

Fundamental frequency is critical to speech perception in noise in combined acoustic and electric hearing^{a)}

Jeff Carroll^{b)}

Hearing and Speech Research Laboratory, Department of Biomedical Engineering, University of California, Irvine, California 92697-5320

Stephanie Tiaden^{c)}

Hearing and Speech Research Laboratory, Department of Biomedical Engineering, University of California, Irvine, California 92697-5320

Fan-Gang Zeng^{d)}

Hearing and Speech Research Laboratory, Departments of Anatomy and Neurobiology, Biomedical Engineering, Cognitive Sciences, and Otolaryngology—Head and Neck Surgery, University of California, Irvine, California 92697-5320

(Received 28 January 2011; revised 11 May 2011; accepted 5 August 2011)

Cochlear implant (CI) users have been shown to benefit from residual low-frequency hearing, specifically in pitch related tasks. It remains unclear whether this benefit is dependent on fundamental frequency (F0) or other acoustic cues. Three experiments were conducted to determine the role of F0, as well as its frequency modulated (FM) and amplitude modulated (AM) components, in speech recognition with a competing voice. In simulated CI listeners, the signal-to-noise ratio was varied to estimate the 50% correct response. Simulation results showed that the F0 cue contributes to a significant proportion of the benefit seen with combined acoustic and electric hearing, and additionally that this benefit is due to the FM rather than the AM component. In actual CI users, sentence recognition scores were collected with either the full F0 cue containing both the FM and AM components or the 500-Hz low-pass speech cue containing the F0 and additional harmonics. The F0 cue provided a benefit similar to the low-pass cue for speech in noise, but not in quiet. Poorer CI users benefited more from the F0 cue than better users. These findings suggest that F0 is critical to improving speech perception in noise in combined acoustic and electric hearing.

© 2011 Acoustical Society of America. [DOI: 10.1121/1.3631563]

PACS number(s): 43.66.Hg, 43.71.Ky, 43.71.Bp, 43.64.Me [KG]

Pages: 2054–2062

I. INTRODUCTION

Cochlear implant (CI) users often have difficulty perceiving speech in noisy backgrounds despite their good performance in quiet. The difficulty in CI speech recognition in noise has been related to CI users' inability to discriminate talkers and to segregate different speech streams (e.g., Shannon *et al.*, 2004; Vongphoe and Zeng, 2005; Luo and Fu, 2006). In acoustic hearing, one cue used to segregate multiple speech streams is the difference in voice pitch or fundamental frequency (F0) between talkers (Brox and Nootboom, 1982; Assmann and Summerfield, 1990; Drullman and Bronkhorst, 2004). In addition to the absolute difference in F0, frequency and amplitude modulations in the F0 and harmonics allow for the segregation of talkers (McAdams, 1989; Marin and McAdams, 1991; Summerfield and Culling, 1992; Binns and Culling, 2007).

A different line of research in speech reading also revealed the importance of F0 in visual speech recognition.

For example, earlier work showed that the mere presence of voice pitch significantly improved a listener's ability to lip-read speech (Rosen *et al.*, 1981; Breeuwer and Plomp, 1984). Grant *et al.* (1984) systematically demonstrated that cues to voicing duration, F0 frequency modulations, and low-frequency amplitude modulations all benefited speech reading. While speech reading and CI processing have some similarities, they also have differences. Similar to speech reading, the CI does not transmit the F0 cue effectively. Contrary to the visual cues in speech reading, the CI transmits mostly speech cues related to temporal envelopes. For example, /b/ and /d/ are easily discernable using the visual cue, but very difficult using the envelope cue (Van Tasell *et al.*, 1987). Despite these differences, it is not clear whether the benefit of adding F0 to speech reading can be extended to CI listening.

A good representation of pitch is useful for speech segregation in noise when the speech and noise cannot be spatially separated. In a normal cochlea, pitch is coded by the place of excitation (von Békésy, 1960; Greenwood, 1990), the temporal fine structure of the neural discharge (Seebeck, 1841), or is weakly coded by the envelope of a modulated carrier (Burns and Viemeister, 1976, 1981). Overall, CIs lack good representation of pitch. The limited number of electrodes and the broad spread of electrical current in CIs

^{a)}Portions of this work were presented at the 2007 Conference on Implantable Auditory Prostheses, Granlibakken Conference Center, Lake Tahoe, CA.

^{b)}Current address: SoundCure, 33 Arch Street, Boston, MA 02110.

^{c)}Current address: GluMetrics, 15375 Barranca Parkway, Irvine, CA 92618.

^{d)}Author to whom correspondence should be addressed. Electronic mail: fzen@uci.edu

provide for only a very coarse place pitch cue. The fixed pulse rate does not provide any temporal fine structure cue, leaving only the weak temporal envelope cue for pitch. Although many methods have been proposed to enhance pitch representations in CIs, e.g., by increasing spectral resolution through virtual channels (Tonwshend *et al.*, 1987; McDermott and McKay, 1994) or channel density (Geurts and Wouters, 2004; Carroll and Zeng, 2007), restoring the temporal cue with a varying pulse rate (Shannon, 1983; Jones *et al.*, 1995), or by enhancing temporal envelope modulations (Green *et al.*, 2002, 2004, 2005; Carroll *et al.*, 2005; Vandali *et al.*, 2005; Laneau *et al.*, 2006), all have shown limited functional benefit.

