



Published in final edited form as:

Hear Res. 2008 January ; 235(1-2): 143–156.

Processing F0 with Cochlear Implants: Modulation Frequency Discrimination and Speech Intonation Recognition

Monita Chatterjee¹ and Shu-Chen Peng

Cochlear Implants and Psychophysics Laboratory, Dept. of Hearing and Speech Sciences, University of Maryland, College Park, College Park, MD 20742

Abstract

Fundamental frequency (F0) processing by cochlear implant (CI) listeners was measured using a psychophysical task and a speech intonation recognition task. Listeners' Weber fractions for modulation frequency discrimination were measured using an adaptive, 3-interval, forced-choice paradigm: stimuli were presented through a custom research interface. In the speech intonation recognition task, listeners were asked to indicate whether resynthesized bisyllabic words, when presented in the free field through the listeners' everyday speech processor, were question-like or statement-like. The resynthesized tokens were systematically manipulated to have different initial F0s to represent male vs. female voices, and different F0 contours (i.e., falling, flat, and rising). Although the CI listeners showed considerable variation in performance on both tasks, significant correlations were observed between the CI listeners' sensitivity to modulation frequency in the psychophysical task and their performance in intonation recognition. Consistent with their greater reliance on temporal cues, the CI listeners' performance in the intonation recognition task was significantly poorer with the higher initial-F0 stimuli than with the lower initial-F0 stimuli. Similar results were obtained with normal hearing listeners attending to noiseband-vocoded CI simulations with reduced spectral resolution.

Keywords

modulation frequency discrimination; speech intonation; prosody; cochlear implants; voice pitch

INTRODUCTION

Perhaps the greatest challenge for cochlear implants (CIs) today is the delivery of information about the pitch of spectrally complex signals. The spectral degradation of speech information provided by CIs removes the fine structure information necessary for the extraction of pitch. The shallow insertion depth of the electrode means that the signal is also shifted in tonotopic place. While this has not yet been shown to have a negative impact on pitch perception by CI listeners, work by Oxenham et al (2004) suggests that normal hearing listeners have difficulty processing temporal-envelope based pitch information when the carrier signal is similarly shifted to the base. Although the fundamental frequency (F0) information in everyday sounds may not be critical for speech recognition in quiet, it contributes greatly to our ability to hear and appreciate the musical quality of sounds. In addition, F0 cues are considered to be important

¹Corresponding Author: Cochlear Implants and Psychophysics Laboratory, Dept. of Hearing and Speech Sciences, 0100 LeFrak Hall, University of Maryland, College Park, MD 20742, Email: mchatterjee@hesp.umd.edu, Voice: (301) 405 7716, Fax: (301) 314 2023.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

in auditory scene analysis (Brokx and Nootboom, 1982; Assmann and Summerfield, 1990: for a review, see Darwin and Carlyon, 1995). The F0 contour in speech also provides critical cues for the processing of prosodic information, e.g. in lexical tone recognition. In spoken English, speech prosody may mark linguistic contrasts at various units such as words, phrases, or sentences; such contrasts are often referred to as speech intonation (Ladd, 1996; Lehiste, 1970, 1976). For example, varying the F0 contour (e.g., rising vs. falling) of a sentence while keeping its syntactic structure unchanged (e.g., “*The boy is on the school bus?*” vs. “*The boy is on the school bus.*”) results in a change in the perceived utterance from a question to a statement.

Although F0 information provides primary cues to listeners’ speech intonation recognition, F0 variation often takes place simultaneously with intensity and duration variations (Cooper & Sorensen, 1981; Freeman, 1982; Ladd, 1996; Lehiste, 1970, 1976). Normally hearing listeners are capable of using the F0, intensity, and duration characteristics of utterances to recognize speech intonation contrasts in a collective manner (Fry, 1955, 1958; Lehiste, 1970, 1976).

Present-day CIs are not designed to specifically deliver F0 information to the listener. An important source of F0 information available to CI listeners is derived from the periodicity cues in the time pattern of the acoustic signal, and transmitted to individual electrodes via the temporal envelope. When listeners are compelled to obtain F0 information from temporal envelope cues in spectrally reduced speech, their performance in tasks that require F0 detection deteriorates (Green et al., 2002; Fu et al., 2004; Qin and Oxenham, 2005; Stickney et al., 2004). These findings are consistent with research showing that the auditory system performs poorly in tasks involving extraction of musical or voice pitch from periodicity cues in the temporal pattern (Burns and Viemeister, 1976; McKay and Carlyon, 1999; Carlyon and Deeks, 2002). Studies by Green and colleagues (Green et al., 2005; Green et al., 2004) suggest that enhancing the temporal fluctuations of spectrally degraded speech may produce small improvements in F0 processing, but certain kinds of enhancements may result in reduced transmission of other speech features. In an interesting study, Vandali et al (2005) were able to demonstrate improved voice pitch coding without diminished speech recognition using an experimental speech processing strategy in which temporal F0 cues were provided to listeners using synchronous, multi-channel envelope information with enhanced modulation depth. Improved spectral coding has also been shown to improve F0 processing by CI listeners (Geurts and Wouters, 2004). However, such strategies are not as yet available to CI listeners. In general, CI listeners using present-day devices do not perform as well as normal hearing (NH) listeners in tasks involving speaker recognition, gender recognition, tone recognition, or intonation recognition, all of which involve F0 processing.

Studies of temporal pitch processing with cochlear implants

Several studies have explored the ability of CI listeners to process pitch conveyed by the periodicity of a pulse train or sinusoid stimulating a single electrode location (Shannon, 1983; Townshend et al., 1987; Zeng, 2002). Chen and Zeng (2004) explored the sensitivity of CI listeners to pulse-rate modulations. These studies collectively indicate that CI listeners have poor resolution for both static and dynamic pitch differences between pulse trains presented at rates higher than 300 – 500 Hz. Similar results were obtained with NH listeners attending to the pitch of unresolved harmonics (Carlyon and Deeks, 2002). While these studies are informative about the temporal pattern processing limitations of the electrically and acoustically stimulated auditory system, they are not entirely relevant to the context of present-day CIs, in which voice pitch information is presented via *temporal modulations* of a carrier pulse train. McKay and colleagues (1994, 1995) have shown that the perception of the pitch of an amplitude-modulated train of electrical pulses is multi-dimensional. Their results indicate that the pitch of a modulated pulse train is dominated by the carrier place and pulse rate at low

modulation depths, but as the modulation depth increases, the perceived pitch is influenced by the lower modulation frequency. Results of a further study by McKay and Carlyon (1999) showed that modulation frequency and carrier pulse rate evoke independent dimensions of pitch, and that the salience of the amplitude-modulated pitch increases with modulation depth. Geurts and Wouters (2001) measured CI listeners' ability to process temporal modulations at 150 and 180 Hz, and the effects of experimental speech processing strategies designed to explicitly transmit F0 information through the temporal envelope. Although they did not observe improved performance with the experimental strategy, one of the interesting findings of their study was that listeners were able to achieve better performance in discriminating between temporal patterns on multi-channel, co-modulated stimuli than on single channel stimuli.

Lexical tone and intonation recognition with cochlear implants

Experiments have shown that CI listeners are able to process F0 cues present in the speech stream to perform lexical tone recognition in tonal languages (Fu et al., 1998). Consistent with what is known about complex pitch perception, it has been shown that the greater the amount of spectral detail available to normal hearing listeners, the better their performance in tone recognition (Luo and Fu, 2006; Lin et al., 2007). Using acoustic stimuli, Wei et al. (2007) have shown correlations between measures of spectral (pure-tone frequency discrimination) and temporal processing (gap detection), and lexical tone recognition in noise by cochlear implant users. Studies of speech intonation processing in English have revealed that CI listeners are able to process intonation cues in speech to some extent (Green et al., 2004; Green et al., 2005). Although F0 is the primary cue in the recognition of both lexical tones and intonation, it is inevitably accompanied by covarying duration and intensity cues in natural utterances. Thus, the fact that CI listeners can process intonation or lexical tone information does not necessarily mean that they can detect F0 changes. In fact, Luo and Fu (2004) found that enhancing the co-varying amplitude envelope information improved tone recognition by normal-hearing listeners attending to spectrally degraded speech.

Objectives of the present study

In this paper, we report on a study of modulation frequency discrimination by CI listeners, measured over the normal range (50 to 300 Hz) of voice pitch frequencies. In addition, the participants performed a speech intonation recognition task, in which they were asked to identify a bisyllabic word ("popcorn") as question-like or statement-like. The word was resynthesized (from the original utterance) to have varying F0 contours; intensity and duration cues were varied. We were interested in the correlations, if any, between the results of psychophysical and intonation recognition tasks.

The speech intonation recognition experiments were also conducted with a small group of NH listeners, who listened to the unprocessed ("full") stimuli as well as noiseband-vocoded versions (Shannon et al., 1995). Although the focus of the present study was not on the comparison between the performance of CI and NH listeners, results obtained with the NH individuals provided a useful framework against which the CI listeners' performance could be evaluated.

