Published in final edited form as:

J Am Acad Audiol. 2015 June ; 26(6): 572–581. doi:10.3766/jaaa.14067.

Discrimination of Stochastic Frequency Modulation by Cochlear Implant Users

Stanley Sheft*, Min-Yu Cheng*, and Valeriy Shafiro*

*Department of Communication Disorders and Sciences, Rush University Medical Center, Chicago, IL

Abstract

Background—Past work has shown that low-rate frequency modulation (FM) may help preserve signal coherence, aid segmentation at word and syllable boundaries, and benefit speech intelligibility in the presence of a masker.

Purpose—This study evaluated whether difficulties in speech perception by cochlear implant (CI) users relate to a deficit in the ability to discriminate among stochastic low-rate patterns of FM.

Research Design—This is a correlational study assessing the association between the ability to discriminate stochastic patterns of low-rate FM and the intelligibility of speech in noise.

Study Sample—Thirteen postlingually deafened adult CI users participated in this study.

Data Collection and Analysis—Using modulators derived from 5-Hz lowpass noise applied to a 1-kHz carrier, thresholds were measured in terms of frequency excursion both in quiet and with a speech-babble masker present, stimulus duration, and signal-to-noise ratio in the presence of a speech-babble masker. Speech perception ability was assessed in the presence of the same speech-babble masker. Relationships were evaluated with Pearson product–moment correlation analysis with correction for family-wise error, and commonality analysis to determine the unique and common contributions across psychoacoustic variables to the association with speech ability.

Results—Significant correlations were obtained between masked speech intelligibility and three metrics of FM discrimination involving either signal-to-noise ratio or stimulus duration, with shared variance among the three measures accounting for much of the effect. Compared to past results from young normal-hearing adults and older adults with either normal hearing or a mild-to-moderate hearing loss, mean FM discrimination thresholds obtained from CI users were higher in all conditions.

Conclusions—The ability to process the pattern of frequency excursions of stochastic FM may, in part, have a common basis with speech perception in noise. Discrimination of differences in the temporally distributed place coding of the stimulus could serve as this common basis for CI users.

Corresponding author: Stanley Sheft, Department of Communication Disorders and Sciences, Rush University Medical Center, Chicago, IL 60612; Stanley_Sheft@rush.edu.

Portions of this article were presented at the American Auditory Society Meeting, Scottsdale, AZ, March 2010.

Keywords

cochlear implant; frequency modulation; speech perception

INTRODUCTION

Speech recognition ability among cochlear implant (CI) users can vary significantly both in quiet and in noise (Firszt et al, 2004; Won et al, 2010). Understanding of the basis for this variation in speech ability is useful not only in terms of device design and settings, but also for development of tests to monitor and enhance rehabilitation progress. With processing through a limited number of spectral channels which transduce temporal variations of only the channel envelope, CIs reduce both the spectral and temporal cues of speech. Differences in the abilities to utilize these diminished cues most likely contribute to the variation in speech recognition performance. Past work with CI users has shown significant relationships between speech perception and psychoacoustic measures of both spectral and temporal resolution (Busby and Clark, 1999; Fu, 2002; Won et al, 2007; Saoji et al, 2009). The studies of spectral processing by Won et al (2007) and Saoji et al (2009) assessed sensitivity in terms of either the depth or rate of modulation of the spectral envelope. In this work, the spectral differences between stimuli of a given experimental trial were, on average, constant over the stimulus duration. Consequently, the detection or discrimination cues were well summarized by differences between the long-term spectra of the stimuli. Speech, on the other hand, is characterized by dynamic spectral variations, with change over time in the spectral content of the signal, a key aspect that conveys linguistic information. Working with Schroeder-phase harmonic complexes that exhibit periodic modulation of instantaneous frequency at a rate equal to the fundamental frequency of the complex, Drennan et al (2008) and Won et al (2010) investigated the ability of CI users to discriminate between upward or downward gliding patterns of dynamic spectral variation. Results showed significant relationships to speech perception at modulation rates of 50 and 200 Hz.

Along with rapid spectral variations in the range evaluated in the Schroeder-phase studies, speech shows prominent slow spectral variations which can be expressed as low-rate frequency modulation (FM). Similar to the envelope spectrum, the FM spectrum of speech is bandpass with greatest amplitude for rates of less than 5 Hz (Sheft et al, 2012). Though generally poorer than performance obtained when temporal envelope cues are present, the intelligibility of speech processed to retain primarily cues related to stimulus FM is significantly above chance (Sheft et al, 2008; Ardoint and Lorenzi, 2010). In terms of specific involvement of low-rate modulation in speech processing, FM information may help preserve signal coherence (Dorman et al, 1975) while also enhancing segmentation of syllables and words (Spitzer et al, 2007). Furthermore, manipulation of the natural intonation contour, that is, the low-rate FM of the voicing fundamental, has been shown to have a detrimental effect on speech recognition in noise (Binns and Culling, 2007; Miller et al, 2010).

