CLEAN VERSION

Intensity discrimination and speech recognition of cochlear implant users

Colette M. McKay[1, 2], cmckay@bionicsinstitute.org

Natalie Rickard[1], natalierickard@gmail.com

Katherine Henshall[1], k.henshall@unimelb.edu.au

1. Bionics Institute, Melbourne, Australia.

2. The University of Melbourne, Department of Medical Bionics, Australia

Running title: Intensity discrimination and speech recognition

Send correspondence to:

Colette McKay

The Bionics Institute of Australia

384-388 Albert St

East Melbourne, 3002

Australia

Phone +61 3 96677541

Email cmckay@bionicsinstitute.org

Number of words

Abstract 219

Intro 752

Discussion 1016

ABSTRACT:

The relation between speech recognition and within-channel or across-channel (i.e. spectral tilt) intensity discrimination was measured in 9 CI users (11 ears). Within-channel intensity difference limens (IDLs) were measured at four electrode locations across the electrode array. Spectral tilt difference limens were measured with (XIDL-J) and without (XIDL) level jitter . Only 3 subjects could perform the XIDL-J task with the amount of jitter required to limit use of within-channel cues. XIDLs (normalized to %DR) were correlated with speech recognition (r = 0.67, p = 0.019) and were highly correlated with IDLs. XIDLs were on average nearly three times larger than IDLs and did not vary consistently with the spatial separation of the two component electrodes. The overall pattern of results was consistent with a common underlying subject-dependent limitation in the two difference limen tasks, hypothesized to be perceptual variance (how the perception of a sound differs on different presentations), which may also underlie the correlation of XIDLs with speech recognition. Evidence that spectral tilt discrimination is more important for speech recognition than within-channel intensity discrimination was not unequivocally shown in this study. However, the results tended to support this proposition, with XIDLs more correlated with speech performance than IDLs, and the ratio XIDL/IDL also being correlated with speech recognition. If supported by further research, the importance of perceptual variance as a limiting factor in speech understanding for CI users has important implications for efforts to improve outcomes for those with poor speech recognition.

Keywords: cochlear implant; intensity discrimination; speech recognition

INTRODUCTION

One of the most persistent challenges in the cochlear implant (CI) field is the high variation in speech perception outcomes, even in adults with post-lingual deafness. Up to a third of such CI users have minimal ability to understand sentences without lipreading (for example words in sentences less than 50% correct in quiet) (Blamey, et al. 2013). Many psychophysical studies have been performed to try to understand the basis of poor speech recognition in individuals. Armed with this knowledge, research could be guided in efforts to overcome the limitations experienced by the CI users with poor speech recognition. In this study, the relation between speech recognition and the ability of CI users to detect within-channel and across-channel (spectral tilt) intensity changes was examined.

Auditory speech recognition relies on multiple acoustic and contextual cues. To perceive speech cues in a complex speech stimulus, the auditory system must employ spectral and temporal analyses of the signal (i.e. spectral and temporal resolution). Although the intelligibility of speech is little affected by overall level changes (provided audibility is not compromised) (e.g. Lu and Cooke 2009), spectral resolution relies on being able to compare intensities across different simultaneous frequency components and temporal resolution relies on being able to follow intensity changes over time within a channel.

Many studies have demonstrated a correlation between spectral ripple detection or discrimination and speech recognition (Anderson, et al. 2011; Anderson, et al. 2012; Drennan, et al. 2016; Henry and Turner 2003; Lawler, et al. 2017; Litvak, et al. 2007; Saoji, et al. 2009; Winn, et al. 2016; Won, et al. 2014; Won, et al. 2011). However, the spectral ripple stimuli are complex and can potentially be discriminated or detected using multiple different types of cues, such as subtle pitch or loudness changes and within-channel amplitude modulations, in addition to resolution of individual ripples. Saoji et al. (2009) measured spectral ripple detection thresholds across a range of spectral modulation frequencies, while controlling loudness cues. They found that the relation between ripple detection and speech recognition could be wholly accounted for by ability to detect ripples of the lowest ripple density (in which the task is similar to detecting a broad spectral tilt), and not by the rate at which detection deteriorated as ripple density increased (related to spectral resolution). An inference from that study could be that the ability to detect relative intensity changes across well-separated channels in a multi-electrode stimulus is a basic psychophysical ability underlying spectral ripple detection. A study by Anderson et al (2012) compared spectral ripple detection and discrimination thresholds in the same CI users. Although performance on the two tasks was correlated, thresholds at the same modulation depth were not equivalent (as expected if the subjects were using identical cues in each task). They found intensity difference limens for broadband noise were not correlated with spectral ripple detection. They inferred that ripple detection did not rely on within-channel cues and therefore there must be a disassociation between the ability to detect small changes of intensity across frequency or across time.

Amplitude modulation detection has also been related to speech perception performance in some studies, either using direct electrical stimulation on single electrodes or broadband acoustic stimulation via the speech processor (Brochier, et al. 2017; Luo, et al. 2008; Won, et al. 2011). In these cases, the parameter correlated with speech recognition was the sensitivity at low modulation frequencies, which is highly correlated with intensity discrimination (Galvin and Fu 2009), rather than the slope or cut-off frequency of the modulation transfer function (related to temporal resolution).

Although spectral and temporal resolution are necessary for speech recognition, the psychophysical studies discussed above suggest that differences in these abilities are not the major factors, in general, that lead to differences in speech recognition among CI users. This may be the case because speech recognition, at least in quiet, does not require extremely fine resolution (e.g. Shannon, et al. 1995), and implantees generally have sufficient resolution ability. In the temporal domain, Fraser and McKay (2012) showed that CI temporal modulation transfer functions had low-pass modulation frequency cut-offs (hence temporal resolution) broadly similar to those of normal hearing subjects. Additionally, there is little evidence that the width of the central temporal integration window (which limits temporal resolution) in CI users differs from that in normal hearing listeners (McKay, et al. 2013).

A hypothesis that arises from the psychophysical studies described above is that subject-dependent variation in intensity discrimination is a psychophysical factor underlying the association of performance in speech recognition tasks and performance in amplitude modulation or spectral ripple detection tasks. It is somewhat surprising that simple intensity discrimination measures have received little attention by researchers who have investigated variation in speech recognition in CI users. In this study, we hypothesized that within-channel (single electrode) intensity discrimination and across-channel relative intensity discrimination (in a dual-electrode 'spectral tilt' task) are correlated with speech recognition. Spectral tilt is a broad-spectrum-based acoustic cue that is important for speech recognition, especially stop consonants (Alexander and Kluender 2008, 2009). Furthermore, compared to normal hearing listeners, people with hearing impairment (Alexander and Kluender 2009) and CI users (Winn and Litovsky 2015) increase the weighting of spectral tilt cues over the finer spectral formant cues . We therefore further hypothesized, based on the study of Anderson et al. (2012) and the importance of spectral tilt cues for CI users, that differences in the discrimination of spectral tilt would be more related to differences in speech recognition than within-channel intensity discrimination .