METHODS

Subjects

Ten adult cochlear implant users (Nucleus-22, Nucleus-24, and Freedom devices) participated in these experiments. Eight subjects (P1 – P8) were late (post-lingually) deafened, and the remaining two subjects (E1 and E2) had early-onset deafness. All participants were native speakers of American English with one exception: subject E1 was exposed to Pennsylvania

Dutch up to school age and was immersed in American English at school. She also had undiagnosed early hearing loss and did not receive hearing aids until she was 8 years old. Based on her account, we would describe her deafness as early-onset. Table I lists some relevant information about the participants. Table II displays further information about their speech processing strategies and stimulation parameters used in the psychophysical experiments. Reliable psychophysical data were obtained in nine of the ten participants (all except P5, who had considerable difficulty with the task and limited availability for training). All 10 participated in the speech intonation experiments. Other than P5, all participants had some prior experience with psychophysical and speech perception experiments. All procedures were approved by the Institutional Review Board at the University of Maryland, College Park.

Stimuli and Procedures

A. Psychophysical Measures

i. Stimuli: Stimuli were generated by the House Ear Institute Nucleus Research Interface, or HEINRI (Robert, 2002). Stimuli were 200-ms long trains of biphasic current pulses presented at 2000 pulses/sec: this pulse rate allowed the transmission of the higher modulation frequencies without perceptual effects of sub-sampling. In pilot studies, lower carrier pulse rates were found to result in sub-sampling effects that were perceptible to the more sensitive CI listeners. All stimuli were presented to electrode 16, in bipolar or monopolar stimulation mode, using pulse phase durations indicated in Table II. The reference level of the pulse trains was fixed at 50% of the dynamic range of the unmodulated carrier, unless otherwise specified. Sinusoidal amplitude modulation was applied to the pulse trains at modulation frequencies of 50, 100, 150, 200, 250 and 300 Hz. Unless otherwise specified, sinusoidal modulation depth was fixed at 20% of the amplitude of the pulse train (in microamperes).

- a. *Dynamic range (DR)* was measured for the unmodulated carriers using the following methods: A 3-down, 1-up, 2-alternative, forced-choice task was used to measure detection threshold (microamperes), resulting in an estimate of the 79.4% correct point on the psychometric function. The mean of two repetitions was calculated. To measure the Maximum Acceptable Level, the listener was asked to raise the level of the signal by pressing the “Up Arrow” on the computer keyboard, until the signal reached a loudness that was maximally tolerable. The listener was encouraged to adjust the level by incrementing and decrementing it (by means of the Up and Down arrows) until it seemed to reach the upper limit of comfortable loudness. This procedure was repeated twice, and the mean of the two was calculated (in microamperes). The DR (in microamperes) was calculated as the linear difference between the MAL and the threshold.
- b. *Modulation frequency discrimination* was measured as follows: the reference (unmodulated) level of the carrier was fixed at 50% of the DR. The carrier pulse train was sinusoidally amplitude modulated (SAM). Unless otherwise stated, the modulation depth was fixed at 20%: at this depth, the peaks were well below the MAL, the valleys were well above detection threshold, and the stimuli sounded comfortably loud. Modulation frequency was varied in a 3-down, 1-up, 3-interval, forced choice, adaptive procedure. The listener was asked to indicate which of the three intervals was different from the other two. The modulation frequency of the signal was adaptively varied until the necessary number of reversals had been achieved. A minimum of 8 reversals was required, step size was changed after the first 4 reversals, and the mean of the last four reversals was calculated as modulation frequency discrimination threshold. The step sizes were adjusted according to the listener’s sensitivity: generally, initial and final step sizes were 4 and 2 Hz or 2 and 1 Hz at low modulation frequencies, and 8 and 4 Hz or 6 and 3 Hz at high modulation frequencies.

For listeners who were very sensitive, the smaller step sizes were used at all modulation frequencies. The total number of trials in a run was 55. If 10 reversals were achieved before 55 trials were over, the run was truncated. If the minimum of 8 reversals were not achieved within the 55 trials, the run was aborted. Correct/incorrect feedback was provided. In all cases, the order of conditions was randomized. The final result was calculated as the mean of at least two repetitions.

- c. *Modulation detection thresholds*: Modulation detection thresholds were measured for each modulation frequency using a 3-down, 1-up, 3-interval, forced-choice, adaptive procedure. In 2 of the 3 intervals, the signal was unmodulated. In the third interval (presented randomly), the signal was modulated, and the listener's task was to indicate which of the three intervals sounded different. The modulation depth was adaptively varied until the required number of reversals was reached. A minimum of 8 reversals was required, step size was changed after the first 4 reversals, and the mean of the last four reversals was calculated as modulation detection threshold. The maximum number of trials was 55. If 10 reversals were reached before 55 trials were over, the run was truncated. Correct/incorrect feedback was provided. The order of presentations was randomized. The final result was calculated as the mean of at least two repetitions.
- d. *Roving*: To examine possible effects of any loudness cues that might result from changing the modulation frequency, modulation frequency discrimination thresholds were measured (section C under Results) using methods identical to those described above, but under conditions in which the level of the carrier in the standard signal was roved within ± 0.5 dB of the reference level. This resulted in a variation of approximately 7 current levels in the device, including the reference level. This level of rove was highly detectable to the listeners (in general, JNDs for current amplitude tend to be of the order of 1 clinical unit for CI listeners at moderate loudness levels). Listeners were instructed to ignore any perceived changes in loudness and to listen for a difference in the pitch, or the tone, of the sound. Again, the order of conditions was randomized during the experiment.
- e. *Loudness-Balancing*: Loudness-balancing was used in two instances. In one case (Section C under Results), the effect of removing any loudness differences between stimuli modulated at different rates was measured in three subjects: in this case, stimuli presented to the same electrode were loudness-balanced across modulation rates. In a second case, the effect of changing the electrode location was measured in three subjects (Section D under Results). In this case, the reference (unmodulated) stimuli were loudness-balanced across electrodes. Loudness balancing was performed using a double-staircase, 2-down, 1-up, 2-interval, forced-choice, adaptive procedure. The reference was fixed at one modulation frequency (50 Hz) or one electrode (electrode 14 or 16), depending on the experiment. The listener was asked to judge which of the two stimuli (the reference or the comparison) was louder. In the descending track, when the comparison was judged louder two times in succession, its level was decreased: when the reference was judged louder once, the level of the comparison was increased. In the ascending track, when the reference was judged louder two times in succession, the level of the comparison signal was increased: when the comparison signal was judged louder once, its level was decreased. No feedback was provided. The two tracks were interleaved in random fashion. A minimum of 8 reversals was required for each track, step size was changed after the first 4 reversals, and the mean of the last 4 reversals was calculated. As in other adaptive procedures, the maximum number of trials was 55. If 10 reversals were achieved within 55 trials, the run was truncated and the mean of the last 4 reversals was calculated. At the end of the two interleaved tracks, the descending track provided

the level of the comparison at which it was judged louder than the reference 70.71% of the time, while the ascending track provided the level of the comparison at which it was judged softer than the reference 29.29% of the time.

B. Speech intonation recognition task

i. Subjects: All 10 CI participants performed the intonation recognition task. In addition, four young NH listeners also participated in this task.

ii. Stimuli: A bi-syllabic word “popcorn” which was originally produced by a female speaker as a “statement” was used for resynthesis. This token was resynthesized by varying the F0-related acoustic parameters, i.e., F0 height and F0 variation, using the Praat software (version 4.3; Boersma and Weenink, 2004). For F0 height, tokens had a flat F0 contour with initial F0s at 120 and 200 Hz for the vocalic portions of both syllables to simulate male and female speakers. Each of the F0 height-adjusted tokens was further manipulated to vary the F0 contour from initial to final F0 in nine steps (−1.00, −0.42, −0.19, 0, 0.17, 0.32, 0.58, 1.00, and 1.32 octaves). For this manipulation, a smooth, linearly rising, level, or linearly falling F0 contour was imposed on the entire word, from the beginning to the end. For the 120-Hz tokens, the target final F0s were 60, 90, 105, 120, 135, 150, 180, 240, and 300 Hz; for the 200-Hz tokens, the target final F0s were 100, 120, 150, 175, 200, 225, 300, 400, and 500 Hz. Note that while statements are associated with a falling or flat F0 contour (parameter values ≤ 0 octaves), questions are associated with a rising contour (parameter values > 0 octaves).

The resulting set of resynthesized stimuli ($N = 18$; 2 steps of F0 heights * 9 steps of F0 variations) were also varied in the intensity and duration of the second syllable to eliminate any non-F0 related acoustic cues. For the intensity rove, the second syllable was amplified or attenuated so that its peak amplitude was at a certain dB above or below the reference. The change in amplitude ranged from −10 to 10 dB in 5 dB steps; 0 dB refers to the original level without any manipulation. Similarly, the duration of the second syllable was varied relative to its reference duration. Ratios of the varied durations to the reference duration were 0.65, 1.00, 1.35, or 1.70; 1.00 refers to the original ratio without any manipulation. The entire stimulus set comprised 360 tokens (i.e., 1 bi-syllabic word * 2 F0 heights * 9 F0 variations * 5 intensity ratios * 4 duration ratios).