FM pattern discrimination must underlie any robust contribution of FM processing to speech perception. Past studies of low-rate FM discrimination by CI users have used either prosodic

contours or exemplars of Chinese lexical tones as stimuli. Green et al (2005) and Chatterjee and Peng (2008) evaluated the ability of CI users to distinguish a rising from falling intonation contour. The contours were associated with sentences that were either questions or declarative statements. Though representing a pattern of FM, recognition does not necessarily require processing of the modulation pattern per se, with detection of the direction of a primarily monotonic pitch change indicating the correct response. For CI users, periodicity cues from the temporal envelope of speech can convey information regarding the voicing fundamental. In the work of Chatterjee and Peng (2008), significant correlation between the Weber fractions for rate discrimination of amplitude modulation and intonation recognition were interpreted to indicate an association between temporally based pitch perception and intonation contour discrimination. Studies of discrimination of Chinese lexical tones by CI users have also used monotonic rising and falling intonation patterns with the addition of both a flat contour and ones with a midpoint inflection (Ciocca et al, 2002, Fu et al, 2004; Wei et al, 2007). Results from Fu et al (2004) showed significant association between recognition ability for these four intonation patterns and Chinese sentence recognition. Wei et al (2007) evaluated Mandarin tone recognition by CI users in the presence of a noise masker. Along with a progressive improvement in recognition performance with increasing signal-to-noise ratio (SNR), results indicated significant correlations between tone recognition in noise and both gap detection and frequency discrimination abilities. These associations were taken to suggest the use of multiple acoustic cues in the recognition task.

To study FM discrimination in acoustic hearing (i.e., not CI users) for complex patterns that go beyond frequency glides or contours with a single inflection, Sheft et al (2012) evaluated whether difficulties in speech perception by older listeners relate to a deficit in the ability to discriminate stochastic—that is random or uncertain —FM patterns that fluctuate at a slow average rate. The approach incorporated two procedural aspects intended to affect the relationship of psychoacoustic result to speech ability. The first was the introduction of stimulus uncertainty through the use of samples of 5-Hz lowpass noise as the frequency modulators. Collins et al (1994) suggested that by incorporating aspects of natural speech, use of stimulus uncertainty in psychoacoustic paradigms increases predictive power regarding listener performance in speech tasks. The second aspect related to control of cues so that discrimination was reliant on the temporal pattern of frequency fluctuation rather than the differences in frequency excursion or modulation rate typically assessed in FM discrimination studies. The basis for this approach came from the work of Remez et al (1994) that, as an alternative to the detection of discrete cues as a basis for speech perception, emphasized the importance of the pattern of the temporal modulations of speech. In the study of Sheft et al (2012), thresholds were measured in terms of frequency excursion, stimulus duration, and SNR in the presence of a speech-babble masker. Significant differences between young and older adult listeners were obtained in most measures, with results showing significant correlation between FM pattern-discrimination thresholds measured in terms of the SNR and masked speech intelligibility.

Using the procedures of Sheft et al (2012), this study examined the relationship between the ability to discriminate among stochastic patterns of low-rate FM and speech perception for CI users. Since no previous work has evaluated discrimination ability of CI users for

complex patterns of FM, the five separate conditions of Sheft et al (2012) were used. Across conditions, FM discrimination ability was estimated in terms of frequency excursion, stimulus duration, or SNR in the presence of a speech-babble masker. As a threshold metric, frequency excursion provides a measure of sensitivity in terms of the magnitude of pattern deviation from the carrier frequency. This metric also allows for consideration of performance in relationship to past studies of pitch discrimination. Speech contains longer durations of modulation associated with prosody and syllabic rate as well as more rapid changes such as formant transitions. Discrimination thresholds in terms of stimulus duration may suggest a limiting factor in modulation perception by CI users. Speech perception in noise is a dominant problem that confronts CI users. Results obtained with the SNR threshold metric offer an evaluation of stochastic FM discrimination ability in a similar situation.

In all conditions of the study, the noise modulators were drawn from a common sampling distribution with the same maximum frequency excursion, duration, and carrier frequency for the contrasting stimuli of a given discrimination trial. These constraints were used so that discrimination could rely on only the temporal pattern of frequency fluctuation. Therefore, this work contrasts with past studies of modulation processing by CI users both by using complex patterns and not allowing for discrimination based on change in modulation statistics such as rate or depth of excursion, or direction of pitch change as in past studies of speech prosody.

METHOD

Participants

Thirteen CI users (seven females) were tested. Background information for the participants is given in Table 1. All participants were postlingually deafened native speakers of American English who had at least one year of implant experience, and showed no signs of decreased mental status, scoring at least 28 on the Mini-Mental State Index (Folstein et al, 1975). Among participants, discrimination in quiet of 200 consonant-nucleus-consonant words ranged from 33% to 91% correct (Table 1), which is within the performance range commonly reported in studies of open-set word recognition by CI users (Firszt et al, 2004; Won et al, 2010). The experimental protocol was approved by the Institutional Review Board of the Rush University Medical Center.

Psychoacoustic stimuli and procedure

Five conditions measured thresholds for discriminating between 1-kHz pure tones frequency modulated by low-pass noise. A consequence of the modulation is that the instantaneous frequency of the stimulus follows the amplitude pattern of the noise modulator. In all conditions, modulator bandwidth was fixed at 5 Hz, resulting in an average fluctuation rate of roughly 4 Hz, similar to the syllabic rate of speech. Conditions measured the ability to discriminate stimuli based on differences in the temporal pattern of frequency fluctuation about the 1-kHz carrier frequency rather than variations in long-term or peak stimulus characteristics. To maintain a constant maximum frequency excursion (ΔF), modulator peak magnitude was the same for all stimuli of a given discrimination trial. Two methods were

used in stimulus generation. In the first, the contrasting modulators of the 1-kHz carrier of each discrimination trial were independent samples drawn from the same underlying noise distribution so that they shared common average noise statistics (Fig. 1). This method was used in the first three FM conditions (Delta F, Masked Delta F, and SNR) described below. The second method paired a single random sample with its time reversal scaled by 21.0 following the cumulative integration of time-domain FM. This approach leads to a time reversal of the instantaneous frequency function so that the long-term amplitude spectra of the stimuli of a given trial are identical. To avoid cuing by differences in amplitude spectra at brief durations, the second method was used in the fourth and fifth FM conditions (Duration Down and Duration Up) which estimated thresholds in terms of stimulus duration.