In contrast, combined acoustic and electric stimulation can provide significant benefits for speech recognition in noise and melody recognition in both simulated (Dorman *et al.*, 2005; Qin and Oxenham, 2006; Chang *et al.*, 2006) and actual combined acoustic and electric hearing users (von Ilberg *et al.*, 1999; Turner *et al.*, 2004; Kong *et al.*, 2005). Most studies have evaluated the role of low-frequency acoustic information in CI listeners with sharply sloping audiograms or by simulating this type of hearing loss in normal-hearing listeners. These studies have suggested that the benefit of combined acoustic and electric stimulation is a result of the F0 cue present in the low-frequency acoustic signal (Brown and Bacon, 2009a,b; Cullington and Zeng, 2010; Zhang *et al.*, 2010a). Brown and Bacon (2009b) examined the role of various F0 cues in simulated electro-acoustic-stimulation and found that both frequency and amplitude modulation cues contributed to the combined hearing benefit. On the other hand, Kong and Carlyon (2007), also using a CI simulation, found conflicting results in that frequency modulation in F0 offered no additional benefit over a flat F0 contour, suggesting that merely the presence of voicing may account for the combined hearing benefit.

The present study examined the role of different F0 cues in three speech recognition experiments using both simulated and actual implant listeners. Experiment 1 replicated previous work (Brown and Bacon, 2009b) by acoustically presenting the target speech F0 only and comparing the F0 only condition to the more commonly used method of adding low-pass filtered speech information to the CI. In addition, exp. 1 expanded on previous work by investigating the effect of presenting the masker F0 only, or both the target and masker F0s, on speech recognition in a competing talker background. exp. 2 investigated the contributions of the FM and AM cues independently, including the use of a stationary sinusoid to cue voicing information. Experiment 3 compared the effect of F0 with the low-passed condition in actual CI users to assess the relative contributions of these acoustic cues to combined acoustic and electric hearing.

II. EXPERIMENT 1

A. Methods

1. Subjects

21 normal-hearing subjects (11 females and 10 males) between the ages of 19 and 60 yr old (mean 23.9 yr) partici-

pated in this study. The University of California Irvine's Institutional Review Board approved all experimental procedures. Informed consent was obtained from each subject. Subject audiograms were collected prior to testing to confirm the absence of any hearing deficiency [20 dB hearing level or lower for octave frequencies 250 to 8000 Hz]. All subjects were native English speakers and were compensated for their participation.

2. Stimuli

Subjects were tested on sentence intelligibility in the presence of a single competing talker using a cochlear implant simulation combined with unprocessed, low-frequency acoustic information. The sentences were chosen from the IEEE sentence set with each sentence containing five target words (Rothausser *et al.*, 1969). The target was a male with an average F0 of 102 ± 33 Hz. The masker was a female speaking the sentence "It was hidden from sight by a mass of leaves and shrubs" (F0 = 238 ± 36 Hz). This sentence was chosen because it is longer than all target sentences, ensuring that no part of any target sentence would be unmasked. The use of a fixed single sentence as a masker certainly reduced the amount of informational masking, but it allowed a listener to "listen-in-the-dips," or between the masking words. Another reason for the use of a single masker sentence was that neither actual nor simulated CI listeners can differentiate the female masker from the male target, making typical instruction, for example, asking the subject to ignore a specific masking talker or to learn the characteristics of a target talker difficult, if not impossible (e.g., Freyman *et al.*, 2001). The present single masker sentence was used previously to demonstrate the masking difference between a steady-state noise and a competing voice (Stickney *et al.*, 2004), and represented an experimental condition between the traditional informational masking studies (e.g., Freyman *et al.*, 2004) and the traditional "listening-in-the-dips" studies using amplitude-modulated noises (e.g., Jin and Nelson, 2006).

The CI simulation was created using a six-band noise-excited vocoder (Shannon *et al.*, 1995) with cutoff frequencies assigned according to the Greenwood map between 250 and 8700 Hz (Greenwood, 1990). The target and masker were combined at a specified SNR and processed together. Combined stimuli were passed through third order bandpass elliptical filters, full-wave rectified and low-pass filtered to obtain the envelope, and used to modulate a broadband speech-shaped noise. The stimulus was then passed through the original analysis filter again to eliminate sidebands induced outside the channel due to modulation. Low-frequency information was then added to the processed signal and delivered either as a simple low-passed version of the speech (simulating low-frequency residual hearing) or as an explicit representation of the F0. The low-passed speech was derived from a third order elliptical filter with a cutoff at 500 Hz applied to the combined target and masker signal allowing access to F0 as well as some first formant (F1) information. The low pass information significantly overlapped with the low end of the CI simulation frequency range. This combination is consistent with CI users with a full insertion

electrode array, who have residual contralateral hearing (e.g., subjects in exp. 3), but not with hybrid users with a short insertion electrode array, who have residual ipsilateral hearing. As the lowest channel in the CI simulation will not preserve the fine frequency information, the overlap should not present a redundancy of pitch information. The F0 was generated for the target and masker separately and presented either individually or combined. Frequency information of the F0 contour was extracted using the STRAIGHT program (Kawahara *et al.*, 1999). The output only contained information during voiced speech; unvoiced segments were represented as silence. The amplitude information for the F0 contour was extracted by low-passing the original signal with a third order elliptical filter with a cutoff equal to the highest frequency in the F0 contour, full-wave rectifying, and low passing at 20 Hz to minimize the introduction of sidebands to the time-varying frequency contour. All stimuli were presented monaurally at an overall level of 75 dB sound pressure level through Sennheiser HDA 200 Audiometric headphones. This level was verified by a Bruel and Kjaer type 2260 Modular Precision Sound Analyzer coupled to an artificial ear.