The NH listeners were also tested with acoustic noiseband-vocoder simulations of cochlear implant speech (Shannon et al., 1995). The simulations were created using software made available by Dr. Qian-Jie Fu (TigerSpeech Technology, House Ear Institute). The noiseband vocoders were implemented as follows: The speech signals (input frequency range 200 – 7000 Hz) were bandpass filtered into 4 and 8 frequency bands (−24 dB/octave). Corner frequencies for each analysis band were based on Greenwood’s frequency-to-place map (Greenwood, 1990). The temporal envelope from each band was extracted by half-wave rectification and lowpass filtering (cutoff at 400 Hz, −24 dB/octave). The extracted envelope from each band was used to amplitude-modulate bandlimited white noise. For each channel, the modulated noise was spectrally limited by the same bandpass filter used for frequency analysis. The output signals from all channels were summed, and the total r.m.s. level at the output was adjusted to match the r.m.s. of the unprocessed input signal.

iii. Procedures: The speech intonation recognition task was conducted in a sound-treated booth (IAC). All speech stimuli were presented via a single loudspeaker (Tannoy Reveal) at 65 dBA. All listeners were tested using their clinically assigned speech processors and settings. Each of the resynthesized speech tokens was judged by the listener as either a “statement” or a “question,” using a single-interval, 2 alternative force-choice procedure. In all experiments, the entire stimulus set (i.e., 360 resynthesized bi-syllabic words) was tested in one experimental

run, and all participants completed two runs of resynthesized stimuli. No feedback was provided during testing. Prior to testing, a small set of stimuli similar to the test stimuli was used to familiarize listeners with the intonation recognition task. The experiment was controlled by means of a Matlab-based experimental user interface.

The NH listeners performed the task under three conditions, with the unprocessed, 8-channel noiseband-vocoded, and 4-channel noiseband-vocoded stimuli. The order of the conditions was randomized across listeners.

Results of the intonation recognition experiment were analyzed by deriving cumulative d' scores from the psychometric functions (Macmillan and Creelman, 2005). The cumulative d' score obtained from each listener's psychometric function was used in subsequent analyses.

RESULTS

Psychophysical results

A. Modulation frequency discrimination—Figure 1 shows the modulation frequency discrimination (MFD) thresholds (Weber fraction expressed as a percentage) obtained in the nine CI users who participated (i.e., all except P5). Considerable variety was observed in the shape of the functions, as well as in listeners' sensitivity to modulation frequency differences. Listener P4 had higher MFD thresholds than all the others. Listeners E1 and E2 performed very differently from each other in the task: E1 had very low thresholds in this task, while E2 had higher thresholds at all modulation frequencies. The thick solid line shows the median thresholds. The median function captures the basic shape of the function: it has a bandpass-to-lowpass characteristic, with the Weber fraction increasing (sensitivity decreasing) at high modulation frequencies. At the lowest modulation frequency (50 Hz), the median Weber fraction increases slightly. While the decline in sensitivity at high modulation frequencies is consistent with the low-pass characteristic of the temporal modulation transfer function in NH and CI listeners (Viemeister, 1979; Shannon, 1990), the increase in the Weber fraction at 50 Hz may be related to the small number of modulation periods available within the 200 ms signal.

B Effects of modulation salience

i) Modulation detection thresholds: As the modulation depth was fixed at 20% in all measures of MFD above, it is possible that for some listeners with poor modulation detection thresholds (MDTs) at some modulation frequencies, the salience of the modulation was not strong enough to allow for good discrimination in the MFD task. To examine this possibility, we measured MDTs at each modulation frequency in a number of subjects. Figure 2 shows the relationship between modulation detection threshold at each modulation frequency, and the corresponding MFD thresholds at that modulation frequency, for each of the nine participants in the study. Each symbol represents a different participant. The intermittent lines show the result of a linear regression for each subject's data. The solid line shows a linear function with the median intercept and slope obtained from the linear regressions for all the subjects' data. A positive relationship exists between the MDT and MFD in most cases. Exceptions were subjects E1 and P4, who had extremely low and extremely high MFD thresholds respectively: the intercepts and slopes of the linear regression with each of their data sets, were clear outliers.

ii) Increasing Modulation Depth: While modulation detection thresholds may provide one measure of the effect of modulation salience on the results, a more direct approach would be to measure the effect of increasing modulation depth on MFDs. Figure 3 shows results obtained with four of the participants (E1, E2, P2 and P8) in conditions of increasing modulation depth,

from 5% to 30%, with increments of 5%. The maximum modulation depth was limited to 30% in this experiment in order to ensure that the stimuli did not become uncomfortably loud.

The experiment was conducted with a 0.5 dB rove (see below) on the reference signal. In Figure 3, the abscissa shows the modulation depth in terms of percent modulation above modulation detection threshold. The Weber fraction is shown on the ordinate for reference modulation frequencies of 50, 150, 250 and 300 Hz. A log scale was chosen for these plots to accommodate the wide range spanned by the data across subjects. If modulation salience played an important role in MFD, we would expect these functions to decline with increasing modulation depth. However, overall, convincing effects were not observed. A one-way repeated-measures analysis of variance (ANOVA) was performed on the pooled data at each modulation frequency tested. Results showed no significant effects of modulation depth on MFD Weber fractions at any of the modulation frequencies.

C. Effects of loudness and intensity-roving—As the reference modulation frequency is increased, the number of peaks and valleys presented to the listener within the 200-ms signal also increases. It is possible, therefore, that just as the loudness of a low-rate pulse train increases with pulse rate, the loudness of an amplitude-modulated pulse train also increases with modulation frequency, and that subjects use loudness cues rather than modulation frequency cues to perform the MFD task. To investigate the effects of possible loudness increases, two sets of measurements were made: i) stimuli were loudness balanced across modulation frequency and ii) effects of roving the intensity on the MFD thresholds.

a. Loudness balancing: The stimuli were loudness-balanced using a double staircase adaptive procedure as described under Methods. The reference was always the signal, amplitude modulated at 50 Hz. The modulation depth was fixed at 20%. The mean of at two repetitions was calculated. Only listeners P2, P4 and P8 were available to participate in these experiments. If loudness increased with increasing modulation frequency, the functions should show a consistent decline (indicating the need for lower current levels to balance the 50 Hz reference modulation frequency). Consistent changes in loudness were not observed with increases in the modulation frequency. To facilitate data analysis, each loudness-balanced amplitude was converted into percent difference from the reference amplitude at 50 Hz. A one-way repeated-measures ANOVA was conducted to test for differences in the loudness-balanced levels due to modulation frequency. As Mauchly's test indicated that the sphericity assumption was violated, a Greenhouse-Geisser adjustment was used. The analysis did not show a significant effect of modulation frequency on the loudness-balanced levels ($F[1.88, 3.77] = 2.733$; $p = 0.184$).

b. Roving: As the loudness-balancing results showed relatively small changes in loudness with modulation frequency, measurements were repeated in listeners E1, E2, P1, P2, P4, P7 and P8 with a small rove applied to the reference signal, as described under Methods. The Weber fraction for MFD was measured at 50, 100, 150, 200, 250 and 300 Hz. As observed in Fig. 1, the task became very difficult for some subjects with increasing modulation frequency, and results could not be obtained with all subjects at all modulation frequencies: therefore, statistical analysis was performed on the results obtained at each modulation frequency, separately. A one-way, repeated-measures ANOVA revealed no significant effects of roving on MFD Weber fractions at any of the reference modulation frequencies tested ($F[1,6] = 0.044$ at 50 Hz; $F[1,6] = 0.95$ at 100 Hz; $F[1,5] = 0.713$ at 150 Hz; $F[1,5] = 3.625$ at 200 Hz; $F[1,4] = 0.233$ at 250 Hz; $F[1,4] = 0.024$ at 300 Hz; $p > 0.05$ in all cases). All subjects indicated that the roving was very noticeable, and some noted that it made the task more difficult. These results suggest that loudness differences did not play an important role in modulation frequency discrimination within the range of parameters used in our study.

D. Effects of cochlear location—All measures described above were obtained on a single, apical electrode. It is of considerable interest to examine the dependence of MFD on cochlear place. Across-place variability has been shown in different psychophysical measures in CI listeners (Pfungst et al., 2007, 2004). Therefore, modulation frequency discrimination was measured at four electrode locations in listeners P2, P4 and P8. Unmodulated carriers on electrodes 8, 10, 14 and 16 were loudness-balanced [the reference electrode was either electrode 14 or electrode 16, at 50% of its dynamic range]. Note that electrodes are numbered 1–22 from base to apex; therefore, electrodes 8, 10 and 14 are 8, 6 and 2 electrodes apart from electrode 16 in the basal direction. Measurements of the MFD function were made as described previously at the loudness-balanced level on each channel. With listeners P2 and P8, the reference stimulus was roved ± 0.5 dB; listener P4 found the tasks so difficult that roving was not applied, and even without roving, MFD thresholds could not be measured above 50 Hz for electrodes 10 and 8. Results are shown in Figure 4: each panel shows results obtained with a different participant. The results suggest that the electrode place does have an effect on modulation frequency discrimination thresholds. A 2-way ANOVA conducted on the results obtained with P2 showed significant effects of modulation frequency ($F[5,24] = 20.447$), $p < 0.001$) and electrode location ($F[3,24] = 6.035$, $p < 0.005$), but no significant interactions between the two. Post hoc Bonferroni tests indicated that the Weber fraction obtained at the 300 Hz modulation frequency was significantly different from all others, and that the Weber fractions obtained on electrode 16 was significantly larger than the results obtained on electrodes 10 and 8. Analyses conducted with results obtained with P8 showed significant effects of frequency ($F[5,20] = 60.858$), $p < 0.001$) and electrode location ($F[3, 20] = 62.389$, $p < 0.001$), as well as a significant interaction between the two ($F[11, 20] = 18.602$, $p < 0.001$). Post hoc Bonferroni tests showed that thresholds obtained on electrodes 14 and 16 were significantly lower than results obtained on other electrodes, but not from each other. A one-way ANOVA conducted on results obtained with listener P4 showed significant effects of electrode location ($F[3,4] = 57.432$, $p < 0.005$). Post hoc Bonferroni tests indicated that thresholds obtained with electrodes 16 and 14 were significantly different from (lower than) results obtained with electrodes 10 and 8, but not from each other. Thus, results with all three listeners showed significant effects of electrode place.