The ability to discriminate stochastic FM was measured with a cued two-interval forced-choice adaptive procedure with 500 msec between intervals. On each trial, a randomly selected noise modulator was used as the cue modulator of the 1-kHz carrier. With random selection, one of the two observation intervals that followed the cue was identical to the cue; the other interval was generated with a contrasting modulator. The listener's task was to indicate which observation interval differed from the cue with correct-response feedback provided after each trial. Across the five conditions, thresholds were estimated in terms of three independent variables, F , SNR, and duration. Using a two-down one-up tracking rule that converged on 70.7% (Levitt, 1971), threshold estimates were based on the average of the last 8 of the required minimum of 12 tracking reversals with run length always at least 40 trials. With geometric step sizes used in the tracking of both F and duration (see below for additional details), reversal averages for these conditions were geometric means. In the SNR condition, threshold was estimated as the arithmetic mean of reversal values in dB. Concerned with clinical feasibility of the psychoacoustic measures, all data from each listener were collected in a single session with one threshold estimate per condition. Results thus represent neither trained nor asymptotic performance levels. For practice, listeners initially completed the first condition without scoring. Each subsequent scored run began with eight practice trials.

There were five FM conditions with data collected from each listener in the following order. Full details are given in Sheft et al (2012) with the conditions briefly summarized here. In the first condition (Delta F), the maximum frequency excursion of the FM tones was varied to determine the minimum F needed to discriminate 500 msec stochastic FM patterns. Beginning at 750 Hz in the adaptive procedure, F was varied geometrically to estimate threshold. F was initially changed by a factor of 1.46 that was reduced to 1.10 following the second tracking reversal. The second condition (Masked Delta F) repeated the first condition in the presence of a continuous four-talker speech-babble masker at an SNR of 16 dB. Using the same speech-babble masker and with F fixed at 750 Hz, the level of the 500 msec FM tones was varied in the third condition (SNR) to determine the SNR needed to discriminate the pattern of frequency fluctuation. Initial tracking step size was 6 dB which was reduced to 2 dB following the second reversal.

The fourth condition (Duration Down) estimated the shortest duration needed to discriminate FM pattern. Conducted in quiet without a masker present, the durations of all stimuli on any given trial were the same with this common duration adaptively varied

between trials so that thresholds were measured in terms of duration. The condition thus did not assess duration discrimination, a task which requires a duration difference between the stimuli that are to be discriminated, but rather FM pattern discrimination with duration used as the threshold metric. In the Duration Down condition, F was fixed at 750 Hz, and the stimulus duration of the FM tones on the first trial was 500 msec. Between trials, duration was initially changed by a factor of 1.46 with factor size reduced to 1.10 following the second tracking reversal.

At briefer durations, listeners do not follow the frequency fluctuations of the stimuli, but rather perceive stimuli globally in terms of overall characteristics such as timber. While trial-by-trial feedback helps participants to adapt listening strategy, cue transition can still affect performance. To better estimate performance limits with duration as the independent variable, the fifth condition (Duration Up) repeated the fourth, but began the tracking procedure with stimulus duration at 11 msec. Though discrimination cues can change with duration in both the Duration Down and Duration Up conditions, again stimulus duration per se cannot cue discrimination with stimuli on each trial having the same duration.

All testing was conducted in a double-walled soundproof booth. Stimuli were played through a single loudspeaker (Lifeline Amplification System, model LLF-50) positioned at 0° azimuth 1 m in front of participants who listened only through their CI set at a comfortable level with their other ear plugged. In the two Delta F conditions, the level of the modulated tone was 65 dB sound pressure level (SPL) measured at head level. In the two Duration conditions, the level of longer stimuli was 65 dB SPL, with adjustment to maintain the energy level at that of a 100-msec stimulus at 65 dB SPL once duration tracked below 100 msec. In the SNR condition, masker level was 65 dBA with FM tone level varied to estimate threshold. In all FM conditions, stimuli were shaped with 50 msec cosinusoidal rise/fall ramps, unless stimulus duration was less than 100 msec, in which case the rise/fall time was equal to half the stimulus duration.

Speech Measures

Speech perception was evaluated in terms of sentence recognition in the presence of a four-talker speech-babble masker. Four lists of ten sentences from the Bamford-Kowal-Bench Speech-in-Noise test (BKB-SIN; Etymic Research, 2005) were used. The sentences, approximately at a first-grade reading level, were originally constructed from language samples taken from young hearing-impaired children (Bench et al, 1979). Across each list, the SNR decreased in 3-dB steps from 21 to 26 dB. Based on the number of the 124 key words (31 words from 4 lists) correctly repeated, results were converted to the metric SNR-50, the estimated SNR needed for 50% correct response. As in the psychoacoustic conditions, speech testing was conducted using a single loudspeaker in a double-walled soundproof booth. The presentation level of the target sentences was 65 dBA measured at the position of the participant's head, with the level of the speech-babble masker varying with SNR.