METHODS

Subjects and equipment

Nine adult cochlear implant users participated in this study. All were users of Nucleus-family cochlear implants, manufactured by Cochlear Ltd. Two subjects with bilateral implants provided data from each ear. All participants signed an informed consent form and the project was granted ethical approval from the Human Ethics Committee of the Royal Victorian Eye and Ear Hospital. The details of the participants' implants, clinical speech processor types, hearing history and etiology are contained in Table 1. Parameters of their clinical speech processing strategies are shown in Table 2.

Psychophysical measurements were carried out using direct electrical stimulation via ImPResS software, interfaced with the implant via a SPEAR processor (Zakis and McDermott 1999). The software defined the stimulus parameters, ran the experimental procedure, and collected subject responses via a response box. The processor sent the coded instructions for each stimulus directly to the implanted electronics via the usual radio frequency link.

Speech recognition measurements were undertaken using direct audio input (DAI) from a sound card to the participants' own speech processors. The calibration of the equivalent input sound level for DAI testing was performed by matching the output current levels on an oscilloscope using an implant in a box. All participants used the ACE or SPEAK strategy (see Table 2), and the speech processors were set to their 'everyday' program. DAI prevented the use of residual hearing in the speech tasks and also facilitated the separate testing of each ear in two of the participants who were implanted bilaterally. The speech recognition and psychophysical tasks were carried out in at least three separate sessions, which lasted for a maximum of 1.5 hours with breaks.

Speech Recognition

Speech recognition in quiet was assessed with consonant-vowel nucleus-consonant (CNC) words. One list of 50 CNC words at an equivalent level of 65 dBA was presented, and the subjects' responses were scored as the percentage of phonemes correctly identified.

Sentence recognition was assessed using CUNY sentences in quiet, and in multi-talker babble at signal-to-noise ratios (SNRs) of +15, +10, and +5 dB. In each condition, one sentence list was presented and scored as percentage of words correct (out of approximately 100 words in each sentence list). The order of sentence testing was the quiet condition first followed by increasingly difficult noise conditions. No more-difficult conditions were tested if the score at one SNR dropped below 20% correct. Since the absolute speech recognition ability of subjects varied greatly in quiet, and some subjects could not complete the tests with low SNRs, there were no fixed SNRs that would not suffer from either floor or ceiling effects in the analysis. Therefore, a parameter was derived (denoted SNR-H) from each individual function of percent correct versus SNR that represented the SNR at which the speech score was reduced to half of the score in quiet. The SNR-H was therefore, unlike a speech reception threshold, not an absolute speech recognition performance measure, but a measure of the effect of noise on an individual's own speech recognition performance.

Intensity Discrimination

Electrode selection

Four active electrodes that spanned the electrode array were selected for evaluation of intensity discrimination measurements: two in the mid-array (10, 13) that were 3 electrodes apart, and two (3, 20) that were very distant. The electrode selection was motivated by optimal sampling of locations and separations. The electrodes were activated in monopolar mode (return electrode outside the cochlea) in all cases except for subject S4 who used bipolar + 1 mode (intracochlear return electrode 2 electrodes apical to the active electrode). Within-channel intensity discrimination was assessed on each of the four active electrodes using single-electrode stimuli, and spectral tilt discrimination was assessed using dual-electrode stimuli (stimulus pulses interleaved on two electrodes) using each of the two sets of electrodes.

Stimulus parameters

Each single-electrode stimulus was a 500-ms-duration constant-current train of biphasic pulses. As the intention was to correlate intensity discrimination with speech recognition with the participants' own processors, the rate of stimulation, mode, phase duration and interphase gap were all set to the same values as in their clinical map. Subject S4 used an early model implant and a speech processor that used a mixture of phase duration and current level to encode amplitude. For this subject, a fixed phase duration of 100 μs was used in the psychophysical task, a value representative of the range in his speech processor. The speech processor parameters for all participants are listed in Table 2. The dual-electrode stimuli consisted of two interleaved single-electrode stimuli with the inter-electrode delay adjusted to make the overall pulses evenly distributed in time.

Creating the reference stimuli

Reference current levels for the reference stimuli in the discrimination tasks were set such that a) each electrode of a pair evoked equal loudness and b) the dual-electrode reference stimulus was approximately at the mid-point in its dynamic range, and c) the reference current level on an electrode was the same in dual- and corresponding single-electrode reference stimuli. The second criterion ensured that there was sufficient room to produce spectral tilts in the dual-electrode stimulus, as these were made by adding the same number of current levels to pulses on one electrode of the pair as the number of current levels subtracted from pulses on the other electrode of the pair. The third criterion ensured that the within- and across-channel discrimination measures were obtained with equal reference current levels. To achieve these three criteria, single-electrode stimuli were first created with twice the rate of the clinical processor (same overall rate as the dual-electrode stimuli). The threshold and maximum comfortable level of these stimuli were determined, and the current level of one of each pair (3 or 10) was set to midway between the threshold and maximum level. A loudness balance procedure was then used to find the current level on the other electrode of each pair (20 or 13) that evoked an equal loudness to that on electrodes and 3 or 10, respectively. The dual-electrode reference stimuli were then created (using the same overall rate) using the loudness balanced current levels. This procedure was based on the predictions of the loudness model of McKay et al. (2003) that the dual-electrode stimulus would be at around 50% of its dynamic range with equal contributions to loudness of the two electrodes, without having to re-adjust the current levels after creating the dual-stimulus to achieve this. The single-electrode reference stimuli were then constructed to have the same per-electrode rate (the clinical rate) and the same currents as the dual-electrode stimulus. That is, the single-electrode stimuli were exactly the components of the related dual-electrode stimulus.

In the above procedure, thresholds were determined using a 3-interval 3-alternative adaptive procedure with a two-down, 1-up adaptive rule. Subjects were asked to nominate which of the three intervals (randomly assigned) contained the sound. Step sizes were 4 current levels (CL) for 2 turns, and 2 CL for 8 turns, with threshold defined as the average CL at the last 6 turn points. Thresholds were measured twice and averaged. Maximum comfortable loudness was determined using a visual category scale with seven categories between ‘not heard’ and ‘too loud’. Maximum comfortable level was defined as the CL at which the subject nominated ‘very loud but tolerable’. Loudness balancing was completed using an adaptive 2-interval forced choice task. The target and reference stimuli were assigned randomly to the two intervals and the task of the subject was to choose the louder sound. The target stimulus was adjusted with a 1-down 1-up rule, with the same step sizes and number of turn points as the threshold task, and the balanced CL was defined as the average CL at the final 6 turn points. The balance procedure was performed a total of 4 times, with each stimulus being the fixed reference twice. The final balanced current level was determined using the average CL difference between the two stimuli over the four balancing runs.