Speech intonation recognition results

Psychometric functions obtained with the 10 CI listeners in the speech intonation recognition task are shown in Fig. 5A. Results represent overall performance averaged over the 120-Hz and 200-Hz initial F0 heights. The thick solid line shows the mean results obtained with the CI listeners. The psychometric functions show a monotonic shape, with bias and slope varying across participants.

Figure 5B shows the mean psychometric functions obtained with the CI listeners, alongside those obtained with the NH listeners attending to the three conditions (full speech, 8-channel noiseband-vocoded speech, and 4-channel noiseband-vocoded speech). The NH listeners' psychometric function obtained with the full speech stimuli was near-ideal, with almost no bias and a steep slope crossing the "category boundary" at the 0-octave F0 ratio. However, the NH listeners performed more poorly with the 8- and 4-channel noiseband-vocoded stimuli: it is apparent that reduced spectral resolution results in shallower slopes in the psychometric function.

For further analyses, the "proportion question" scores were converted into Z-scores and the total cumulative d' was calculated from each listener's psychometric function. All analyses described below were conducted using the derived total cumulative d' scores.

To investigate the relevance of results obtained with the re-synthesized "popcorn" stimuli to the listeners' everyday listening experience, results obtained in the intonation recognition task

with naturally produced sentences (developed by Peng, 2005) were compared with the cumulative d' scores obtained above. The stimuli were comprised of 120 sentences produced by 3 male and 3 female talkers in two contrasting intonation styles: statement (“*The girl is on the playground.*”) and question (“*The girl is on the playground?*”). Performance by the CI listeners in this task was highly correlated with performance measured with the bisyllabic “popcorn” stimuli ($r = 0.861$, $p < 0.01$). Note that the naturally produced sentences carried co-varying intensity and duration cues, while these cues were systematically roved in the “popcorn” stimuli. The significant correlation observed between the two sets of results validates the relevance of the “popcorn” stimuli to realistic listening, and also confirms the dominance of the F0 cue in intonation recognition.

Relation between modulation frequency discrimination and intonation recognition measures

The different panels of Figure 6 show the correlations between the psychophysical results obtained with different modulation frequency ranges and the cumulative d' calculated from the speech intonation data. The correlations were not substantially different when analyzed separately according to initial F0-height: hence, those analyses are not shown here. Psychophysical MFD thresholds obtained at 50 Hz did not correlate significantly with the intonation results, perhaps because modulation frequency discrimination tended to be easier at 50 Hz and the participants all performed relatively well at this modulation frequency. Similarly, MFD thresholds above 200 Hz did not correlate significantly with the intonation results: these thresholds were very high for most CI listeners. However, MFD results averaged between 50 and 100 Hz (Fig. 6A), 100 and 200 Hz (Fig 6B), and 50, 100, 150 and 200 Hz (Fig 6C), were all significantly correlated with the intonation results. In addition, MFD results obtained at 100 Hz were also significantly correlated with the intonation results (Fig. 6D). In all cases, an exponential function fit the data better than a linear function. In part, this could have been due to the results obtained with listener P4, who performed relatively poorly in the psychophysical task at modulation frequencies above 50 Hz. To explore the validity of the correlations observed in Fig. 6, we repeated the analysis without including results from subject P4. The correlation observed in Fig. 6A with results averaged over 50–100 Hz was no longer significant as a result of this exclusion (Fig. 7A). However, the other correlations remained significant (Figs. 7B–D) and once again, the exponential fit provided a stronger correlation than the linear fit.

Relation between performance in intonation recognition, modulation frequency discrimination, and other tasks

It is possible that the correlations between the modulation frequency discrimination thresholds and the intonation recognition measure were trivial. That is, those who are better at one task may simply be better at any other task. To investigate this possibility, the participants’ performance in the intonation recognition task was compared with their performance in vowel and consonant recognition tasks (Table I). No significant correlations were observed between results obtained in these different tasks. This was true for correlations including the two participants with early-onset deafness as well as for correlations excluding these two participants. However, the vowel and consonant recognition scores were significantly correlated with each other ($r = 0.816$; $p < 0.01$). The cumulative d' obtained in the intonation task was also not significantly correlated with the measures of modulation sensitivity (MDTs shown in Fig. 2) obtained in the same listeners. Additionally, no significant correlation was observed between the modulation frequency discrimination measures (Weber fractions) and vowel and consonant recognition scores obtained in these listeners (both including and excluding E1 and E2). However, if results obtained with subject P4 were excluded from the analyses (his mean Weber fraction was 2.18 s.d.s from the overall mean calculated across subjects and could be considered an outlier), significant correlations were observed between the Weber fraction (mean calculated between 50 and 200 Hz) in the modulation frequency

discrimination task and both vowel ($r = 0.913$; $p < 0.05$) and consonant ($r = 0.834$; $p < 0.05$) recognition. We note that in the case of the correlations observed with performance in the intonation recognition task, the correlations were *stronger* when subject P4 was included. Thus, underlying factors determining these relations are likely to be different in the two cases.

Effects of F0 height, and comparisons with NH listeners' data

Results were also analyzed according to F0 height. Figure 8A shows the results of this analysis in CI listeners: performance was clearly poorer with the 200-Hz initial F0 stimuli than with the 120-Hz initial F0 stimuli. This was verified by a paired t-test, which showed a significant difference between the two conditions ($p < 0.05$). Based on the psychophysical results, we would expect to see poorer performance with the stimuli with a higher starting F0 if CI listeners were relying primarily on temporal envelope cues. As anticipated, CI listeners rely more on temporal information in this task. However, individual results varied considerably (Fig. 8B) across the CI listeners.

Although the number of NH listeners in this study was too small to allow for extensive analyses, some comparisons were made between results obtained with the NH and the CI subjects. The mean overall cumulative d' was lower for CI subjects (Mean = 3.25, s.d. = 1.53) than for the NH subjects (Mean = 4.67, s.d. = 0.93): however, the difference between the two groups did not reach statistical significance. In contrast to results obtained with the CI listeners, results obtained with the four NH listeners showed no significant differences between the two initial-F0 conditions in a paired t-test ($p = 0.20$). However, the same NH listeners' results did show a significant difference ($p < 0.01$ in a paired t-test) between the two conditions with the 4-channel noiseband-vocoded speech. Further, the difference was in the same direction as that obtained with the CI listeners: i.e., poorer performance (lower cumulative d') was obtained with the 200 Hz than for the 120 Hz initial-F0 stimuli. The 8-channel noiseband-vocoded condition did not produce significant differences between the two initial F0-conditions ($p = 0.85$) in a paired t-test.

DISCUSSION

The results presented here provide a fairly comprehensive view of CI listeners' ability to utilize temporal envelope patterns in the normal voice pitch range. Significant correlations were observed between psychophysical sensitivity to modulation frequency changes and performance in the speech intonation recognition task. Another important finding was that CI listeners' performance in the intonation recognition task with the lower (120 Hz) initial-F0 stimuli was significantly poorer than their performance in the same task with the higher (200 Hz) initial-F0.

The psychophysical results are of some interest in themselves. Consistent with our expectations, listeners' sensitivity to differences in modulation frequency tended to follow a low-pass filter shaped function, worsening with increasing modulation frequency. The two listeners with early-onset deafness had results that were very different from each other, with listener E1 being one of the most sensitive of participants and listener E2 being one of the least sensitive. It should be noted, however, that while E2 was most likely congenitally deafened, E1 had a more progressive hearing loss. However, she did not begin to use hearing aids until she was 8 years of age. The post-lingually deafened participants also showed considerable variation in sensitivity.

The lack of a strong effect of modulation depth (salience) was surprising. Although loudness considerations did not allow us to increase modulation depth beyond 30%, the decrease of modulation depth down to 5% did not appear to have a significant negative impact on performance. It is possible that increasing the depth would have revealed some effects. Based

on McKay et al.'s (1995) and McKay and Carlyon's (1999) findings that the pitch of an amplitude-modulated carrier follows the modulator more and more strongly as the depth of the signal increases, we had initially hypothesized that the MFD thresholds would be strongly dependent on the depth of modulation. Thus, these results may appear to be inconsistent with expectations based on McKay and colleagues' findings. However, one important difference between our experiments and those of McKay and colleagues is that in the present study, participants were not specifically asked to judge or compare the pitch of the sounds: instead, they picked the one that was different from the other two in the trial (they were instructed to listen for a change in pitch or sound quality, and to ignore loudness changes). It is possible that a specifically pitch-based task, as was adopted by McKay and colleagues, would have shown greater dependence on modulation depth. However, the majority of participants in our study would have had considerable difficulty with such a task: in our experience, even some of the most sensitive cochlear listeners seem to have trouble with pitch judgments. This does not necessarily mean that these CI listeners cannot use F0 cues: they may have learned to use other F0-related percepts (sound quality changes like roughness) to perform tasks such as intonation recognition. The lack of a significant correlation between MDTs and performance in the intonation detection task, also suggests that modulation salience is not critical for listeners to be able to utilize information in the temporal pattern effectively.