Analysis

Psychoacoustic results from all but the SNR condition were logarithmically transformed before data analysis. Comparisons between conditions were conducted with repeated-measures *t*-tests. Relationships between psychoacoustic and speech performance were evaluated with Pearson product-moment correlation analysis. With no past work indicating a beneficial effect of hearing loss on either FM or speech-in-noise abilities, one-tailed significance testing was used. The resampling method of Troendle (1995) was used to control for family-wise error with multiple hypothesis testing in correlation analysis. In the iterative procedure, the data matrix was repeatedly shuffled with the minimum *p* value saved from each random ordering. Following 10,000 iterations, the proportion of times that these saved values were less than the *p* value of the original data were used as the *p* value adjusted for multiple comparisons. For psychoacoustic measures that showed significant pairwise correlation to BKB-SIN SNR-50 thresholds, the combined prediction by these variables of speech thresholds was assessed with multiple linear regression. Finally, commonality analysis (Nimon, 2010) was used to determine the unique and common contributions among the psychoacoustic variables in accounting for the relationship to speech performance.

RESULTS

Results from the psychoacoustic testing are shown as box plots in Figure 2. In general, there was a wide performance range across participants in all conditions. For the two Delta F conditions (left panel), in quiet individual thresholds varied by a factor of nearly 18 from under 30 Hz to almost 480 Hz, with the addition of the speech-babble masker at an SNR of 16 dB shifting the range to roughly 100–650 Hz. Average threshold as assessed by geometric mean was 125.4 Hz in quiet and 265.1 Hz in the presence of the masker. Significance of the masking effect was confirmed by a paired-samples *t*-test on logarithmically transformed delta F thresholds [$t_{(12)} = -2.70$, $p = 0.019$, $d = 0.99$]. In the FM SNR condition (middle panel), thresholds spanned a range of almost 16 dB with a mean value of 2.36 dB. For the two Duration conditions (right panel), thresholds varied by a factor of roughly 10 and 14 in the Duration Down and Duration Up conditions, respectively. Though the two Duration conditions were identical except for starting stimulus duration, the geometric mean threshold was much larger in the Duration Down (91.9 msec) than Duration Up (21.5 msec) condition. A paired-samples *t*-test on logarithmically transformed thresholds showed this difference significant [$t_{(12)} = 5.76$, $p < .001$, $d = 1.64$].

Speech-in-babble thresholds from BKB-SIN testing ranged from 6.75 to 20 dB (Table 1), with the average threshold 14.70 dB. The same babble masker was used in the FM SNR condition as in the speech testing, with both assessing a threshold value in dB. Possible utility of FM cues to benefit performance in the speech condition requires that FM discrimination thresholds are at worst no poorer than the speech thresholds. The average FM SNR threshold was 12.36 dB lower than the average BKB-SIN SNR-50 threshold, with a paired-samples *t*-test confirming the difference as significant [$t_{(12)} = 11.11$, $p < 0.001$, $d = 2.61$]. It is important to keep in mind that this difference on its own does not demonstrate involvement of FM discrimination ability in the speech task. Furthermore, the comparison between speech and FM SNR thresholds reflects not only difference in signal type, but also

in procedure with BKB-SIN testing using an open-set response format in contrast to the two-interval forced-choice paradigm of the psychoacoustic protocol.

To evaluate relationships between psychoacoustic and BKB-SIN performance, results were submitted to Pearson product-moment correlation analysis with a resampling method used to control family-wise error. Between speech results and the five FM measures, there were three significant correlations. Correlations between the SNR-50 thresholds and the FM SNR, Duration Down, and Duration Up results were 0.66 ($p = 0.025$), 0.68 ($p = 0.028$) and 0.58 ($p = 0.05$), respectively. Scatterplots of the relationships are shown in Figure 3. In multiple regression estimating SNR-50 thresholds, use of these three measures led to a significant prediction model ($F_{(3,12)} = 41.45$, $p = 0.028$) with an R^2 value of 0.618. Despite differences in threshold metric, processing of stochastic FM across the three psychoacoustic conditions must have some similarity. For the three FM measures, commonality analysis was used to estimate unique and common contributions to the model. The largest contribution (35%) to the regression equation was from the shared variance among the three FM measures, with variance common to the SNR and Duration Down conditions, the SNR and Duration Up conditions, and the two Duration conditions accounting for 12%, 11%, and 7%, respectively, of R^2 . Unique contributions from each predictor were 20%, 13%, and 2% for the Duration Down, SNR, and Duration Up results, respectively. Thus, among the three psychoacoustic variables that showed significant pairwise correlation to masked speech thresholds, variance contributed by Duration Up performance to predicting speech ability was largely shared with the other FM metrics.