Single-electrode intensity difference limens

Single-electrode intensity difference limens (IDLs) were determined using a 2-interval forced choice task with an adaptive 2-down 1-up rule (asymptoting to the CL at which the target was considered louder than the reference with 71% probability). The task of the subject was to choose the louder sound. The same step sizes and turn points were used as in the loudness balancing task, with the average CL of the last 6 turn points averaged. The procedure was repeated twice more and the average of the 3 resultant differences between reference and target CLs was defined as the IDL.

Dual-electrode (spectral tilt) difference limens

Spectral tilt DLs (denoted XIDL) were measured using dual-electrode stimuli. The reference stimulus was the one constructed to have equal loudness contributions from each component electrode. The target stimulus was one in which pulses on one electrode had an increased CL, and pulses on the other electrode had a CL decreased by the same amount (see Figure 1). Multiple targets with differing size of adjustments of CLs were constructed. For each adjustment size, there were two targets with differing direction of adjustment.

The task for the XIDLs was a 4-interval 4-alternative forced choice task, in which the subject was asked to nominate the interval with the randomly assigned 'different' (target) sound. The XIDL was determined using the method of fixed stimuli and a psychometric function (percent correct versus current adjustment). Each psychometric function included at least 6 targets (at least 3 pairs of targets: 3 adjustment sizes and each size with 2 directions of adjustment). In each experimental run, all the targets were included and selected in random order without replacement for each trial in the run, until 30 responses were obtained for each target. An example psychometric function is shown in Figure 2, with illustration of how the XIDL was defined. The different directions of adjustment are shown separately as different sides of the function. The XIDL was defined as the average of the two linearly interpolated CL adjustments (absolute values) for 71% correct identification from the two sides of the function. The range of adjustment sizes to use for each participant's psychometric function test run with interleaved targets was first determined by considering the size of the single-electrode IDLs and then running some test trials with increasing-sized adjustments (including both directions of adjustment in a run) until the targets were identified for more than 8 out of 10 trials). That target was then used as the one with the maximum CL adjustment in the psychometric function, and the psychometric function test run was finally performed with at least three levels of adjustment including two smaller ones. When plotting the psychometric functions a non-data point was included (an open square symbol in Figure 2) representing 25% chance score when the target had zero adjustment relative to the reference.

In the XIDL task, both within- and across-channel cues can be used by the subject to identify the target. The subject may be able to attend to one or both changes in current at each single-electrode site (within-channel cues) to identify the target. If so the XIDL should be similar to or better than the IDLs on the component electrodes. This would be particularly true for the distant electrode pair (3, 20) if the two areas of neural excitation were resolved by the listener. This benefit of listening to within-channel cues would be partially offset by the fact that the XIDL was measured using a 4IFC task whereas the IDL was measured using a 2IFC task[1]. On the other hand, if the subject were attending to the change in *relative* current between the two electrodes, they would perceive the target stimulus as the one with a different pitch or timbre than the reference stimulus.

It was assumed, to a first approximation, that overall loudness cues did not significantly contribute to the XIDL task, as each target had an equal increase and decrease in CLs compared to the reference stimulus. McKay et al (2003) showed that, for high rates and moderate or low levels in the dynamic range, the loudness growth functions are linear on a log-log scale (log loudness versus CL), and thus the two CL changes in opposite directions would nearly completely balance out. The assumption that the positive and negative changes in loudness contribution from each electrode in the target dual-stimulus were balanced relied on the assumption that the two component electrodes had similar slopes of loudness growth with

[1] 4IFC DLs would be approximately 1.2 times 2IFC DLs if the same stimuli and discrimination were tested in both cases. This ratio was estimated by fitting Weibull functions to two alternative psychometric functions.

increased CL. This assumption was deemed to be likely satisfied if the two electrodes had a similar dynamic range (DR). In some cases, the electrode selection for the psychophysical task was altered to ensure that electrodes had similar DRs. Electrode 3 was replaced by electrode 6 for subject S7-L, and Electrodes 3 and 20 were replaced by electrodes 6 and 21 for subject S8 and by electrodes 4 and 21 for subject S7-R. It was, however, likely that small overall loudness differences remained between target and references. It was also assumed in the XIDL measurement that any such residual differences in overall loudness between reference and targets would be smaller than the overall loudness difference limen for the dual-electrode stimulus. This assumption was likely to hold unless the threshold adjustments in the XIDL task were very large compared to the component IDLs.

A third experimental measure was undertaken (XIDL-J) that aimed to limit the possible use of the within-channel cues in the spectral tilt discrimination task. The procedure was the same as for the XIDL task, except for the addition of level jitter to each interval in the task. The purpose of the level jitter was to limit the use of *within-channel* changes of level. This amount of jitter was greater than that needed to limit *overall* loudness cues. The range of level jitter was calculated via the methods of Dai and Micheyl (2010) based on the maximum within-channel level adjustment for targets in each psychometric function and a maximum unwanted probability for use of within-channel level cues of 50% correct (since our threshold XIDL-J was defined as the adjustment for 71% correct). Thus, each target stimulus in a single psychometric function test run had the same level jitter applied, but each subject had differing (minimum but sufficient) amounts of level jitter based on the size of the actual level adjustment needed to obtain at least 71% correct in the psychometric function. The spectral tilt difference limen with jitter was calculated from the psychometric function (see Figure 2) in the same way as the XIDL.

Normalization of difference limens

Since the subjects had varying modes of stimulation, pulse durations, and dynamic ranges, it was necessary to normalize the difference limens to limit the effect of differing individual loudness growth functions. All difference limens were therefore expressed as a percentage of dynamic range (%DR) before analyzing their relation to speech recognition. For single-channel IDLs, the DR was the difference between threshold (T) and comfortable (C) levels in the subject's clinical map for the same electrode. This measure of DR was used because it is directly relevant to the intensity variations produced at the electrodes by the speech signal via the speech processor, and the purpose of the experiment was to seek associations between the measured IDLs and speech recognition *with their usual speech processor*. An exception was made for S4, whose processor map used mixed CL and phase duration to encode amplitude. In his case, the DRs of the single-electrode stimuli determined in the psychophysical task (with a fixed phase duration of 100 μs) were used. For XIDL and XIDL-J measurements, the DR was calculated as the average of the two relevant single-channel DRs.

RESULTS

Table 3 shows the speech recognition and psychophysical results for each subject. It can be seen that the XIDL values were about three times larger than the component IDL values, showing that subjects found the across-channel task more difficult than the within-channel task. Only 3 subjects could reach XIDL-J threshold with both directions of level adjustment as, in the remaining subjects, the jitter range required to limit within-channel cues exceeded the dynamic range of the dual stimulus.