Significant place-of-excitation effects were observed in each of the subset of three CI listeners who were able to participate in that experiment. In everyday listening through the device, CI listeners obtain temporal envelope cues for voice pitch on all channels that would respond to the voiced portion of speech. Thus, temporal pitch perception by CI listeners must be a multi-channel process. It is unclear how the information from individual channels contributes to the perception of complex temporal pitch through multiple channels: i.e., do the most sensitive channels dominate? Geurts and Wouters (2001) suggest that multiple channels may "cooperate" to enhance sensitivity to F0. Vandali et al. (2005) observed improvements in CI listeners' ability to rank the pitch of sung vowels (without accompanying decrements in speech perception) using a strategy that explicitly provided temporal modulations derived from the speech envelope synchronously across all channels. These results suggest that at least in some cases, multi-channel F0 information may be useful in transmitting complex pitch cues. It is not known how across-channel modulation *interference* effects may influence temporal pitch perception through co-modulated multi-channel stimuli: it has been demonstrated that CI listeners have difficulty in modulation detection tasks in the presence of competing fluctuations on other channels (Richardson et al., 1998; Chatterjee, 2003; Chatterjee and Oba, 2004).

The strong correlation of the MFD Weber fractions with performance in the intonation recognition task is perhaps not surprising at first glance: indeed, it is well known that CI listeners must rely heavily on temporal pattern processing for complex pitch perception. However, various factors may have contributed to increased variance in the results. For instance, the carrier rate used in the psychophysical experiments was 2000 pps, higher than that used by any of the listeners' speech processors. In addition, three of the listeners (P3, P4 and P5) used the N-22 device, with carrier rates of 250 pps per channel. One of these listeners (P3) performed very well in the intonation recognition task. This listener had also participated in pilot studies using carrier rate as a variable. At low carrier rates and high modulation frequencies, he was able to detect modulation frequency changes (presumably by detecting differences in the sub-sampled temporal patterns), even at modulation rates well below half the carrier rate. It is possible that he has learned to utilize such cues, which are likely to be present in his everyday listening. On the other hand, it is also true that the harmonic structure of complex sounds provides some spectral cues in the spectral envelope. Some CI listeners may have been able to utilize such cues in the intonation recognition task. Further, the participants used different speech processing strategies and had different volume and sensitivity settings. Also, the stimulation mode applied was similar to, but not identical to, each

individual's speech processor. Given these and other sources of variation that were introduced by the use of the relatively uncontrolled stimuli delivered through the listener's own speech processor, the strong correlations observed in this study indicate that temporal pattern processing is indeed important for CI listeners attending to voice pitch differences. The fact that exponential functions described the relationship better than linear functions suggests that a minimum level of sensitivity to the temporal pattern is necessary, but not sufficient, to achieve excellent performance in the intonation recognition task.

Consistent with the notion that temporal envelope cues are critical for the CI listeners' performance in the intonation recognition task, we found that performance with the "male" F0 stimuli (i.e. with the initial F0 at 120 Hz) was significantly better than performance with the "female" F0 stimuli (i.e. with the initial F0 at 200 Hz). These findings were consistent with the previous findings by Green et al. (2002, 2004), showing declining usefulness of the temporally-based pitch cue at high F0 frequencies. The poorer modulation frequency discrimination thresholds at higher reference modulation frequencies also support this conclusion. We may therefore infer that spectral cues did not play an important role in this task for most of the CI listeners. The small group of NH listeners included in this study provided results that supported this conclusion: when listening to 4-channel noiseband-vocoded stimuli, their performance was similar to that of the CI listeners', with significantly poorer performance with the 200 Hz initial-F0 intonation contour. However, there was no significant effect of initial F0 height on listeners' intonation recognition with greater spectral resolution. Taken together, these results indicate the dominance of temporal patterns for pitch perception in spectrally reduced speech.

These findings in the speech intonation results did not, however, relate directly to the psychophysical findings. For instance, MFD thresholds obtained at the lower modulation rates were not observed to exhibit stronger correlations with intonation recognition performance with the lower initial-F0 stimuli. This lack of a direct correspondence points to the limitations inherent in such comparisons; the many sources of variability involved in experiments with CI listeners with acoustic stimuli through their speech processor, might have obscured observable correlations.

CONCLUSION

The present study examined CI users' Weber fractions for modulation frequency discrimination on an apical channel, and its relation to their speech intonation recognition using F0 cues. The primary finding was that the Weber fraction for modulation frequency discrimination was strongly correlated with CI listeners' ability to identify speech tokens as questions or statements using F0 information. The psychophysical measures were robust to changes in modulation depth and did not appear to be greatly influenced by loudness differences between the stimuli; however, strong place-of-stimulation effects were observed in a subset of three CI listeners. These individuals had greater difficulty recognizing the intonation with the higher initial-F0 stimuli than with the lower initial-F0 stimuli. This pattern was mirrored in results obtained in the group of four NH listeners attending to spectrally reduced speech. The results of this study indicate that those CI listeners who are more sensitive to temporal patterns tend to be better able to make efficient use of the F0 cues available through their speech processor.

Acknowledgements

We thank the cochlear implant listeners who participated in this study for their ongoing support of our research. We are very grateful to Dr. Qian-Jie Fu for the use of his software for processing the stimuli and for controlling the phoneme recognition tasks. We thank Kelly Hoffard for her help with some of the data collection. Mark E. Robert and Shubhasish B. Kundu are thanked for software support. We thank two anonymous reviewers for helpful comments on an the manuscript. This work was supported by NIH/NIDCD grant nos. R01 DC 04786 (PI: MC) and P30-DC004664 (PI: Robert Dooling).

References

- Assmann PF, Summerfield Q. Modeling the perception of concurrent vowels: Vowels with different fundamental frequencies. *J Acoust Soc Am* 1990;88:680–697. [PubMed: 2212292]
- Brox JPL, Nootboom SG. Intonation and the perceptual separation of simultaneous voices. *Phonetics* 1982;10:23–36.
- Burns EM, Viemeister NF. Nonspectral pitch. *J Acoust Soc Am* 1976;60:863–869.
- Carlyon RP, Deeks JM. Limitations on rate discrimination. *J Acoust Soc Am* 2002;112(3 Pt 1):1009–1025. [PubMed: 12243150]
- Chatterjee M. Modulation masking in cochlear implant listeners: envelope versus tonotopic components. *J Acoust Soc Am* 2003;113(4):2042–2053. [PubMed: 12703715]
- Chatterjee M, Oba SI. Across- and within-channel envelope interactions in cochlear implant listeners. *JARO* 2004;5(4):360–375. [PubMed: 15675001]
- Chen H, Zeng FG. Frequency modulation detection in cochlear implant subjects. *J Acoust Soc Am* 2004;116(4):2269–2277. [PubMed: 15532658]
- Cooper, WE.; Sorensen, JM. *Fundamental Frequency in Sentence Production*. New York: Springer-Verlag; 1981.
- Darwin, CJ.; Carlyon, RP. Auditory Grouping. In: Moore, BCJ., editor. *Hearing*. Academic Press; San Diego, CA: 1995.
- Freeman, FJ. Prosody in perception, production, and pathologies. In: Yoder, DE., editor. *Speech, Language, and Hearing: Pathologies of Speech and Language*. 2. W. B. Saunders; Philadelphia, PA: 1982.
- Fry DB. Duration and intensity as physical correlates of linguistic stress. *The Journal of the Acoustical Society of America* 1955;27:765–768.
- Fry DB. Experiments in the perception of stress. *Language and Speech* 1958;1:126–152.
- Fu QJ, Chinchilla S, Galvin JJ. The role of spectral and temporal cues in voice gender discrimination by normal-hearing listeners and cochlear implant users. *J Assoc Res Otolaryngol* 2004;5(3):253–60. [PubMed: 15492884]
- Fu QJ, Zeng FG, Shannon RV. Importance of tonal envelope cues in Chinese speech recognition. *J Acoust Soc Am* 1998;104(1):505–510. [PubMed: 9670541]
- Green T, Faulkner A, Rosen S, Macherey O. Enhancement of temporal periodicity cues in cochlear implants: effects on prosodic perception and vowel identification. *J Acoust Soc Am* 2005;118(1):375–385. [PubMed: 16119358]
- Green T, Faulkner A, Rosen S. Enhancing temporal cues to voice pitch in continuous interleaved sampling cochlear implants. *J Acoust Soc Am* 2004;116(4 Pt 1):2298–2310. [PubMed: 15532661]
- Green T, Faulkner A, Rosen S. Spectral and temporal cues to pitch in noise-excited vocoder simulations of continuous-interleaved-sampling cochlear implants. *J Acoust Soc Am* 2002;112:2155–2164. [PubMed: 12430827]
- Guerts L, Wouters J. Better place-coding of the fundamental frequency in cochlear implants. *J Acoust Soc Am* 2004;115(2):844–852. [PubMed: 15000196]
- Guerts L, Wouters J. Coding of the fundamental frequency in continuous interleaved sampling processors for cochlear implants. *J Acoust Soc Am* 2001;109:713–726. [PubMed: 11248975]
- Ladd, DR. *Intonational Phonology*. Cambridge University Press; Cambridge: 1996.
- Lehiste, I. *Suprasegmentals*. MIT Press; Cambridge, MA: 1970.
- Lehiste, I. Suprasegmental features of speech. In: Lass, NJ., editor. *Contemporary Issues in Experimental Phonetics*. Academic Press; New York: 1976.
- Lin YS, Lee FP, Huang IS, Peng SC. Continuous improvement in Mandarin lexical tone perception as the number of channels increased: a simulation study of cochlear implant. *Acta Otolaryngol* 2007;127(5):505–514. [PubMed: 17453477]
- Luo X, Fu QJ. Contribution of low-frequency acoustic information to Chinese speech recognition in cochlear implant simulation. *J Acoust Soc Am* 2006;120(4):2260–2266. [PubMed: 17069321]
- Luo X, Fu QJ. Enhancing Chinese tone recognition by manipulating amplitude envelope: implications for cochlear implants. *J Acoust Soc Am* 2004;116(6):3659–3667. [PubMed: 15658716]