3. Procedure

Subjects were presented with the combined speech and asked to ignore the masking sentence. Subjects responded verbally regarding what they heard and responses were scored as either completely correct or incorrect for the five target words. The stimuli were presented at an initial SNR of 0 dB and adapted with a one-up, one-down decision rule corresponding to a 50% correct point on the psychometric function (Levitt, 1971). This value has been referred to in the literature as SNR for 50% correct, or commonly as the speech reception threshold (SRT). Given that SRT is defined as the absolute level in quiet in which the subject scores 50% correct and not as the relative level in noise, this nomenclature can be misleading. In this text, this value will be referred to as the speech reception threshold in noise (SRTn) to differentiate it from SRT. The SNR was increased in 5 dB steps without changing the first sentence until the subject could correctly identify all target words. The SNR was decreased for a correct response and increased for an incorrect response by 5 dB for the first two reversals and 2 dB for the remainder of the test. Each subject was tested on two independent lists of ten sentences each selected randomly for each condition. Lists were never duplicated for a given subject. No practice was given. The SRTn was determined as the average of the third through final reversals. Subjects were tested on five conditions (four experimental plus 1 control) named for the type of low-frequency information presented: low-passed information below 500 Hz from the combined target and masker sentences (lowpass), the F0 of the target speaker only (target F0), the F0 of the masker speaker only (masker F0), or a combination of the target and masker F0s (target + masker F0). Values were compared against the control of a CI simulation alone (sim. only). All ten runs (five conditions \times two lists) were presented in random order for each subject. Since SRTn required that a sentence be identified as completely correct, conditions degraded

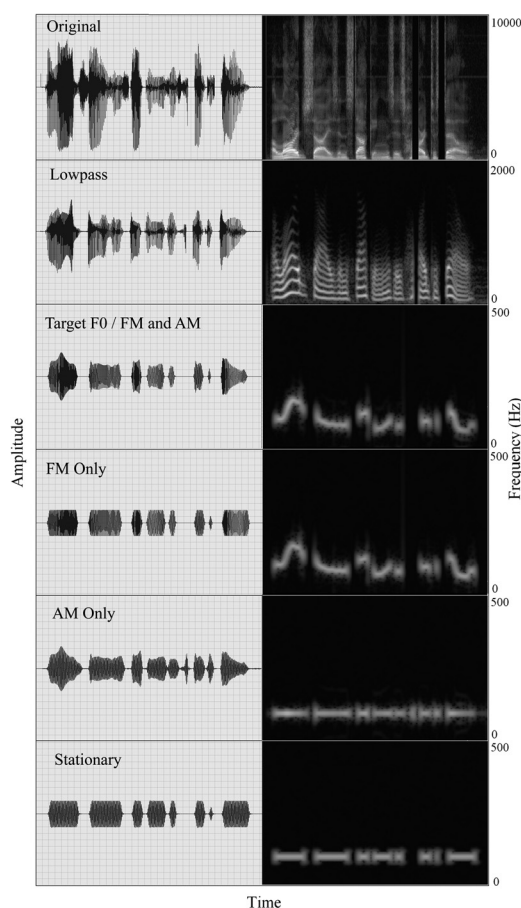


FIG. 1. Waveforms (left panels) and spectrograms (right panels) of a sample sentence. The rows from top to bottom show the original speech, a low-passed version at 500 Hz (lowpass), the F0 contour (named target F0 in exp. 1 or FM and AM in exp. 2), the frequency modulation of the F0 contour (FM only), the amplitude modulation of the F0 contour (AM only), or a sine wave cueing the presence of voicing (stationary).

sufficiently to prevent correct identification of the first sentence (i.e., lowpass and target F0) could not be tested with the SRTn protocol. Percent correct scores were collected for five of the subjects for the lowpass condition at an SNR of 0 dB and the target F0 condition in quiet. The top three panels of Fig. 1 show waveforms and spectrograms for a sample sentence, unprocessed, low-passed at 500 Hz, and the F0. It should be noted that the Lowpass condition is shown without the masker for clarity, but in the study, the lowpass condition always had a masker present. The left panels show the time varying amplitude waveform while the right panels show the time varying frequency contours.

All subjects were tested on all five conditions to allow for a within-subjects analysis. Data were analyzed using a one factor repeated measures analysis of variance (ANOVA) with the type of acoustic information as the independent variable.

B. Results and discussion

Figure 2 shows the SRTn values in dB for the five conditions. Error bars indicate the standard error of the mean. The CI simulation alone provided for an average SRTn of 11.6 dB, similar to values obtained from other simulation

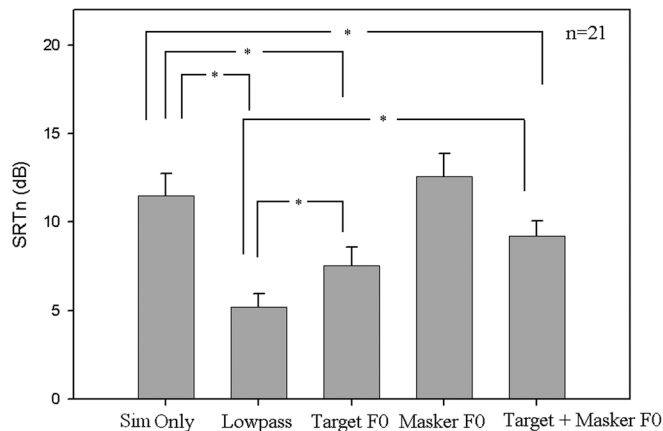


FIG. 2. Speech recognition as a function of different low-frequency listening conditions in simulated combined hearing. SRTn values represent the SNR at which the subjects scored 50% correct for each condition. Asterisks indicate statistical significance ($p < 0.05$).

studies (e.g., Chang *et al.*, 2006; Kong and Carlyon, 2007). Presentation of the low-passed speech improved the CI simulation alone by 6.3 dB, within the range of 5–10 dB improvement reported previously (Kong *et al.*, 2005; Kong and Carlyon, 2007). Target F0 and the combined target and masker F0s also led to an improvement of 4.0 and 2.1 dB SRTn, respectively. In contrast, presentation of the masker F0 provided no benefit over the CI simulation and actually increased SRTn by 1 dB.

A main effect was seen for the type of low-frequency information applied [$F(4,80) = 16.3, p < 0.05$]. *Post hoc* analyses with a Newman–Keuls correction indicated that the addition of lowpass information, the target F0, or the target + masker F0s lead to significantly better results than a CI simulation alone ($p < 0.05$). Lowpass information also provided for significantly better performance than the target F0 or target and masker F0s ($p < 0.05$). There was no statistical difference between the sim. only condition and the addition of the masker F0, nor between the target F0 and the target and masker F0s.