Effect of electrode separation

It could be hypothesized that within-channel cues would be easier to use in the dual stimuli when there was a larger separation between the electrodes, and therefore XIDL values would be smaller for the electrode 3/20 pair compared to the electrode 10/13 pair if within-channel cues were being used. However, the mean size of the XIDLs were very similar being 29.9%DR and 28.5%DR for electrode pairs 10/13 and 3/20 respectively (see Table 3). A Wilcoxin Signed Rank test confirmed that there was no significant systematic difference in size between the XIDLs at different separations ($Z_{(9)} = -1.478$, $P = 0.16$). Furthermore, a Student's paired t-test showed that the ratio XIDL/IDL was not significantly different at the two separations ($t_{(9)} = 0.961$, $P = 0.36$), with differences varying among ears between +2.7 and -3.2. These results suggest that either subjects were not consistently using within-channel cues in the XIDL task or the pulse trains on the component electrodes were not resolved even in the large-separation case. Since there was no consistent effect of electrode separation, the two XIDLs for different separations were averaged for analysis of their relation to speech recognition. One subject (S9) only completed psychophysics on electrode pair (10, 13). In this case, the results from the one electrode pair were used in the analysis of relation to speech recognition, with the justification that there was no consistent effect of separation among the other 10 ears.

Influence of electrical dynamic range

Since all difference limens were normalized as %DR, the relation between DR and other measurements was first explored. Both speech recognition measures were not significantly correlated with the mean DR averaged across the four active electrodes (CNC words; $r_{(10)} = 0.38$, $P = 0.25$: SNR-H; $r(10) = -0.22$, $P = 0.52$). The average dynamic range was also not correlated with the un-normalized (in CL) average difference limens (IDLs; $r_{(10)} = -0.002$, $P = 0.99$: XIDLs; $r_{(10)} = -0.10$, $P = 0.77$). The latter analysis confirms previous reports that showed intensity difference limens were not correlated with dynamic range (e.g. Nelson, et al. 1996). Although the DR measure has some degree of variability due to the subjective nature of "comfortable loudness", the differences in DR between subjects is largely determined by the loudness growth slope, especially in these particular subjects, who vary in stimulation parameters (pulse duration, rate, mode) that affect DR and loudness slope similarly. Therefore, to explain the very low correlations between DRs and DLs the DLs must be significantly influenced by a factor other than loudness growth slope. This factor is the variance of the evoked stimulus percept as explained below.

According to signal detection theory, sensitivity (d-prime) to a current level change can be described as follows:

$$d' = \Delta L/\sqrt{\delta} \quad , \qquad (1)$$

Where ΔL is the average change in loudness (thus associated with the loudness growth function) and δ is the variance of the percept. A difference limen corresponds to a fixed value of d'. In the case of intensity DLs, δ is the variance of the loudness percept evoked by different instances of the stimulus. The variance has both peripheral and central contributions, from variability in the response of the auditory nerve or higher neural pathways, to sensory noise and uncertainly in the central loudness judgement decision. If δ were constant across subjects, then differences in d' would be governed by differences in ΔL (i.e. the loudness growth slope) and hence the raw IDL (in CL) would be inversely correlated with loudness growth slope. If this were the case, the IDLs would be at least moderately correlated with DR. The fact that no such correlation was found implies that the difference in δ across the subjects was a major factor influencing the differences in IDLs (in CL).

Relation between IDLs and XIDLs

A regression between XIDLs and their component IDLs (both in %DR) showed that they were highly correlated with each other ($r_{(20)}$ = 0.92, P < 0.001), as illustrated in Figure 3. The regression equation showed that XIDLs were on average 2.76 times larger than IDLs (with a regression constant not significantly different from zero). The high correlation suggests that the two measures are strongly influenced by a common mechanism. A strong correlation would be expected if subjects were using within-channel intensity discrimination to do the XIDL task; however, the larger XIDLs compared to IDLs weigh against this scenario, since the dual-electrode task could potentially provide two independent within-channel cues. An alternative common mechanism may be the influence of internal perceptual variance on both of the difference limen tasks, since it was inferred above that this variance has a significant influences on the differences between subjects in IDLs.

Since only 3 subjects could do the XIDL-J task with both directions of level adjustment, an alternative measure - the ratio XIDL/IDL - was derived, which may help to differentiate the within- and across-channel processing that occurred in the XIDL task. Since the regression line of XIDL versus IDL went through the origin, the variation in this ratio represents a measure of the variation of each XIDL from the regression line (where the ratio is a fixed value of 2.76). The ratio XIDL/IDL is likely to limit the influence of mechanisms common to both measures (such as variance and within-channel processing), and thus highlight any additional differences specific to across-channel processing ability among subjects.

Relation between difference limens and speech recognition

Table 4 shows the correlations between the two speech recognition scores (CNC words in quiet and SNR-H) and the three difference limen measurements (IDL, XIDL, and XIDL/IDL). Of all the difference limen measures, XIDLs accounted for the most variance (47%) in both speech recognition measures, as illustrated in Figure 4, with significant correlations with SNR-H and CNC scores. Within-channel IDLs were not significantly correlated with either speech score, although there was a trend for correlation with SNR-H scores. The ratio XIDL/IDL was significantly correlated with CNC word scores and showed a trend for correlation with SNR-H scores, as illustrated in Figure 5. However, a forward stepwise regression showed that IDL or XIDL/IDL did not explain any additional variance in either speech test score after accounting for XIDL.

Further qualitative analysis was performed on the XIDL-J scores by dividing the subjects into three groups by categories (see Table 3). Group A (S1, S2, S3) comprised those that achieved XIDL-J measures for dual-electrode targets with currents adjusted in both directions (i.e. at least 71% correct on the psychometric function within the limits of current adjustment available). Group B (S6, S8, S9) achieved 71% correct in one direction of adjustment only. Group C (S4, S5-L, S5-R, S7-L, S7-R) could not achieve 71% correct in either direction of adjustment. The mean CNC score was higher in Group A (90.7%) than in Group B (67%) or Group C (42.9%) and the mean SNR-H was lower in Group A (6.0 dB) than in Group B (7.2 dB) or Group C (14.6 dB). Although there is insufficient power to analyze these results statistically, the trend in the data supports the notion that cross-channel discrimination is important for speech recognition. .

DISCUSSION

The results of this experiment have shown that speech recognition (phonemes in CNC words) and the effect of noise on sentence recognition (SNR-H) are both significantly associated with the ability to discriminate relative level changes (spectral tilt) across two electrode positions (XIDL). Further, it can be deduced from a combination of two results (the high correlation of

IDL and XIDL, and the non-correlation of IDL with DR), that the differences among CI users of IDL, and hence XIDL, are significantly influenced by differences in internal perceptual variance.