- Macmillan, NA.; Creelman, CD. *Detection Theory. A User's Guide. 2.* Lawrence Erlbaum Associates; Mahwah, NJ: 2005.
- McKay CM, Carlyon RP. Dual temporal pitch percepts from acoustic and electric amplitude-modulated pulse trains. *J Acoust Soc Am* 1999;105(1):347–357. [PubMed: 9921661]
- McKay CM, McDermott HJ, Clark CM. Pitch-matching of amplitude-modulated current pulse trains by cochlear implantees: the effect of modulation depth. *J Acoust Soc Am* 1985;97(3):1777–1785. [PubMed: 7699159]
- McKay CM, McDermott HJ, Clark CM. Pitch percepts associated with amplitude-modulated current pulse trains in cochlear implantees. *J Acoust Soc Am* 1994;96(5 Pt 1):2664–2673. [PubMed: 7983272]
- Oxenham AJ, Bernstein JG, Penagos H. Correct tonotopic representation is necessary for complex pitch perception. *Proc Natl Acad Sci USA* 2004;101(5):1421–1425. [PubMed: 14718671]
- Peng, SC. Unpublished Dissertation. University of Iowa; Iowa City, IA: 2005. Perception and Production of Speech Intonation in Pediatric Cochlear Implant Recipients and Children with Normal Hearing.
- Pfingst BE, Xu L, Thompson CS. Effects of carrier pulse rate and stimulation site on modulation detection by subjects with cochlear implants. *J Acoust Soc Am* 2007;121(4):2236–2246. [PubMed: 17471737]
- Pfingst BE, Xu L, Thompson CS. Across-site threshold variations in cochlear implants: relation to speech recognition. *Audiol Neurootol* 2004;9(6):341–352. [PubMed: 15467287]
- Qin MK, Oxenham AJ. Effects of envelope-vocoder processing on F0 discrimination and concurrent-vowel identification. *Ear and Hearing* 2005;26(5):451–460. [PubMed: 16230895]
- Robert, ME. *House Ear Institute Nucleus Research Interface User's Guide.* House Ear Institute; Los Angeles: 2002.
- Shannon RV. Multichannel electrical stimulation of the auditory nerve in man. I. Basic psychophysics. *Hear Res* 1983;11:157–189. [PubMed: 6619003]
- Shannon RV, Zeng FG, Kamath V, Wygonski J, Ekelid M. Speech recognition with primarily temporal cues. *Science* 1995;270(5234):303–304. [PubMed: 7569981]
- Stickney GS, Zeng FG, Litovsky R, Assmann P. Cochlear implant speech recognition with speech maskers. *J Acoust Soc Am* 2004;116(2):1081–1091. [PubMed: 15376674]
- Vandali AE, Sucher C, Tsang DJ, McKay CM, Chew JWD, McDermott HJ. Pitch ranking ability of cochlear implant recipients: A comparison of sound-processing strategies. *J Acoust Soc Am* 2005;117(5):3126–3138. [PubMed: 15957780]
- Townshend B, Cotter N, Van Compernelle D, White RL. Pitch perception by cochlear implant subjects. *J Acoust Soc Am* 1987;82:106–115. [PubMed: 3624633]
- Wei C, Cao K, Jin X, Chen X, Zeng FG. Psychophysical performance and Mandarin tone recognition in noise by cochlear implant users. *Ear Hear* 2007;28:62S–65S. [PubMed: 17496650]
- Zeng FG. Temporal pitch in electric hearing. *Hearing Res* 2002;174:101–106.

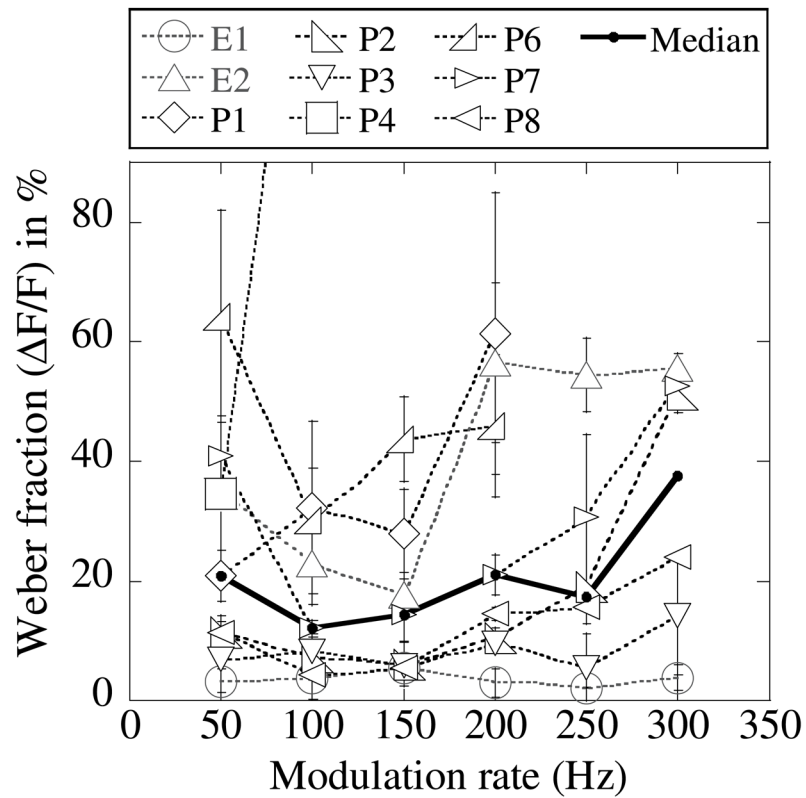


Fig. 1.

Weber fractions (%) plotted as a function of reference modulation frequency. Each symbol corresponds to a different subject. Results for listener P4 are not shown above the 50 Hz reference. Error bars represent ± 1 s.d. from the mean. The thick solid line shows the median values of the results.

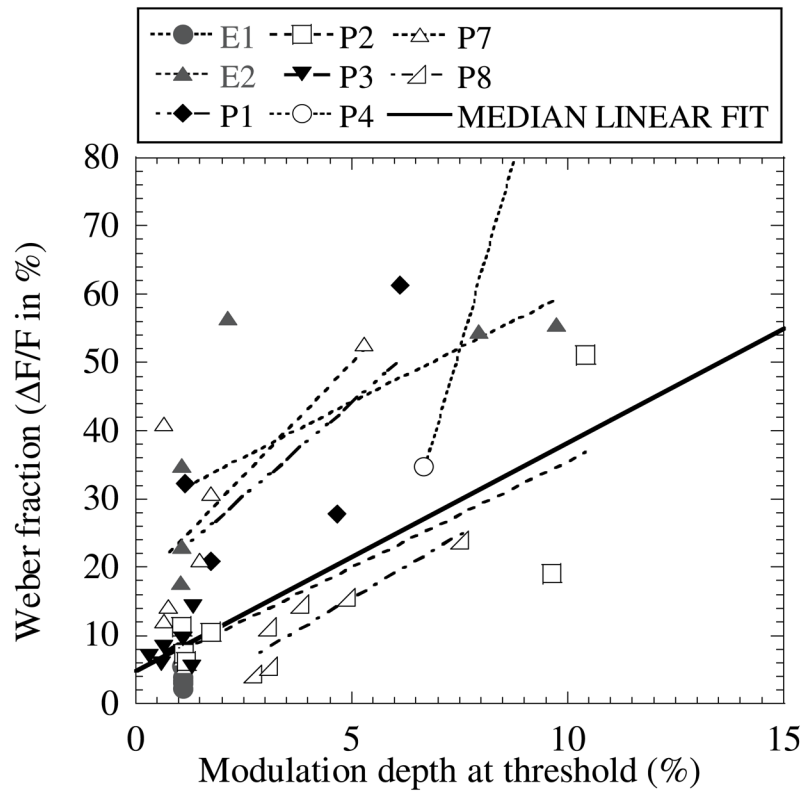


Fig. 2. Weber fractions (%) obtained in the MFD experiment vs. threshold modulation depth (%). Each symbol shows data obtained with an individual subject; each data point corresponds to results obtained at a different modulation frequency. The intermittent lines show the linear regression to each individual subject's data. A line with the mean intercept and slope is also indicated (solid line). Data for listener P4 were truncated.

