rising accents (L+H*) by our NH listeners. The difference in pitch span was however also important, as NH listeners reported more questions overall for stimuli with expanded rises regardless of rise startpoint. The narrow rises were more likely to be interpreted as statements. These results therefore also support those reported in [7].

As in [10], the CI users again displayed a difficulty in perceiving differences in the alignment of F0 turning points. Encouragingly, the CI users did *generally* pattern with the NH listeners when interpreting these two tunes as either questions or statements, although earlier rise onsets with expanded rises were *not* interpreted as questions to the same extent as they were by the NH group. It may be that the CI users did not associate the early rise startpoints with rising accents, as is possible in the case of the NH listeners.

As the CI users of this study were seen to have some difficulty in perceiving aspects of intonation that were perceivable for NH listeners, this brings into CI users' performance in complex discourse interactions. An obvious limitation of this study is that listeners were provided with only a binary choice between question and statement, so we could not assess whether expanded and narrow rises are interpreted differently in terms of signalling directives or acknowledgements. Understanding all of these differences are important to speech understanding, and so future studies investigating how CI users' perceive and produce intonational tunes in interactive discourse are warranted.

5. REFERENCES

1. Fletcher, J., Stirling, L., Mushin, I., Wales, R. 2002. Intonational rises and dialog acts in the Australian English map task. *Lang Speech*, 45(3): 229-253.
2. McGregor, J., Palethorpe, S. 2008. High rising tunes in Australian English: The communicative function of L* And H* pitch accent onsets. *Australian Journal of Linguistics*, 28(2): 171-193.
3. Fletcher, J., Loakes, D. 2010. Interpreting rising intonation in Australian English. In *SP-2010*, 100124:1-4, ISBN:978-0-557-51931-6.
4. Fletcher, J., Wales, R., Stirling, L., Mushin, I. 2002. A dialogue act analysis of rises in Australian English Map Task dialogues. In *SP-2002*, 299-302.
5. Beckman, M., Hirschberg, J. Shattuck-Hufnagel, S. 2005. The original ToBI system and the evolution of the ToBI framework, in *Prosodic Typology*, S.-A. Jun, Editor. Oxford University Press: Oxford. 9-54.
6. Fletcher, J., Grabe, E., Warren P. 2005. Intonational variation in four dialects of English: the High Rising Tune, in *Prosodic typology*, S.-A. Jun, (Ed), Oxford University Press: Oxford. 390-409.

7. Fletcher, J., Harrington, J. 2001. High-rising terminals and fall-rise tunes in Australian English. *Phonetica*, 58(4): 215-229.
8. Moore, B. 2007. *Cochlear hearing loss*. 2nd ed., Chichester, England; John Wiley & Sons.
9. Harrison, R., Gordon, K., & Mount, R. 2005. Is there a critical period for cochlear implantation in congenitally deaf children? *Developmental psycho-biology*, 46(3), 252-261.
10. Peng, S.C., Tomblin, J.B. & Turner, C. 2008. Production and perception of speech intonation in pediatric cochlear implant recipients and individuals with normal hearing. *EarHear*, 29(3), 336-351.
11. Holt, C. & McDermott, H. 2013. Discrimination of intonation contours by adolescents with cochlear implants. *International Journal of Audiology*, 52(12), 808-815.
12. Boersma, P., Weenink, D. *Praat: Doing phonetics by computer*. 2007. p. Retrieved February 18, 2007 from http://www.praat.org/.
13. Warren, P. 2005. Patterns of late rising in New Zealand English: Intonational variation or intonational change? *Language Variation and Change*, 17(02): 209-230.