To characterize contributions of the lowpass and F0 information to speech recognition, percent correct scores were collected. The lowpass alone condition provided for $26.8 \pm 5.6\%$ recognition of keywords, and only $6.0 \pm 3.7\%$ correct for sentences. This value falls within the expected region (Chang *et al.*, 2006; Kong and Carlyon, 2007). Chang *et al.* demonstrated 51% of keywords and 10% with sentences for HINT sentences that contain more contextual information than the IEEE sentences. Kong and Carlyon showed only 15% with words in quiet, but their stimuli were low-passed at 125 Hz and masked above 500 Hz. On the other hand, the target F0 alone provided for 0.2% word recognition with only 1 subject being able to identify 1 word.

Although the target F0 itself provided essentially no word recognition ability, a significant proportion of the benefit seen with the addition of low-frequency information to a CI simulation can be attributed to the extraction of the F0 of the target speaker. The improvement of the lowpass condition over the target F0 or target and masker F0 conditions demonstrated that a further 2–4 dB benefit could be obtained

from the presence of multiple harmonics or vowel formant information, when measuring the SRTn. Using a percent correct metric, Cullington and Zeng (2010) demonstrated a similar improvement trend in combining electric stimulation with progressively more low-frequency acoustic information. Optimizing the benefit available from the target F0 in a speech processing strategy will require knowledge of what acoustic components are responsible for the improvement in speech segregation. Experiment 2 explicitly evaluated the contributions from FM and AM separately.

III. EXPERIMENT 2

A. Methods

1. Subjects

The same twenty-one subjects as in exp. 1 were tested in exp. 2.

2. Stimuli

Three conditions were tested to determine the relative benefit of frequency and amplitude information of the F0: Frequency modulation information only (FM only), amplitude modulation information only (AM only), or stationary. These conditions were directly compared to the target F0 condition from exp. 1, which contains both frequency and amplitude modulation information, and the sim. only condition as a control. As each condition was aimed at determining the contribution of a specific physical component of the target F0, no condition contained the F0, or other low-frequency components of the masker. The FM only condition was generated by using the time varying frequency information from the F0, without applying an envelope. The AM only condition was generated by using the extracted envelope from the F0 as in exp. 1 to amplitude modulate a sine wave with a frequency equal to the average F0 value for that sentence. The stationary condition was generated as a sine wave with a frequency equal to the average F0 for the sentence, and a flat envelope, and represented merely a presence of voicing without any of the dynamic changes normally present. An example of the FM only, AM only, and stationary conditions are shown in the bottom three panels of Fig. 1. As in exp. 1, F0 information was only present when the sentence was voiced. Since the FM only and stationary conditions did not have a speech envelope, 20 ms cosine squared ramps were applied to each onset and offset to reduce the presence of transients. The target F0 condition was renamed “FM and AM” for consistency with the three new stimuli. All stimuli were presented at 75 dB sound pressure level, as in exp. 1, with the same calibration techniques.

3. Procedure

SRTn values were determined for the three new stimuli with an identical procedure to exp. 1. All subjects were tested on all three new conditions to allow for a within-subjects analysis. Sentence lists from exp.1 were not repeated. Data were analyzed using a one factor repeated measures ANOVA with the type of additive acoustic information as the independent variable.

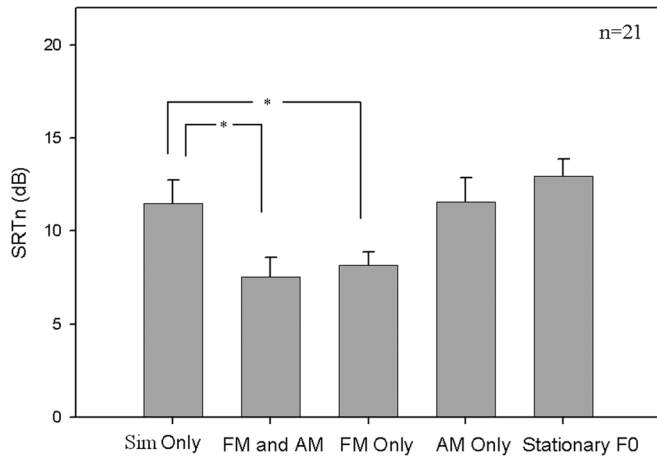


FIG. 3. Speech recognition as a function of different F0 cues in simulated combined hearing. SRTn values represent the SNR at which the subjects scored 50% correct for each stimulus condition. Asterisks indicate statistical significance ($p < 0.05$).

B. Results and discussion

Figure 3 shows the SRTn values for the five stimuli with the Sim Only and the F0 (renamed FM + AM) data being replotted from exp. 1. A main effect was seen for the type of F0 information [$F(4,80) = 9.8$, $p < 0.05$]. *Post hoc* analyses with a Newman–Keuls correction indicated that the FM only condition produced better performance than the sim. only condition (a 3-dB benefit, $p < 0.05$), compared to the 4-dB benefit seen with both FM and AM components. The FM only and FM and AM conditions were not statistically different. On the contrary, the AM only or the stationary F0 conditions failed to produce a significant benefit. These results suggest that the FM component present in the F0 was both necessary and sufficient to produce the benefit seen from the fully extracted F0 contour.

IV. EXPERIMENT 3

A. Methods

1. Subjects

Nine cochlear-implant users between the ages of 48 and 81 yr old, including five females and four males, participated in this study. Three subjects used a Cochlear Corp. Nucleus 24 device, three used an Advanced Bionics Corp. HiRes 90k, two used an Advanced Bionics Corp. Clarion CII, and one used a Med El Combi 40+. Subject experience ranged from 6 months to 6.5 yr (mean = 3.8 yr). The University of California, Irvine Institutional Review Board approved all experimental procedures. Informed consent was obtained from each subject. Subject audiograms were collected prior to testing to determine the level of contralateral residual hearing (Fig. 4). All subjects were native English speakers and were compensated for their participation.