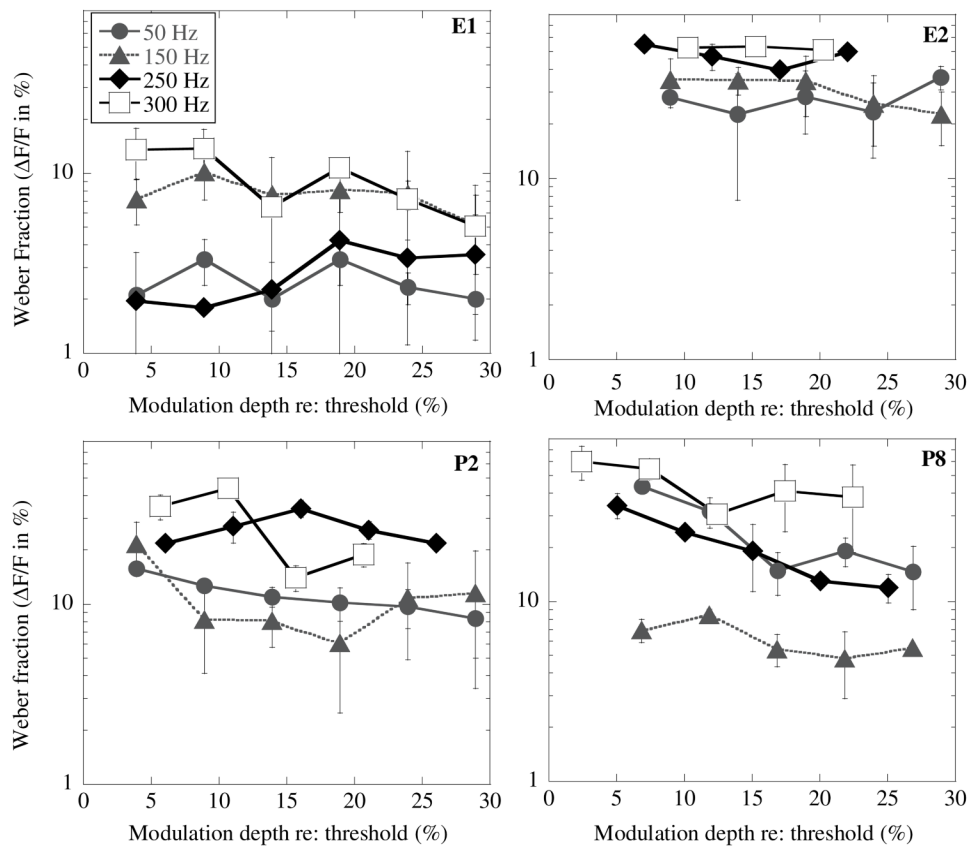


Fig. 3. Weber fractions (%) obtained in the MFD task as a function of modulation salience (the difference between modulation depth and modulation threshold). Each panel shows results obtained with a different subject. Symbols correspond to the different reference modulation frequencies (see inset in top left-hand panel).

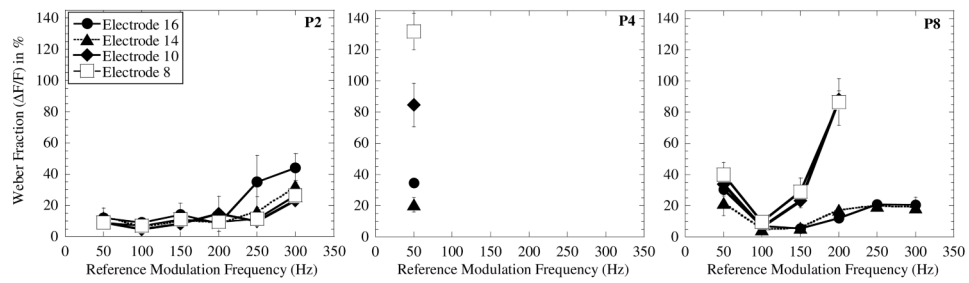


Fig. 4. Weber fractions (%) obtained in the MFD task as a function of reference modulation frequency. The parameter is electrode location (see inset). Each panel shows results obtained with a different subject. Error bars show ± 1 s.d. from the mean.

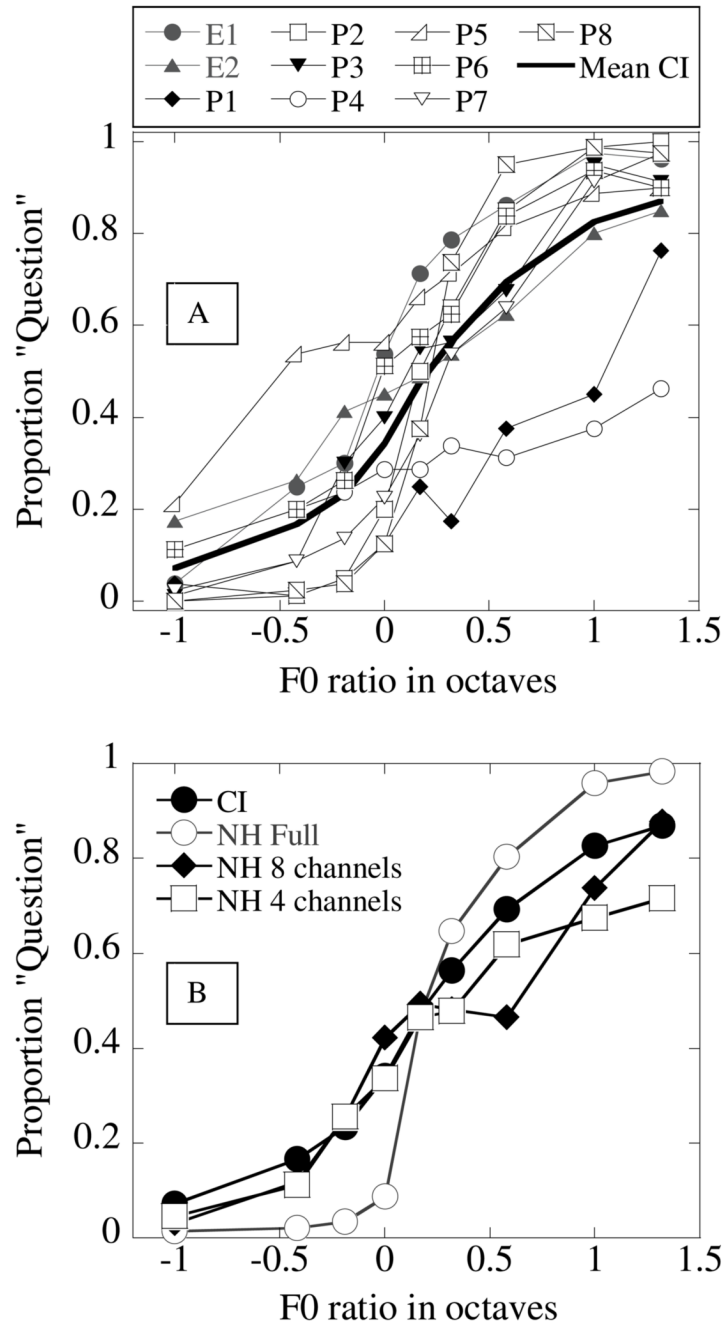


Fig. 5. Fig. 5A. Psychometric functions (proportion “question” vs. F0 ratio in octaves) obtained in the speech intonation recognition task with the 10 CI listeners. The different symbols show results obtained with individual subjects. The thick solid line shows the mean function. Fig. 5B. Mean psychometric functions obtained in the speech intonation task with CI (filled circles) and NH listeners attending to full (open circles), 8-channel (filled diamonds), and 4-channel (open squares) noiseband-vocoded speech, respectively.

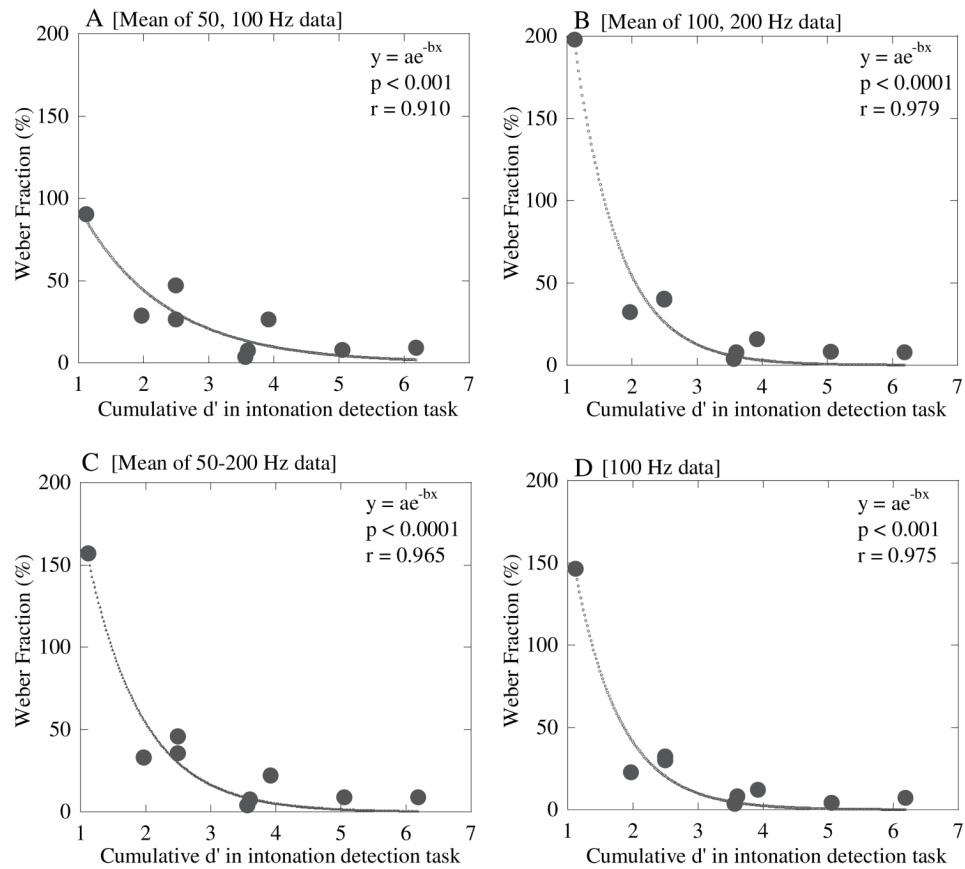


Fig. 6. Scatterplots showing the relation between the Weber fraction calculated in different ways and the cumulative d' obtained from the speech intonation results in CI listeners. The line represents the best-fitting exponential function to the data.