In this study several techniques aimed to separate out the influences of within- and across-channel intensity processing on speech recognition. Our measure of XIDL-J was intended to separate these two processes. However, the participants with poor speech recognition had very large XIDLs that did not allow the use of a sufficient jitter range. Support for the suggestion that participants were not using within-channel cues in the XIDL task comes from the facts that a) there was no effect of electrode separation in the XIDL task and b) the XIDLs were nearly three times the size of the IDLs. Furthermore, the ratio XIDL/IDL, which should limit the use of cues common to both tasks (IDLs and perceptual variance) was also correlated with speech recognition (phonemes in CNC words). The modest power of this analysis, and the reliance on several assumptions, make unequivocal interpretations impossible. However, the overall pattern of results lend support to the proposal of Anderson et al (2012) that spectral tilt difference limens may rely on different processes than within-channel intensity difference limens, and that the former is more important in speech recognition.

Provided that the subjects were actually doing across-channel processing in the XIDL task, the proposition that perceptual variance was a limiting factor in both tasks could explain why XIDLs were on average nearly three times larger than the single-channel IDLs. If the subjects were judging the relative level of the two component channels in the XIDL task, as proposed, then the variance in this judgement would be twice as large as the variance present in a single channel task. On the other hand, if the subjects were doing the XIDL task by monitoring a single channel or two separate single channels, the XIDLs would be predicted to be more similar to the IDLs.

In summary, the results are consistent with poor XIDLs being associated with poor recognition of phonemes in CNC words, and with a large effect of noise on sentence recognition (SNR-H). Furthermore, the results are consistent with poor XIDLs being associated with large internal perceptual variance. It is not surprising that poor XIDLs are associated with poor speech recognition, especially phoneme recognition, as being able to recognize spectral shape cues is important for vowel recognition and for features of consonants such as place of articulation. More surprising is the evidence that internal perceptual variance may have a significant influence on differences in speech recognition among CI users. Although it has been noted before that within-channel IDLs are not correlated with loudness growth slope, and hence must be largely determined by perceptual variance, (e.g. Nelson et al. 1996), it can be deduced from the results of this study that perceptual variance may also influence speech recognition.

It is indeed very plausible that perceptual variance could have an influence on speech recognition, either by directly limiting the ability to recognize spectral shape, and/or less directly by limiting the ability of the CI user to learn new phoneme representations after implantation. The perceptual variance may alter the perceived spectral shape for different instances of the same phoneme, making it harder to adapt to the new auditory input.

A psychophysical experiment such as the one in this study cannot differentiate the peripheral and central sources of the perceptual variance. Poor cochlear health may be associated with high variance in peripheral neural responses (Shepherd and Hardie 2001) or higher sensory noise throughout the auditory pathways. On the other hand, cortical plastic changes due to auditory deprivation have been strongly associated with speech recognition after cochlear implantation (e.g. Lazard, et al. 2014; Lazard, et al. 2012), and deterioration in central decision mechanisms may contribute to the perceptual variance or uncertainty.

Implications for psychophysical markers of poor speech perception

The proposal that perceptual variance is likely to be a factor underlying the relation between speech recognition and intensity discrimination in CI users has important implications for interpretation of the relation of speech with other psychophysical measures, as well as for the aim of overcoming limitations in speech recognition. Every psychophysical measure that uses a standard discrimination task of comparing a reference and target stimulus will be affected by perceptual variance in a way similar to intensity discrimination. This is especially true for absolute measures in spectral and temporal resolution tasks, as each relies to a large extent on intensity discrimination, whether within or across-channels. It is noteworthy that the low-frequency cut-off frequencies in temporal or spectral modulation transfer functions have *not* been shown to be correlated with speech recognition (Fraser and McKay 2012; Saoji et al. 2009). This lack of correlation would be expected if perceptual variance and not spectral or temporal resolution *per se* was the factor affecting performance, particularly if the perceptual variance is not dependent on spectral or amplitude modulation frequency.

A corollary of the proposal that perceptual variance is a dominant factor limiting speech recognition, is that efforts to optimize outcomes for those with poor speech recognition would more be more effective if focused towards understanding the sources of this variance so that they can be limited, rather than focusing on methods to improve frequency or temporal resolution.

Future research efforts should focus on methods to differentiate the effects of perceptual variance from other perceptual processes. For example, Azadpour and McKay (2012) developed a novel method to measure spectral resolution that limited the influence of overall loudness differences or spectral shift. Using loudness models and comparisons with single-electrode IDLs, they showed that the ability of subjects to discriminate spectrally rippled and flat stimuli was highly influenced by perceptual variance, and that the large deterioration of performance with reduction in overall level was likely to be due to increased variance at lower stimulus levels.

CONCLUSIONS

The overall pattern of results of this study is consistent with poorer spectral tilt discrimination being associated with poorer speech recognition and with a larger effect of noise on sentence recognition. Furthermore the results are consistent with poorer spectral tilt discrimination being associated with larger internal perceptual variance.

It was concluded that perceptual variance is an important factor that limits both within- and across-channel intensity discrimination and therefore may be a common factor, at least in CI users, underlying the relation between spectral tilt discrimination and speech recognition.

Evidence that cross-channel discrimination is more important for speech recognition than within-channel discrimination in CI users was not unequivocally shown in this study. However, the results were not inconsistent with this proposition, supported by the facts that XIDLs were more correlated with speech performance than IDLs, and that the ratio XIDL/IDL was also correlated with speech recognition.

ACKNOWLEDGEMENTS

The ImpResS software was jointly developed by the University of Melbourne and Medical Research Council (UK), and the SPEAR research processor was developed by the Hearing Cooperative Research Centre in Melbourne. The research was supported by veski and Lions fellowships to the first author. The Bionics Institute acknowledges the support of the Victorian Government through its Operational Infrastructure Support program.

CONFLICT OF INTEREST

The authors declare that they have no conflict of interest.

REFERENCES

Alexander J M, Kluender K R (2008) Spectral tilt change in stop consonant perception. J Acoust Soc Am 123:386-96.

Alexander J M, Kluender K R (2009) Spectral tilt change in stop consonant perception by listeners with hearing impairment. Journal of speech, language, and hearing research : JSLHR 52:653-70.

Anderson E S, Nelson D A, Kreft H, Nelson P B, Oxenham A J (2011) Comparing spatial tuning curves, spectral ripple resolution, and speech perception in cochlear implant users. J Acoust Soc Am 130:364-75.

Anderson E S, Oxenham A J, Nelson P B, Nelson D A (2012) Assessing the role of spectral and intensity cues in spectral ripple detection and discrimination in cochlear-implant users. J Acoust Soc Am 132:3925-34.

Azadpour M, McKay C M (2012) A psychophysical method for measuring spatial resolution in cochlear implants. J Assoc Res Otolaryngol 13:145-57.

Blamey P, Artieres F, Baskent D, Bergeron F, Beynon A, Burke E, Dillier N, Dowell R, Fraysse B, Gallego S, Govaerts P J, Green K, Huber A M, Kleine-Punte A, Maat B, Marx M, Mawman D, Mosnier I, O'Connor A F, O'Leary S, Rousset A, Schauwers K, Skarzynski H, Skarzynski P H, Sterkers O, Terranti A, Truy E, Van de Heyning P, Venail F, Vincent C, Lazard D S (2013) Factors affecting auditory performance of postlinguistically deaf adults using cochlear implants: an update with 2251 patients. Audiol Neuro-otol 18:36-47.