2. Stimuli

Subjects were tested with the same sentence sets as in exps. 1 and 2. Acoustic information was presented to each subject's non-implanted ear. While the stimuli were pre-

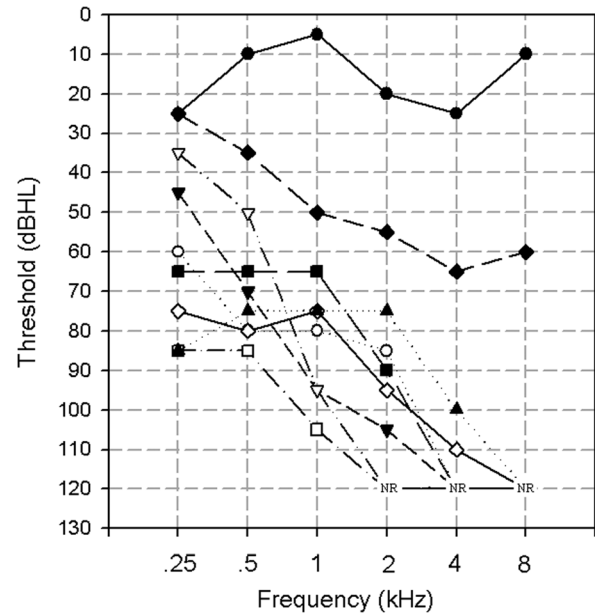


FIG. 4. Audiograms for nine cochlear-implant users with residual hearing in their contralateral ear. NR indicates no response.

sented to the same ear in simulation, previous work suggests that performance is similar between situations when acoustic cues are presented contralateral or ipsilateral to a cochlear implant (e.g., Gifford *et al.*, 2007). Acoustic stimuli were either the lowpass condition (CI + Lowpass) or the target F0 (CI + target F0) condition from exp. 1. While the lowpass condition was meant to simulate the average of the hearing loss that many people with residual low-frequency hearing may have, this filter undoubtedly restricted the performance of some subjects with significantly more residual hearing. This restriction was intended to allow for a more homogeneous population for analysis, and the final data were not meant to indicate the full range of performance that may be possible for those with residual hearing. As in the CI simulation in experiments 1 and 2, the target F0 condition did not contain any low-frequency information regarding the masker. Acoustic stimuli were presented through Sennheiser HDA 200 Audiometric headphones and amplified linearly across frequency to be presented at the subjects' most comfortable level. Subjects did not use their own hearing aids because of variations in the low-frequency response and differences in each individual's hearing aid programming. Stimuli to the CI were played through the implants' direct connect cable, also at the subjects' most comfortable level.

3. Procedure

The SRTn procedure was not used because of the inability of many implant users in pilot data to reach perfect performance for the first sentence in the test for any SNR, including quiet. Target sentences were presented in quiet or mixed with the masker sentence and presented at a fixed SNR of 10 dB. Two lists of ten sentences were presented for each condition, and lists were never duplicated for a given subject. Each condition was tested twice. Outcome measures were percent correct scores for keywords in each condition. No practice was given.

All subjects were tested on all three conditions at both SNRs to allow for a within-subjects analysis. Data were analyzed using a two-factor repeated measures ANOVA, with the noise and the type of low-frequency information as the independent variables.

B. Results and discussion

Figure 5 shows the performance for the three different conditions in quiet and in noise. Scores are plotted as the mean percent correct across subjects with the error bars indicating the standard error of the mean. In quiet, the CI only condition yielded scores ranging from 53–96%, with a mean of 75.4%. On average, the CI + F0 condition reduced the CI only performance by 2.6 percentage points, while the CI + lowpass condition improved by 3.3 percentage points. At 10-dB SNR, the average CI score dropped to 46%, but showed an improvement of 5.3 percentage points for the CI + F0 condition and 8.2 percentage points for the CI + lowpass condition. Performance for lowpass and F0 information were tested without the contralateral CI and were found to have scores of 0.75% (2 subjects identifying 4 and 2 target words out of 100) and 0.125% (1 subject identifying 1 target word out of 100), respectively.

A main effect was seen for both noise [$F(1,8) = 52.2$, $p < 0.05$] and the type of low-frequency information [$F(2,16) = 7.7$, $p < 0.05$]. An interaction was also observed [$F(2,16) = 4.1$, $p < 0.05$] indicating differential effects within the quiet and noise conditions. As such, separate ANOVAs were conducted for each of the quiet and noise conditions. In quiet, a main effect for type of low-frequency information was seen [$F(2,16) = 3.7$, $p < 0.05$]. *Post hoc* analyses with a Newman–Keuls correction showed that both the small benefit from the lowpass information and the small decrease due the F0, compared to the CI alone, were not individually significant. In noise, a main effect of type of low-frequency information was also seen [$F(2,16) = 7.9$, $p < 0.05$]. *Post hoc* analyses with a Newman–Keuls correction showed a significant benefit for both the CI + lowpass and the CI + F0 conditions ($p < 0.05$). The CI + lowpass and CI + F0 conditions

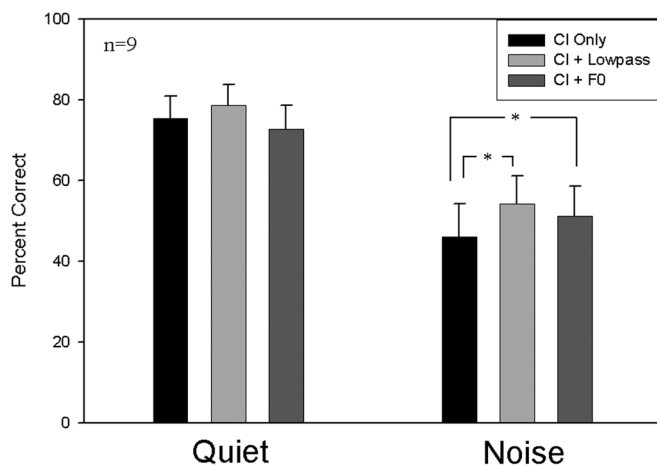


FIG. 5. Cochlear-implant speech recognition in percent-correct scores combined with three low-pass listening conditions in quiet or at a 10 dB SNR. Asterisks indicate statistical significance ($p < 0.05$).

were not statistically different from each other. The values from the lowpass alone or the CI alone were not sufficient to allow for a linearly additive explanation of the benefit seen in noise with the added acoustic information.