DISCUSSION

The study evaluated the ability of CI users to discriminate stochastic patterns of low-rate FM to gain a better understanding of dynamic spectral processing by CI users and possible relationship of this ability to speech perception. Performance was measured in five conditions—Delta F, Masked Delta F, SNR, Duration Down, and Duration Up—using three different threshold metrics. The Delta F conditions measured thresholds in terms of the minimal frequency excursion needed for FM pattern discrimination. Using the same experimental protocol, Sheft et al (2012) measured Delta F performance in quiet from normal-hearing young adults and listeners 60 yr old who had either normal hearing or a mild-to-moderate sensorineural hearing loss. Mean thresholds from the young and older adults were lower than the current results by factors of 3.2 and 1.76, respectively. Past results from CI users show elevated thresholds for FM detection and pitch, rate, and glide discrimination (Chen and Zeng, 2004; Green et al, 2005; Wei et al, 2007; Chatterjee and Peng, 2008; Drennan et al, 2008; Luo et al, 2008; Pretorius and Hanekom, 2008; Kang et al, 2009; Won et al, 2010; Schauwers et al, 2012). Current results extend the finding of threshold elevation for CI users to the processing of complex dynamic spectral patterns.

The mean delta F threshold in quiet was roughly 125 Hz. Past study of frequency discrimination by CI users has indicated large interlistener variability. With pure-tone stimuli ranging from 250 to 4000 Hz and listening through their implant processor, Wei et al (2007) reported a roughly constant frequency difference limen of in the vicinity of 100 Hz. Pretorius and Hanekom (2008), however, observed that thresholds can often vary with small

changes in base frequency due to the relationship between stimulus frequency and processor filterbank characteristic. For complex tones with a fundamental ranging from 262 to 392 Hz, Kang et al (2009) reported discrimination thresholds ranging from 1.0 to 8.0 semitones and a mean threshold of 3.0 semitones. The mean threshold of 125 Hz with a 1-kHz carrier in the current study corresponds to 2.0 semitones. With caution regarding interlistener variability and effects of differences in both method and implant processor, past work can be taken to suggest that the present delta F thresholds represent pattern discrimination thresholds that are reasonably close to thresholds for detecting a single change in frequency. The current task, however, cannot be solved by simply detecting a change in frequency; the same maximum frequency excursion was used in each stimulus presentation of a given trial. Furthermore, due to the stochastic nature of the frequency modulation in this study, many of the frequency fluctuations that fully describe a specific pattern were less than the value of the maximum excursion, the metric used to define threshold. Thus, current results suggest that CI users were able to use near-threshold pattern features in the discrimination task. Though this relatively robust ability to discriminate random patterns of low-rate FM could potentially benefit speech perception, results did not offer evidence of relationship for psychoacoustic thresholds measured in terms of delta F.

A speech-babble masker was present in two of the discrimination conditions. In the Masked Delta F condition, a high SNR of 16 dB was used to minimize energetic masking, so effects might relate more directly to factors such as auditory attention and distraction. Measuring the SNR at which listeners could just discriminate stochastic FM patterns, the FM SNR condition offered evaluation of discrimination ability at low sensation levels, a characteristic of many adverse listening environments in which speech processing is difficult. The mean FM SNR threshold was 2.36 dB. The fixed SNR of 16 dB in the Masked Delta F conditions was thus over 13-dB greater than the threshold SNR needed to discriminate FM pattern with F fixed at 750 Hz in the FM SNR condition. While not intended as a comparison of performance between conditions that used different threshold metrics, the SNR difference suggests that there would have been little energetic masking at the higher SNR of the Masked Delta F condition.

In the Masked Delta F condition, mean threshold was significantly elevated by a factor of 2.11 due to the addition of the masker. This value is roughly similar to the one reported by Sheft et al (2012) for older adults (1.84), with both values much larger than the factor of 1.31 obtained from young normal-hearing adults. Results from the Masked Delta F condition are consistent with past work with CI users evaluating masking of pitch contour discrimination. Working with Mandarin tones distinguished by differences in gradual pitch variation of the fundamental frequency contour, Wei et al (2007) measured recognition ability by CI users in the presence of speech-shaped noise. In that work, results showed that performance progressively improved with SNR up to highest SNR tested of 10 dB. At the highest level used, however, a masking effect was still evident. Wei et al. also reported that recognition performance of CI users was not above chance until the SNR was greater than -5 dB. To achieve mean threshold performance in this study, an SNR of 2.36 dB was needed in the FM SNR condition. This SNR is 20.0 and 9.1 dB higher than that obtained by Sheft et al (2012) for younger and older adults, respectively, from a similar experimental condition, but with the previous work using a smaller F of 400 Hz than the present use of

750 Hz with CI users. Thus, current results are consistent with the high level of auditory difficulty often experienced by CI users due to background noise or interference.

In the Duration conditions, discrimination cues vary from those based on following the pattern of frequency fluctuation with longer duration stimuli to differences in a global stimulus percept (e.g., timber or pitch) at brief durations. Threshold values were much larger in the Duration Down than Duration Up condition, despite the fact that the two were identical except for starting stimulus duration. This result indicates an effect of cue transition on listener performance. With contrasting modulators used on a given trial having identical long-term amplitude spectra, the remaining phase-spectra differences lead to differences in arrival time by frequency for the two acoustic stimuli. To serve as a discrimination cue at brief stimulus durations, auditory processing must in some way window stimuli with a minimum integration time to preserve arrival-time differences which represent a type of cross-spectral asynchrony. Working with Huffman sequences which differ in terms of phase but not amplitude spectra, Patterson and Green (1970) assessed auditory temporal acuity in terms of the minimum stimulus duration that allowed for discrimination, with results indicating that highly trained listeners were able to discriminate stimuli as brief as 2.5 msec. Again using a smaller DF of 400 Hz than the 750 Hz of the current study, the mean Duration Up threshold reported by Sheft et al (2012) for 18 normal-hearing young adults was 9.7 msec, with one standard deviation above the mean at 30.2 msec. In the present work, Duration Up thresholds from four CI users were less than 10 msec with thresholds less than 30 msec obtained from an additional three listeners. As reported for other measures of temporal resolution (Busby and Clark, 1999), current results thus indicate that some but not all CI users show normal temporal acuity based on the Duration Up metric. With other measures of temporal resolution in CI users at times confounded by level and loudness (Luo et al, 2008; Garadat and Pfingst, 2011; Galvin et al, 2014), use of transient stimuli in the Duration Up procedure may offer an alternative approach.