Brochier T, McDermott H J, McKay C M (2017) The effect of presentation level and stimulation rate on speech perception and modulation detection for cochlear implant users. J Acoust Soc Am 141:4097.

Dai H, Micheyl C (2010) On the choice of adequate randomization ranges for limiting the use of unwanted cues in same-different, dual-pair, and oddity tasks. Atten Percept Psychophys 72:538-47.

Drennan W R, Won J H, Timme A O, Rubinstein J T (2016) Nonlinguistic Outcome Measures in Adult Cochlear Implant Users Over the First Year of Implantation. Ear Hear 37:354-64.

Fraser M, McKay C M (2012) Temporal modulation transfer functions in cochlear implantees using a method that limits overall loudness cues. Hear Res 283:59-69.

Galvin J J, 3rd, Fu Q J (2009) Influence of stimulation rate and loudness growth on modulation detection and intensity discrimination in cochlear implant users. Hear Res 250:46-54.

Henry B A, Turner C W (2003) The resolution of complex spectral patterns by cochlear implant and normal-hearing listeners. J Acoust Soc Am 113:2861-73.

Lawler M, Yu J, Aronoff J M (2017) Comparison of the Spectral-Temporally Modulated Ripple Test With the Arizona Biomedical Institute Sentence Test in Cochlear Implant Users. Ear Hear 38:760-766.

Lazard D S, Innes-Brown H, Barone P (2014) Adaptation of the communicative brain to post-lingual deafness. Evidence from functional imaging. Hear Res 307:136-43.

Lazard D S, Vincent C, Venail F, Van de Heyning P, Truy E, Sterkers O, Skarzynski P H, Skarzynski H, Schauwers K, O'Leary S, Mawman D, Maat B, Kleine-Punte A, Huber A M, Green K, Govaerts P J, Fraysse B, Dowell R, Dillier N, Burke E, Beynon A, Bergeron F, Baskent D, Artieres F, Blamey P J (2012) Pre-, per- and postoperative factors affecting performance of postlinguistically deaf adults using cochlear implants: a new conceptual model over time. PloS one 7:e48739.

Litvak L M, Spahr A J, Saoji A A, Fridman G Y (2007) Relationship between perception of spectral ripple and speech recognition in cochlear implant and vocoder listeners. J Acoust Soc Am 122:982-991.

Lu Y Y, Cooke M (2009) The contribution of changes in F0 and spectral tilt to increased intelligibility of speech produced in noise. Speech Commun 51:1253-1262.

Luo X, Fu Q J, Wei C G, Cao K L (2008) Speech Recognition and Temporal Amplitude Modulation Processing by Mandarin-Speaking Cochlear Implant Users. Ear Hear 29:957-970.

McKay C M, Henshall K R, Farrell R J, McDermott H J (2003) A practical method of predicting the loudness of complex electrical stimuli. J Acoust Soc Am 113:2054-63.

McKay C M, Lim H H, Lenarz T (2013) Temporal processing in the auditory system: insights from cochlear and auditory midbrain implantees. J Assoc Res Otolaryngol 14:103-24.

Nelson D A, Schmitz J L, Donaldson G S, Viemeister N F, Javel E (1996) Intensity discrimination as a function of stimulus level with electric stimulation. J Acoust Soc Am 100:2393-414.

Saoji A A, Litvak L, Spahr A J, Eddins D A (2009) Spectral modulation detection and vowel and consonant identifications in cochlear implant listeners. J Acoust Soc Am 126:955-958.

Shannon R V, Zeng F G, Kamath V, Wygonski J, Ekelid M (1995) Speech recognition with primarily temporal cues. Science 270:303-4.

Shepherd R, Hardie N (2001) Deafness-induced changes in the auditory pathway: Implications for cochlear implants. Audiol Neuro-Otol 6:305-318.

Winn M B, Litovsky R Y (2015) Using speech sounds to test functional spectral resolution in listeners with cochlear implants. J Acoust Soc Am 137:1430-42.

Winn M B, Won J H, Moon I J (2016) Assessment of Spectral and Temporal Resolution in Cochlear Implant Users Using Psychoacoustic Discrimination and Speech Cue Categorization. Ear Hear 37:e377-e390.

Won J H, Drennan W R, Nie K, Jameyson E M, Rubinstein J T (2011) Acoustic temporal modulation detection and speech perception in cochlear implant listeners. J Acoust Soc Am 130:376-88.

Won J H, Humphrey E L, Yeager K R, Martinez A A, Robinson C H, Mills K E, Johnstone P M, Moon I J, Woo J (2014) Relationship among the physiologic channel interactions, spectral-ripple discrimination, and vowel identification in cochlear implant users. J Acoust Soc Am 136:2714-25.

Won J H, Jones G L, Drennan W R, Jameyson E M, Rubinstein J T (2011) Evidence of across-channel processing for spectral-ripple discrimination in cochlear implant listeners. J Acoust Soc Am 130:2088-97.

Zakis J A, McDermott H J. (1999) A new digital sound processor for hearing research IEEE Engineering in Medicine and Biology Society Conference; Biomedical Research in the Third Millennium. Melbourne, Australia.

FIGURE LEGENDS

Figure 1. A schematic diagram illustrating the reference and test stimuli for the spectral tilt discrimination task. A and B represent the two component electrodes of the dual-electrode stimulus, with the horizontal line representing equal loudness for the reference stimulus. The test stimulus has electrode A current levels increased by X CL, and electrode B has currents reduced by X CL.

Figure 2. Example psychometric functions to derive XIDL (closed circles) and XIDL-J (open circles) for S2 with dual-electrode stimulation on E10 and E13. The x-axis refers to the increase or decrease of CL on E10 relative to the that in the reference stimulus: E13 always had the opposite adjustment to E10. For example the data point at x = 2 is that for E13 increased by 2 CL and E10 decreased by 2 CL. XIDL and XIDL-J were defined as the mean of the absolute intercepts of the interpolated functions with the 71% correct (dotted) line, then expressed as %DR.

Figure 3. Regression between across-electrode intensity difference limens (XIDLs) and the mean single-electrode difference limens (IDLs) using the same pair of electrodes.

Figure 4. Correlation of XIDLs with speech recognition: upper panel phonemes in CNC words, lower panel SNR-H.

Figure 5. Correlation of the ratio of XIDL/IDL with speech recognition: upper panel phonemes in CNC words, lower panel SNR-H.

Table 1. Details of participants' implants, speech processors, etiology, age, and implant experience.