V. GENERAL DISCUSSION

A. Performance variability

Implant users showed a large degree of variability both in their performance as well as their improvement with added low-frequency information. To find out who benefited from low-frequency information in combination with a CI, correlations were investigated between the combined CI and low-frequency performance and the CI-only performance as well as acoustic thresholds at 250 and 500 Hz. No significant correlations were found with respect to performance and low-frequency thresholds, indicating that more residual low-frequency hearing does not necessarily provide for a greater improvement with a cochlear implant. Figure 6 shows scatter plots comparing CI only scores and the benefit to CI with low-frequency information. Positive values indicate a benefit while negative values represent worse performance with the added low-frequency information. Linear regression lines were fit to each of the four correlations and shown as a solid line. Recall from Fig. 5 that the data represented in panels C and D represent the only sets with statistically significant effects. Panels A, C, and D show correlations with slopes less than 1. This may be due to a ceiling effect in the better performers. The small number of subjects prevented the slopes from being statistically different from 1. While no claims could be made statistically, the trends suggest a need to analyze the data without the best users. Also note the greater subject performance variability in the quiet condition (three performed worse with added lowpass and seven performed worse with added F0) than in the noise condition (only one subject failed to show improvement with added lowpass or F0). This difference in subject variability is likely to account for the statistical significance of the noise condition but not the quiet condition seen in Fig. 5.

Figure 7 breaks down the performance between the good users (CI only scores in quiet better than 85%, $n = 4$) and the average users (CI only scores in quiet less than 85%, $n = 5$). One-factor ANOVA was conducted on each of the quiet and noise conditions for each subject set. No effect of low-frequency information was seen for the good users on each of the quiet or noise sets, signaling a ceiling effect. The average users showed a main effect in quiet with the addition of low-frequency information [$F(2,8) = 4.8$, $p < 0.05$]. *Post hoc* analyses with Newman–Keuls correction showed a significant improvement of 7.8 percentage points with the CI + lowpass condition over the CI only condition. This effect indicates that the failure of the group to reach significance (Fig. 5) was due to the good users' performance at ceiling. The average users also showed a main effect in noise for low-frequency information [$F(2,8) = 9.9$, $p < 0.05$]. *Post hoc* analyses with Newman–Keuls correction showed a significant improvement with both the F0 (8.9 percentage points) and Lowpass conditions (11.6 percentage points) over the CI alone condition. The present analysis suggests