BKB-SIN SNR-50 results showed a performance range typical of past results of speech perception by CI users. Except for a different presentation level of 60 dB HL, Donaldson et al (2009) used the same four-talker speech-babble masker and experimental procedure in BKB-SIN testing of CI users as used in the current study. Across 21 CI users, they reported a range for SNR-50 of 4.3 to 20.9 dB, similar to the range of 6.75 to 20.0 dB obtained in the present work. Group performance, however, was slightly better in the study of Donaldson et al, with their results showing a mean SNR-50 of 11.9 dB in contrast to the current 14.7-dB average. In the current study, FM SNR discrimination thresholds were significantly lower than the listeners' BKB-SIN SNR-50 thresholds. With the same masker used in both tasks, this result indicates that despite the elevated FM SNR thresholds of CI users, speech cues conveyed by low-rate FM can potentially contribute to masked speech intelligibility.

In Pearson pairwise analysis, there were three significant relationships between measures of FM pattern discrimination and speech ability as assessed with BKB-SIN. The absence of significant correlation for two of the five measures indicates that the basis of the significant relationships was not solely procedural due to some general aspect of the psychophysical testing or stimulus type. Using the same five metrics in a study of the effects of aging and

hearing loss, Sheft et al (2012) found that only performance in the FM SNR condition showed a significant correlation to masked speech perception. The greater number of significant relationships in the present work with CI users suggests that the reduced redundancies of speech available for CI users may make relationships more apparent.

The characteristics of specific conditions can be used to infer aspects of the relationship to speech performance. The significant correlation between the FM SNR thresholds and BKB-SIN performance supports speculation that the ability to process low-rate FM may contribute to masked speech intelligibility for CI users. Shortening stimulus duration in the Duration Down condition reduces the number of frequency fluctuations and progressively brings stimuli closer to the durations typical of formant transitions. The significant correlation between Duration Down thresholds and BKB-SIN performance may reflect involvement of fidelity of coding of first formant transitions. This speculation is consistent with the suggestion of Dorman et al (1996) that coding of frequencies in the region of the first formant plays an important role in consonant recognition by CI users. Finally, the correlation between BKB-SIN results and the Duration Up measure shows a relationship between speech perception ability and auditory temporal resolution.

Results from the commonality analysis indicate a limitation to this speculation regarding specific conditions. The greatest contribution to the prediction of speech performance arose from common variance among the three FM measures. Furthermore, unique contribution from the Duration Up results was quite small. The analysis thus showed that despite differences in threshold metric, a common basis in the processing of stochastic FM affected measured relationships to speech perception ability. This common basis, however, did not extend to the two Delta F conditions in which results did not significantly correlate to BKB-SIN thresholds. The important distinction may have been that the two Delta F conditions procedurally adapted on frequency excursion, measuring a threshold value for delta F, while the three metrics which did show relationship to speech ability maintained a fixed suprathreshold value for the maximum frequency excursion across all trials.

In general, CI users have access to temporal and place cues for discrimination of either static or dynamic frequency differences. Temporal cues, coded either by changes in envelope periodicity or pulse-rate differences with some processors, are most robust for frequencies below roughly 300 Hz (Dorman et al, 1996; Chen and Zeng, 2004; Chatterjee and Peng, 2008). With a 1-kHz carrier frequency in the FM conditions of the current work, it is reasonable to assume that temporal cues were not a dominant factor for most participants in any psychoacoustic condition. Drennan et al (2008) argued that discrimination of the dynamic spectral changes of Schroeder-phase stimuli by CI users is based on temporally distributed place coding of changes in the short-term frequency content of the signals. The ability of CI users to discriminate stochastic FM stimuli at a center frequency of 1 kHz may have a similar basis. In the present study, FM SNR thresholds along with the two FM-duration measures correlated with speech perception in the presence of four-talker interference. These significant correlations thus suggest that the ability to discriminate temporal differences in the spatially distributed place pattern of stimulation in response to stochastic FM represents an aspect of auditory processing relevant to speech perception by CI users.

CONCLUSIONS

This study evaluated the ability of CI users to discriminate stochastic patterns of low-rate FM in five conditions with stimulus parameters chosen for possible relevance to speech perception. Using stochastic frequency modulators derived from 5-Hz lowpass noise applied to a 1-kHz carrier, discrimination thresholds were measured in terms of frequency excursion both in quiet and with a speech-babble masker present, stimulus duration, and SNR in the presence of a speech-babble masker. Compared to past results obtained from young normal-hearing adults or older adults with either normal hearing or a mild-to-moderate hearing loss, mean thresholds obtained from CI users were higher in all conditions. Without masking, CI users were able to use near-threshold pattern features in the discrimination task. However, results from the masking conditions were consistent with the high level of auditory difficulty often experienced by CI users due to background noise or interference. For CI users, significant correlations were obtained between masked speech intelligibility and three metrics of FM discrimination involving either SNR or stimulus duration. Despite differences in threshold metric across the three measures, shared variance accounted for much of the effect, indicating a common basis in the processing of the suprathreshold frequency excursions of the stochastic FM. Overall, results confirm that discrimination of differences in the temporally distributed place coding of stochastic FM represents an aspect of auditory processing relevant to speech perception by CI users.