Subject	Etiology	Implant type	Speech processor	Implant experience (years)
S1	Progressive/genetic	CI24RE	CP810	7
S2	Menieres	CI24RE	CP910	6
S3	Otosclerosis	CI24RE	CP810	5
S4	Otosclerosis	CI22	Freedom	22
S5-L	Progressive/genetic	CI24RE	CP810	7
S6	Progressive/genetic	CI24RE	Freedom	6
S5-R	Progressive/genetic	CI512	CP810	4
S7-R	Progressive unknown	CI24M	CP810	15
S8	Progressive/genetic	CI24R	CP810	15
S9	Progressive unknown	CI512	CP900	1
S7-L	Progressive unknown	CI512	CP900	5

Table 2. Details of the clinical map parameters of each participant (Strategy, maxima, rate (per-electrode), phase duration, interphase gap).

Subject	Mode	Strategy	Rate (pps)	maxima	Phase duration (μs)	Interphase gap (μs)
S1	MP	ACE	900	2	200	8
S2	MP	ACE	900	8	37	8
S3	MP	ACE	500	8	25	8
S4	BP+1 (stim level)	SPEAK	250	8	variable	45
S5-L	MP	ACE	900	8	25	8
S6	MP	ACE	900	8	25	8
S5-R	MP	ACE	900	8	25	8
S7-R	MP	ACE	900	8	25	8
S8	MP	ACE	1200	8	25	8
S9	MP	ACE	900	8	25	8
S7-L	MP	ACE	900	8	25	8

Table 3. Results of speech tests and intensity difference limens for each subject. The first two columns are the words and sentence recognition in quiet. The third columns is the SNR at which the sentence score was half that of the score in quiet. In the XIDL-J columns, the X denotes failure to reach 71% correct at adjustment level of 25%DR (at which the jitter level required was too high), and H denotes reaching 71% in one direction of adjustment only. S9 left the study before DLs on electrodes 3/20 were measured.

Subject	CNC %	CUNY %	SNR-H dB	E10 IDL %DR	E13 IDL %DR	E10/13 XIDL %DR	E10/13 XIDL-J %DR	E3 IDL %DR	E20 IDL %DR	E3/20 XIDL %DR	E3/E20 XIDL-J %DR
S1	84.7	100.0	8.4	1.3	1.5	2.4	2.5	1.1	2.5	3.8	7.1
S2	96.0	100.0	3.0	3.7	2.2	2.3	7.2	5.7	2.8	3.3	10.8
S3	91.3	99.5	6.7	5.4	5.4	11.1	27.8	4.0	4.6	5.0	22.4
S4	35.3	26.0	15.5	20.0	6.1	37.0	X	4.9	13.3	25.8	X
S5-L	24.0	14.0	17.5	13.4	17.3	56.9	X	21.8	30.6	81.6	X
S6	82.0	96.1	5.8	2.8	8.2	23.6	X	11.3	5.4	20.9	H
S5-R	31.3	23.0	17.0	10.8	18.9	60.1	X	5.5	1.7	24.5	X
S7-R	48.7	88.0	10.4	13.5	15.2	45.6	X	31.7	17.5	62.5	X
S8	52.7	82.4	8.6	4.6	3.2	18.6	X	3.6	4.5	6.1	H

S9	37.0	40.0	17.4	14.2	8.7	25.8	H	-	-	-	-
S7-L	75.3	100.0	12.4	13.9	18.9	45.1	X	13.0	37.6	51.5	X

Table 4. Linear regression of speech scores predicted by difference limen scores (n = 11).

	IDL	XIDL	XIDL/IDL
CNC words	R = - 0.52, p = 0.10	R = - 0.69, **p = 0.019**	R = -0.67, p = **0.024**
SNR-H	R = 0.58, p = 0.059	R = 0.69, **p = 0.019**	R = 0.58, p = 0.064

Figure 1.

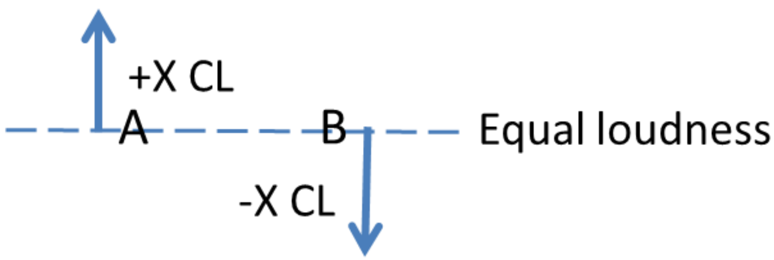
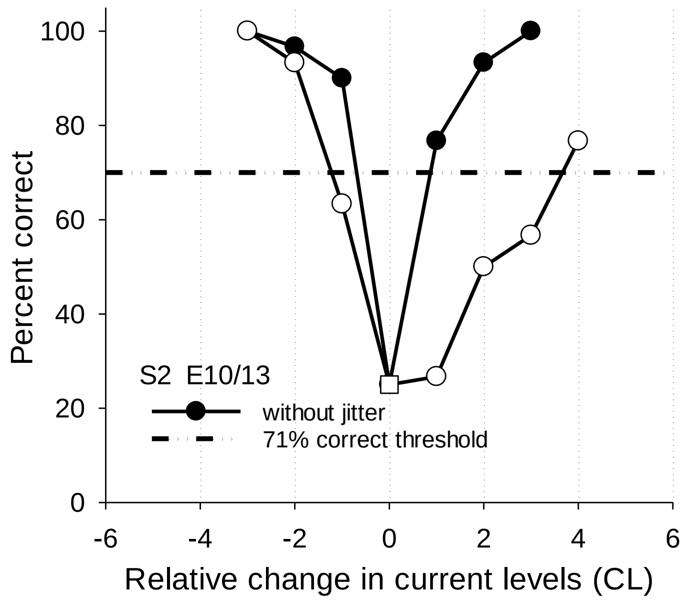
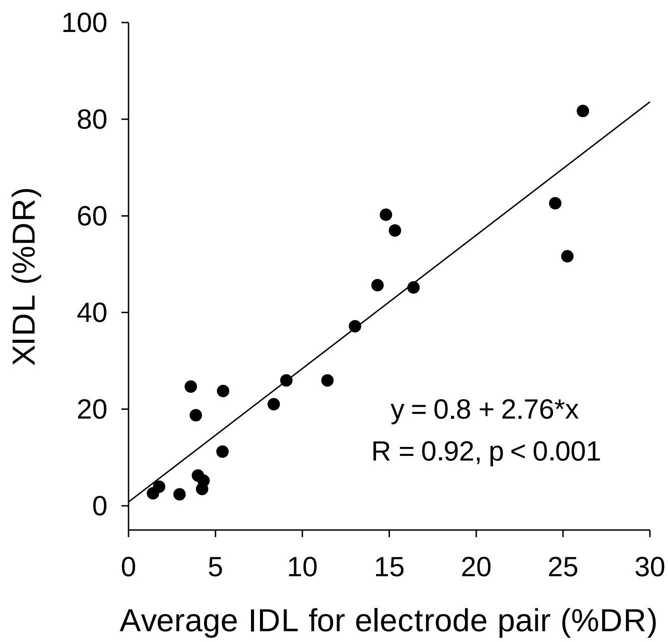