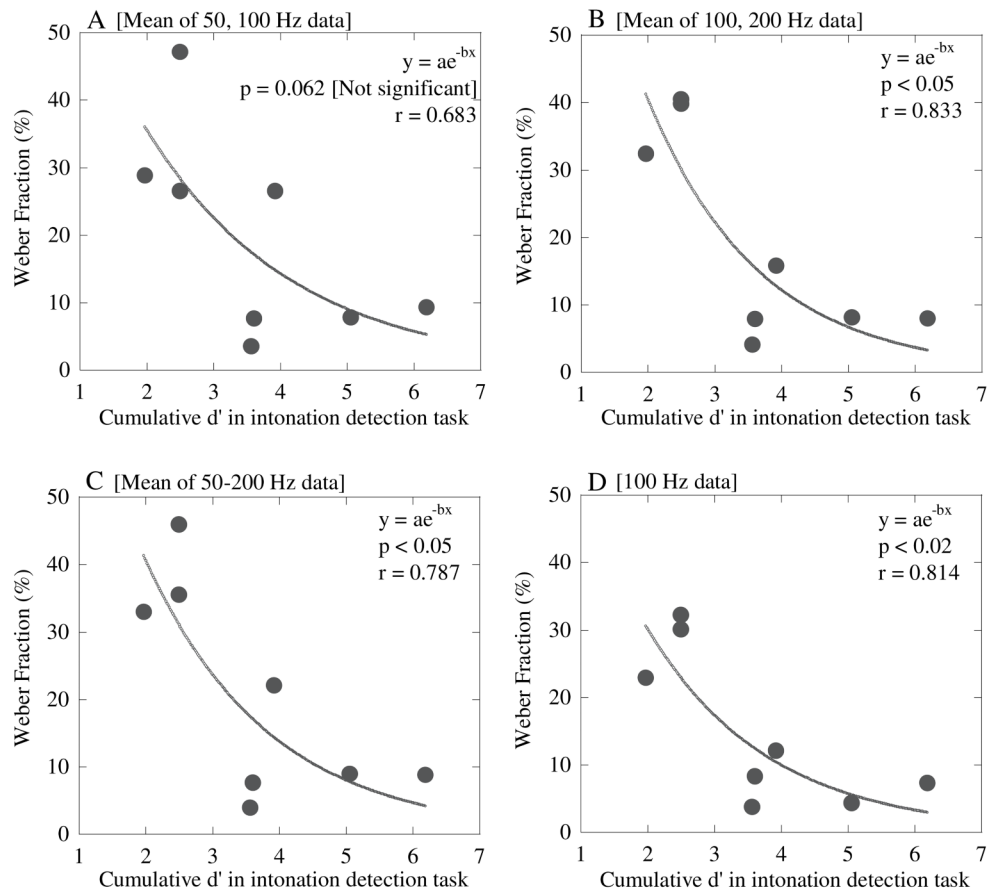


Fig. 7. Scatterplots as in Fig. 6, but excluding results obtained with P4.

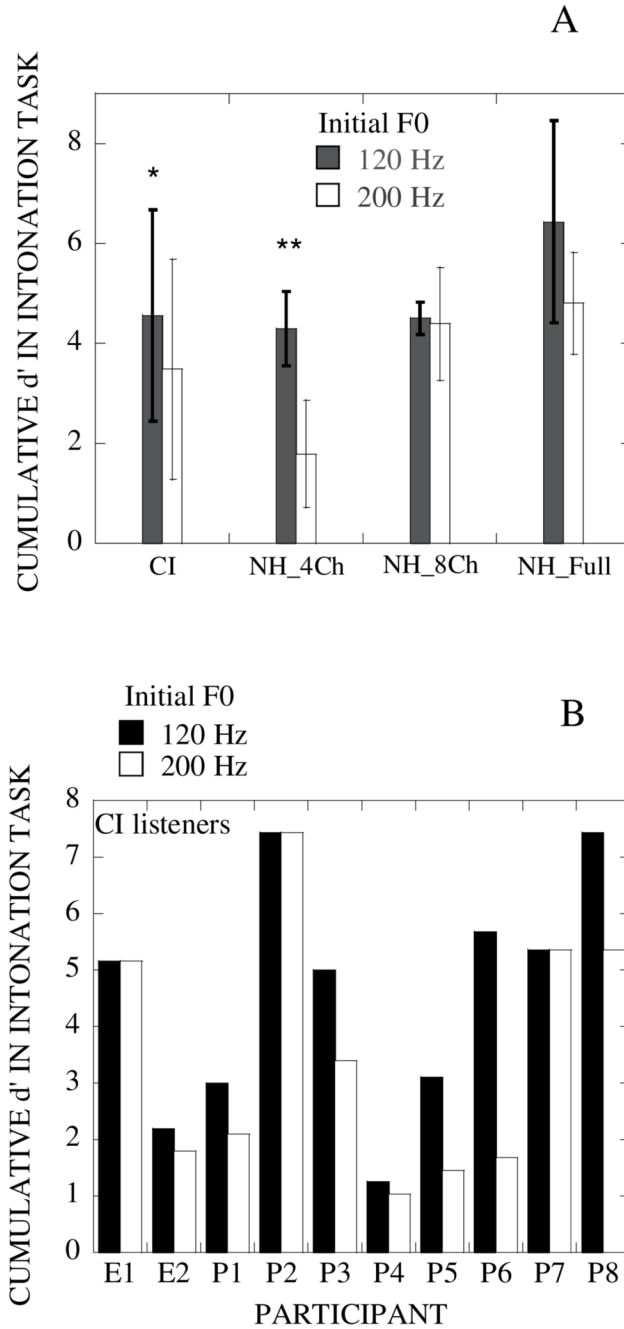


Fig. 8. Fig. 8A. Barplot showing the cumulative d' calculated from results with the 120-Hz and the 200-Hz initial-F0 stimuli. Results are shown for CI listeners, and NH listeners attending to 4-channel and 8-channel noiseband-vocoded speech (NH_4Ch and NH_8Ch) and unprocessed speech (NH_Full), respectively. Single asterisks indicate a significant difference between the two sets of results at the 0.05 level: double asterisks indicate significance at the 0.01 level. Fig. 8B. Barplots showing cumulative d' values obtained with the ten individual CI listeners for the 120-Hz and 200-Hz initial F0 stimuli (filled and patterned bars, respectively).

TABLE 1

Relevant information about the CI participants in the study. Participants with code numbers beginning with E and P had early-onset and late-onset (post-lingual) deafness, respectively. Vowel tests were performed using items from the Hillenbrand et al. (1995) vowel set (12 vowels, 5 male and 5 female talkers) and consonant tests were performed using items recorded by Shannon et al. (1999) (20 consonants, 5 male and 5 female speakers). Vowel and consonant recognition tests were controlled by means of software made available by Dr. Qian-Jie Fu (TigerSpeech Technology, House Ear Institute). Stimuli were presented by a single loudspeaker in a sound-treated booth at 65 dB SPL.

CI Participant	Etiology/type of Hearing Loss	Age at testing	Age at implantation	Gender	Device	Vowel recognition (%)	Consonant recognition (%)
E1	Possibly genetic	56	47	F	N24	25.83	36.50
E2	Congenital	23	16	M	N24	57.50	39.50
P1	Unknown	60	57	M	N24	51.67	64.00
P2	Genetic	63	60	F	N24	85.00	92.50
P3	Trauma	49	35	M	N22	85.00	73.00
P4	Unknown	65	51	M	N22	70.00	75.00
P5	Genetic	70	57	F	N22	70.83	74.00
P6	Unknown	83	78	M	N24	46.67	57.00
P7	German measles and possible ototoxicity	67	66	F	FREEDOM	61.67	64.50
P8	Possibly genetic	64	63	F	FREEDOM	67.50	84.00

Table II Selected speech processor parameters and stimulation parameters for psychophysical experiments.

CI Participant	SPEECH PROCESSOR PARAMETERS		STIMULATION PARAMETERS FOR PSYCHOPHYSICS	
	Processing strategy	Pulse rate (pulses/sec)	Stimulation mode	Pulse phase duration-Interphase gap(µs)
E1	ACE	720	MPI+2	40-6
E2	SPEAK	250	MPI+2	40-6
P1	ACE	900	MPI+2	40-6
P2	ACE	1200	MPI+2	40-6
P3	SPEAK	250	BP+1	100-20
P4	SPEAK	250	BP+1	100-20
P5	SPEAK	250	CG	
P6	ACE	900	MPI+2	40-8
P7	ACE	1800	MPI+2	25-8
P8	ACE	1800	MPI+2	25-8