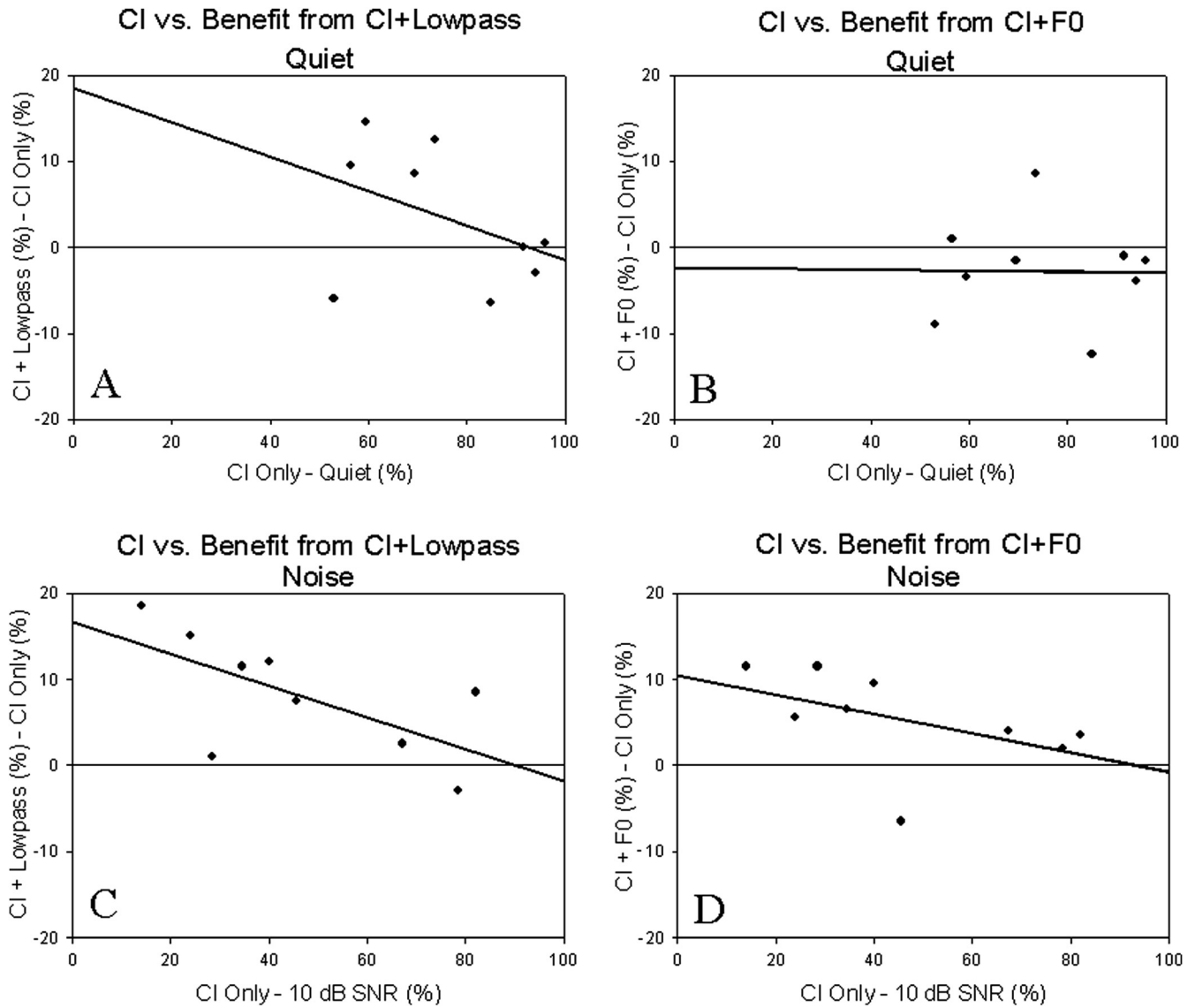


FIG. 6. Correlations between individual CI only scores and benefits of the added low-frequency information under four listening conditions. Solid lines indicate the linear regression to the data.

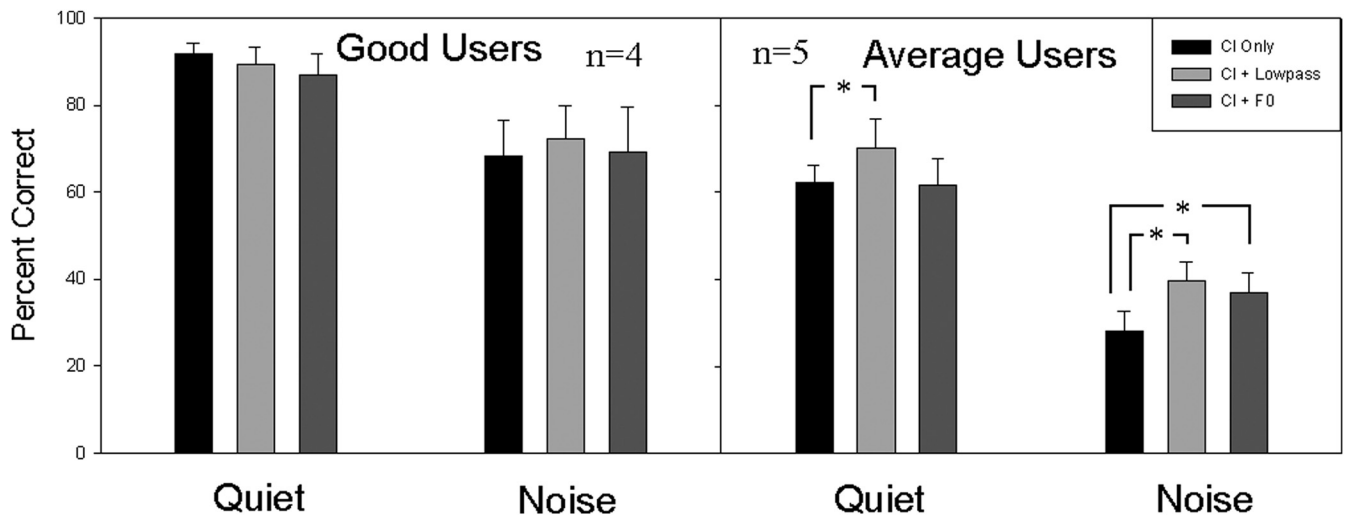


FIG. 7. Performance between good users (CI only scores in quiet better than 85%) and average users (CI only scores in quiet less than 85%). Asterisks indicate statistical significance ($p < 0.05$).

that poorer users may benefit more from the presence of low-frequency information. It is possible that good users could also benefit from the added low-frequency information at poorer SNRs, but this was not tested.

B. The role of F0

The present data showed a clear F0 benefit for both simulated and actual implant users in a sentence recognition task with a competing background voice. The presence of low-passed information at 500 Hz provided for an even greater benefit. Although the present data are in general agreement with a growing body of literature showing the importance of F0 in cochlear-implant speech recognition in competing voices, they differ from previous studies in terms of suggesting which acoustic component of the F0 cue is the most critical. For example, Kong and Carlyon (2007) showed a significant benefit of a stationary tone representing only the voicing information, while Brown and Bacon (2009a) showed a significant benefit of adding the AM component of the F0 cue. Moreover, earlier speech reading and speech perception studies showed a significant effect of AM information type and amount (e.g., Drullman *et al.*, 1994; Grant *et al.*, 1991). The present study failed to show any benefit with either the stationary or the AM information. In addition, the benefit of combined hearing observed here seemed to be less than that observed in other studies (e.g., Kong *et al.* 2005; Brown and Bacon 2009b; Zhang *et al.* 2010b). These discrepancies were most likely due to differences in signal processing, CI subjects, speech materials, task types and difficulty, SNRs, and even listening strategies (e.g., Bernstein and Grant 2009; Brown and Bacon 2009b; Chen and Loizou 2010; Zhang *et al.* 2010a). Taking together, the present study does not suggest that the AM information is not important, rather it is possible that the AM information provided by residual low frequency hearing is redundant or overlaps with that provided by the cochlear implant. It is also possible that amplitude information about F0, or merely the presence of voicing, may offer a benefit at SNRs different from that measured by SRTn in the present study.