Figure 2



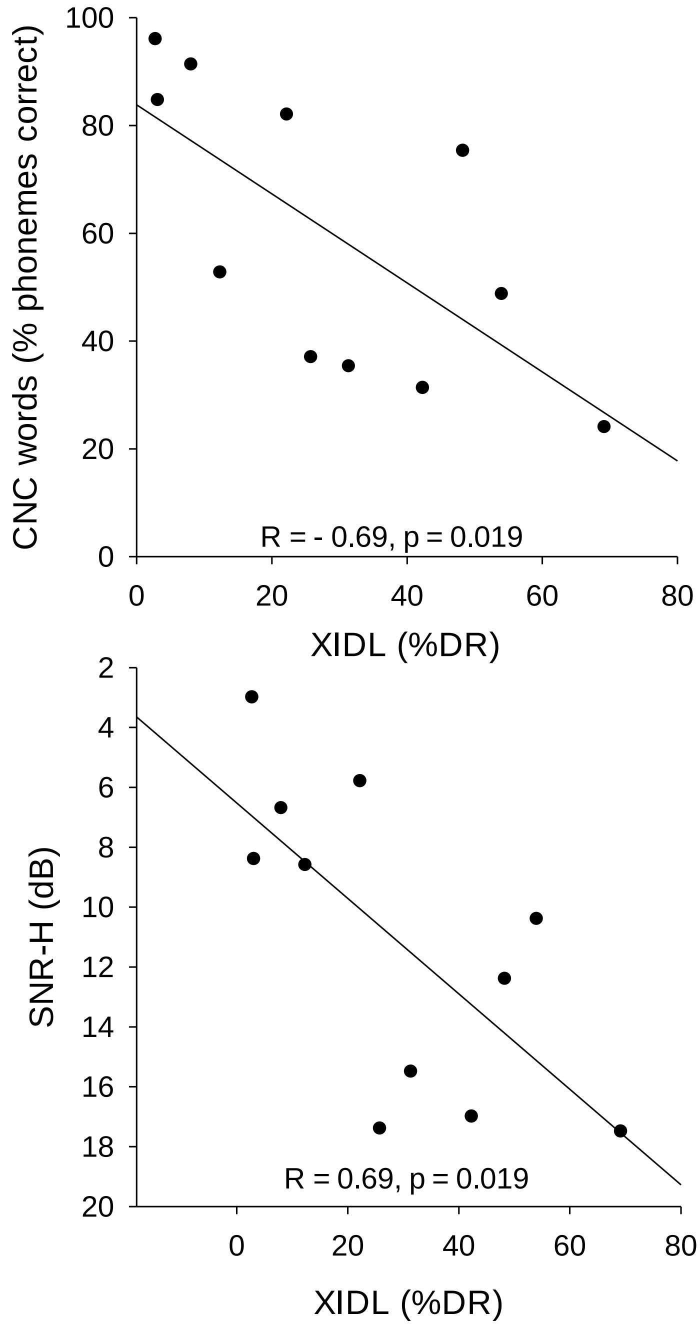
S2 E10/13
without jitter
71% correct threshold
Percent correct
Relative change in current levels (CL)
0
20
40
60
80
100
-6
-4
-2
0
2
4
6

Figure 3



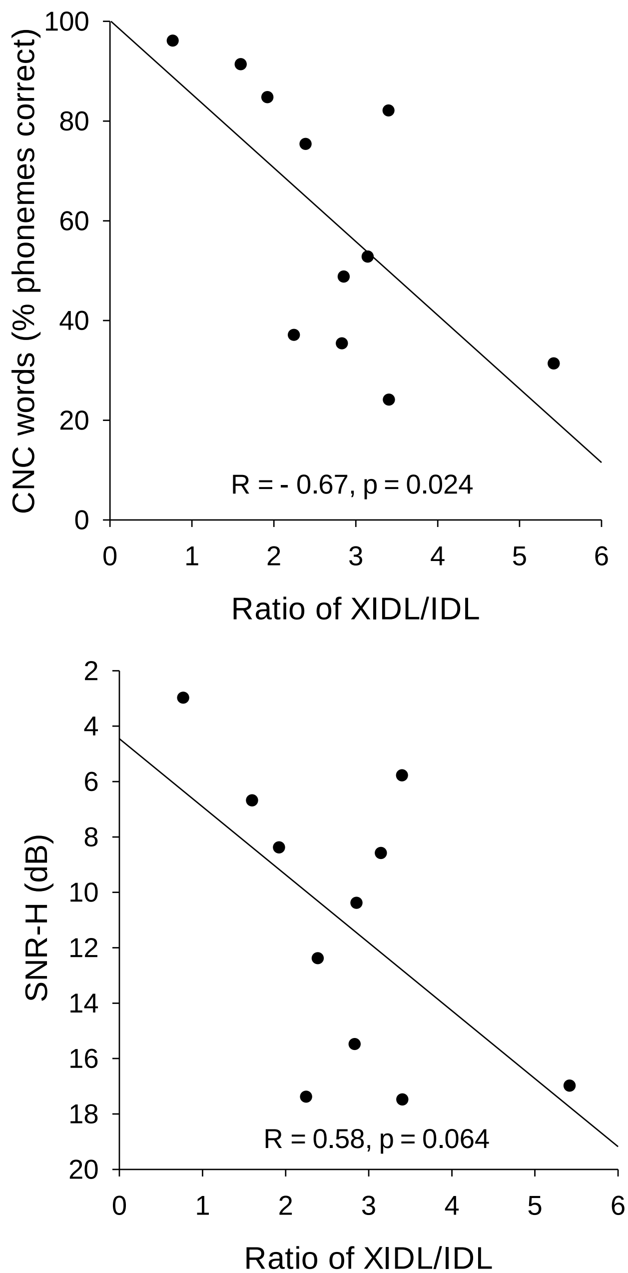
100
80
60
40
20
0
XIDL (%DR)
y = 0.8 + 2.76*x
R = 0.92, p < 0.001
0
5
10
15
20
25
30
Average IDL for electrode pair (%DR)

Figure 4



CNC words (% phonemes correct)
100
80
60
40
20
0
R = - 0.69, p = 0.019
0
20
40
60
80
XIDL (%DR)
SNR-H (dB)
2
4
6
8
10
12
14
16
18
20
R = 0.69, p = 0.019
0
20
40
60
80
XIDL (%DR)

Figure 5



CNC words (% phonemes correct)
100
80
60
40
20
0
R = - 0.67, p = 0.024
0 1 2 3 4 5 6
Ratio of XIDL/IDL
SNR-H (dB)
2
4
6
8
10
12
14
16
18
20
R = 0.58, p = 0.064
0 1 2 3 4 5 6
Ratio of XIDL/IDL