Acknowledgments

This research was supported in part by Award Number R15DC011916 from the National Institute on Deafness and Other Communication Disorders.

Abbreviations

BKB-SIN	Bamford-Kowal-Bench speech-in-noise
CI	cochlear implant
F	maximum frequency excursion
FM	frequency modulation
SNR	signal-to-noise ratio
SPL	sound pressure level

REFERENCES

- Ardoint M, Lorenzi C. Effects of lowpass and highpass filtering on the intelligibility of speech based on temporal fine structure or envelope cues. *Hear Res.* 2010; 260(1–2):89–95. [PubMed: 19963053]
- Binns C, Culling JF. The role of fundamental frequency contours in the perception of speech against interfering speech. *J Acoust Soc Am.* 2007; 122(3):1765–1776. [PubMed: 17927436]
- Bench J, Kowal A, Bamford J. The BKB (Bamford-Kowal-Bench) sentence lists for partially-hearing children. *Br J Audiol.* 1979; 13(3):108–112. [PubMed: 486816]
- Busby PA, Clark GM. Gap detection by early-deafened cochlear-implant subjects. *J Acoust Soc Am.* 1999; 105(3):1841–1852. [PubMed: 10089607]

- Chatterjee M, Peng SC. Processing F0 with cochlear implants: Modulation frequency discrimination and speech intonation recognition. *Hear Res.* 2008; 235(1–2):143–156. [PubMed: 18093766]
- Chen H, Zeng FG. Frequency modulation detection in cochlear implant subjects. *J Acoust Soc Am.* 2004; 116(4):2269–2277. [PubMed: 15532658]
- Ciocca V, Francis AL, Aisha R, Wong L. The perception of Cantonese lexical tones by early-deafened cochlear implantees. *J Acoust Soc Am.* 2002; 111(5):2250–2256. [PubMed: 12051445]
- Collins LM, Wakefield GH, Feinman GR. Temporal pattern discrimination and speech recognition under electrical stimulation. *J Acoust Soc Am.* 1994; 96(5):2731–2737. [PubMed: 7983278]
- Donaldson GS, Chisolm TH, Blasco GP, Shinnick LJ, Ketter KJ, Krause JC. BKB-SIN and ANL predict perceived communication ability in cochlear implant users. *Ear Hear.* 2009; 30(4):401–410. [PubMed: 19390441]
- Dorman MF, Cutting JE, Raphael LJ. Perception of temporal order in vowel sequences with and without formant transitions. *J Exp Psychol Hum Percept Perform.* 1975; 104(2):121–129. [PubMed: 1194865]
- Dorman MF, Smith LM, Smith M, Parkin JL. Frequency discrimination and speech recognition by patients who use the Ineraid and continuous interleaved sampling cochlear-implant signal processors. *J Acoust Soc Am.* 1996; 99(2):1174–1184. [PubMed: 8609301]
- Drennan WR, Longnion JK, Ruffin C, Rubinstein JT. Discrimination of Schroeder-phase harmonic complexes by normal-hearing and cochlear-implant listeners. *J Assoc Res Otolaryngol.* 2008; 9(1):138–149. [PubMed: 18066624]
- Etymotic Research. BKB-SIN Speech-in-Noise Test (Version 1.03) [Audio CD]. Elk Grove Village, IL: 2005.
- Firszt JB, Holden LK, Skinner MW, et al. Recognition of speech presented at soft to loud levels by adult cochlear implant recipients of three cochlear implant systems. *Ear Hear.* 2004; 25(4):375–387. [PubMed: 15292777]
- Folstein MF, Folstein SE, McHugh PR. “Mini-mental state”. A practical method for grading the cognitive state of patients for the clinician. *J Psychiatr Res.* 1975; 12(3):189–198. [PubMed: 1202204]
- Fu QJ. Temporal processing and speech recognition in cochlear implant users. *Neuroreport.* 2002; 13(13):1635–1639. [PubMed: 12352617]
- Fu QJ, Hsu CJ, Horng MJ. Effects of speech processing strategy on Chinese tone recognition by Nucleus-24 cochlear implant users. *Ear Hear.* 2004; 25(5):501–508. [PubMed: 15599196]
- Galvin JJ 3rd, Fu QJ, Oba S, Ba kent D. A method to dynamically control unwanted loudness cues when measuring amplitude modulation detection in cochlear implant users. *J Neurosci Methods.* 2014; 222:207–212. [PubMed: 24269251]
- Garadat SN, Pfingst BE. Relationship between gap detection thresholds and loudness in cochlear-implant users. *Hear Res.* 2011; 275(1–2):130–138. [PubMed: 21168479]
- 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]
- Kang R, Nimmons GL, Drennan W, et al. Development and validation of the University of Washington Clinical Assessment of Music Perception test. *Ear Hear.* 2009; 30(4):411–418. [PubMed: 19474735]
- Levitt H. Transformed up-down methods in psychoacoustics. *J Acoust Soc Am.* 1971; 49(2, part 2):467–477. [PubMed: 5541744]
- Luo X, Fu QJ, Wei CG, Cao KL. Speech recognition and temporal amplitude modulation processing by Mandarin-speaking cochlear implant users. *Ear Hear.* 2008; 29(6):957–970. [PubMed: 18818548]
- Miller SE, Schlauch RS, Watson PJ. The effects of fundamental frequency contour manipulations on speech intelligibility in background noise. *J Acoust Soc Am.* 2010; 128(1):435–443. [PubMed: 20649237]
- Nimon K. Regression commonality analysis: Demonstration of an SPSS solution. *Multiple Linear Regress Viewpoints.* 2010; 36:10–17.

- Patterson JH, Green DM. Discrimination of transient signals having identical energy spectra. *J Acoust Soc Am*. 1970; 48(4):894–905. [PubMed: 5480387]
- Pretorius LL, Hanekom JJ. Free field frequency discrimination abilities of cochlear implant users. *Hear Res*. 2008; 244(1–2):77–84. [PubMed: 18692556]
- Remez RE, Rubin PE, Berns SM, Pardo JS, Lang JM. On the perceptual organization of speech. *Psychol Rev*. 1994; 101(1):129–156. [PubMed: 8121955]
- Saoji AA, Litvak L, Spahr AJ, Eddins DA. Spectral modulation detection and vowel and consonant identifications in cochlear implant listeners. *J Acoust Soc Am*. 2009; 126(3):955–958. [PubMed: 19739707]
- Schouwens K, Coene M, Heeren W, et al. Perception of pitch changes in hearing-impaired adults with aided and unaided hearing loss. *J Hear Sci*. 2012; 2:25–34.
- Sheft S, Ardoint M, Lorenzi C. Speech identification based on temporal fine structure cues. *J Acoust Soc Am*. 2008; 124(1):562–575. [PubMed: 18646999]
- Sheft S, Shafiro V, Lorenzi C, McMullen R, Farrell C. Effects of age and hearing loss on the relationship between discrimination of stochastic frequency modulation and speech perception. *Ear Hear*. 2012; 33(6):709–720. [PubMed: 22790319]
- Spitzer SM, Liss JM, Mattys SL. Acoustic cues to lexical segmentation: a study of resynthesized speech. *J Acoust Soc Am*. 2007; 122(6):3678–3687. [PubMed: 18247775]
- Troendle JF. A stepwise resampling method of multiple hypothesis testing. *J Am Stat Assoc*. 1995; 90:370–378.
- 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(Suppl):62S–65S. [PubMed: 17496650]
- Won JH, Drennan WR, Rubinstein JT. Spectral-ripple resolution correlates with speech reception in noise in cochlear implant users. *J Assoc Res Otolaryngol*. 2007; 8(3):384–392. [PubMed: 17587137]
- Won JH, Drennan WR, Kang RS, Rubinstein JT. Psychoacoustic abilities associated with music perception in cochlear implant users. *Ear Hear*. 2010; 31(6):796–805. [PubMed: 20595901]

Figure 1. Schematic illustration of two examples of stochastic FM showing the contrasting instantaneous frequency functions of the stimuli. In the sample-discrimination procedure with scaling for a common F , discrimination is based on differences in the temporal pattern of frequency fluctuation.

Figure 2.

Box plots showing FM discrimination thresholds from each condition. The line through each box is the median threshold; the upper and lower box edges indicate the 25th and 75th percentiles with error bars showing the 10th and 90th percentiles.

Figure 3.

Relationship between individual BKB-SIN SNR-50 thresholds and performance on the FM SNR, Duration Down, and Duration Up tasks. Symbol shape indicates coding strategy of the individual's CI, with upright triangles, circles, squares, and the inverted triangle for high resolution sequential, advanced combination encoder, simultaneous analog stimulation, and multiple pulsatile stimulation, respectively. In each panel, the best fitting linear regression of the aggregate data are shown as a solid line, with correlation and p value in the upper left corner. p values for the correlations were adjusted with a resampling method to control family-wise error.

CI User Information and Speech Results.

Table 1

Participant	Age	Years of Use	CI Type	Strategy	CNC (% Correct)	BKB-SIN SNR-50 (dB)
S1	43	2	HiRes-90k	HiRes-S	33.0	20.00
S2	22	10	Freedom	ACE	82.5	16.75
S3	31	2	Freedom	ACE	62.5	15.00
S4	66	7	HiRes-90k	HiRes-S	68.0	13.25
S5	77	14	Clarion II	SAS	45.5	14.00
S6	21	9	Clarion II	SAS	43.5	19.50
S7	53	4	Freedom	ACE	62.0	15.50
S8	68	8	Freedom	ACE	69.0	16.75
S9	71	1	Freedom	ACE	63.5	12.00
S10	39	3	Freedom	ACE	91.0	7.50
S11	40	6	Nucleus 24	ACE	43.5	18.50
S12	49	4	Freedom	ACE	59.0	6.75
S13	49	9	Clarion II	MPS	38.0	16.00
Mean					58.5	14.73
Standard deviation					17.3	4.10

Note: ACE = advanced combination encoder; CNC = consonant-nucleus-consonant; HiRes-S = high resolution sequential; MPS = multiple pulsatile stimulation; SAS = simultaneous analog stimulation