C. Signal processing

While the present study shows that F0 can largely account for the benefit of preserving residual hearing, it addresses a theoretical issue but in no way suggests that F0 extraction be the method of signal processing for improving CI speech recognition in noise. First, the benefit over CI alone has been shown to increase as a function of bandwidth (e.g., Cullington and Zeng, 2010; Zhang *et al.*, 2010b), so sole extraction and presentation of the target F0 would not be advantageous. Second, if an implant user with residual hearing has an additional ability to extract other features such as formants or harmonics, then by all means he or she should be allowed to access the information for a greater benefit. Finally, all previously mentioned F0 studies used offline signal processing under laboratory conditions. In real-world listening situations, accurate real-time F0 extrac-

tion in the presence of multiple talkers is technically challenging, and may not be feasible.

VI. CONCLUSION

Under the present experimental conditions, where a single sentence was used to mask target sentence recognition, a significant benefit was found for both the low-pass speech and the F0 cue in simulated and actual combined acoustic and electric hearing. Specifically, the present results showed the following.

- (a) The 500-Hz low-pass speech produced a 6.3 dB benefit in functional signal-to-noise ratio for a cochlear implant simulation.
- (b) The target only F0 also provided a 2.1–4.0 dB benefit for a cochlear implant simulation. Furthermore, the FM but not the AM component accounted for the observed F0 benefit.
- (c) The F0 produced a benefit in noise but not in quiet in actual hybrid implant listeners, with poorer implant users benefiting more from the F0 cue than better users.

ACKNOWLEDGMENTS

We would like to thank all of our subjects for their dedication. We thank Ken Grant and two anonymous reviewers for their helpful comments on a previous version of the manuscript. The IEEE sentences were provided by Dr. Louis Braida at the Sensory Communication Group of the Research Laboratory of Electronics at MIT. This work was supported by the National Institutes of Health, National Institute on Deafness and Other Communication Disorders Grants Nos. R01-DC008858 and P30-DC008369.

- Assmann, P. F., and Summerfield, Q. (1990). "Modeling the perception of concurrent vowels: Vowels with different fundamental frequencies," *J. Acoust. Soc. Am.* **88**(2), 680–697.
- Bernstein, J. G., and Grant, K. W. (2009). "Auditory and auditory-visual intelligibility of speech in fluctuating maskers for normal-hearing and hearing-impaired listeners," *J. Acoust. Soc. Am.* **125**(5), 3358–3372.
- Binns, C., and Culling, J. F. (2007). "The role of fundamental frequency contours in the perception of speech against interfering speech," *J. Acoust. Soc. Am.* **122**(3), 1765–1776.
- Breeuwer, M., and Plomp, R. (1984). "Speechreading supplemented with frequency-selective sound-pressure information," *J. Acoust. Soc. Am.* **76**(3), 686–691.
- Brokx, J., and Nootboom, S. (1982). "Intonation and the perception separation of simultaneous voices," *J. Phonet.* **10**, 23–36.
- Brown, C. A., and Bacon, S. P. (2009a). "Low-frequency speech cues and simulated electric-acoustic hearing," *J. Acoust. Soc. Am.* **125**(3), 1658–1665.
- Brown, C. A., and Bacon, S. P. (2009b). "Achieving electric-acoustic benefit with a modulated tone," *Ear Hear.* **30**(5), 489–493.
- Burns, E., and Viemeister, N. (1976). "Nonspectral pitch," *J. Acoust. Soc. Am.* **60**(4), 863–869.
- Burns, E., and Viemeister, N. (1981). "Played-again SAM: further observations on the pitch of amplitude-modulated noise," *J. Acoust. Soc. Am.* **70**(6), 1655–1660.
- Carroll, J., Chen, H., and Zeng, F. G. (2005). "High-rate carriers improve pitch discrimination between 250 and 1000 Hz," *Conference on Implantable Auditory Prostheses*, Asilomar Conference Grounds, Pacific Grove, CA.
- Carroll, J., and Zeng, F. G. (2007). "Fundamental frequency discrimination and speech perception in noise in cochlear implant simulations," *Hear. Res.* **231**(1-2), 42–53.

- Chang, J. E., Bai, J. Y., and Zeng, F. G. (2006). "Unintelligible low-frequency sound enhances simulated cochlear-implant speech recognition in noise," *IEEE Trans. Biomed. Eng.* **53**, 2598–2601.
- Chen, F., and Loizou, P. C. (2010). "Contribution of consonant landmarks to speech recognition in simulated acoustic-electric hearing," *Ear Hear.* **1**(2), 259–267.
- Cullington, H., and Zeng, F. G. (2010). "Bimodal hearing benefit for speech recognition with competing voice in cochlear implant subject with normal hearing in contralateral ear," *Ear Hear.* **31**(1), 70–73.
- Dorman, M. F., Spahr, A. J., Loizou, P. C., Dana, C. J., and Schmidt, J. S. (2005). "Acoustic simulations of combined electric and acoustic hearing (EAS)," *Ear Hear.* **26**(4), 371–380.
- Drullman, R., Festen, J. M., Plomp, R. (1994). "Effect of reducing slow temporal modulations on speech reception," *J. Acoust. Soc. Am.* **95**, 2670–2680.
- Drullman, R., and Bronkhorst, A. W. (2004). "Speech perception and talker segregation: effects of level, pitch, and tactile support with multiple simultaneous talkers," *J. Acoust. Soc. Am.* **116**(5), 3090–3098.
- Freyman, R., Balakrishnan, U., and Helfer, K. (2001). "Spatial release from informational masking in speech recognition," *J. Acoust. Soc. Am.* **109**(5), 2112–2122.
- Freyman, R., Balakrishnan, U., and Helfer, K. (2004). "Effect of number of masking talkers and auditory priming on informational masking in speech recognition," *J. Acoust. Soc. Am.* **115**(5), 2246–2256.
- Geurts, L., and Wouters, J. (2004). "Better place-coding of the fundamental frequency in cochlear implants," *J. Acoust. Soc. Am.* **115**(2), 844–852.
- Gifford, R., Dorman, M., McKarns, S., and Spahr, A. (2007). "Combined electric and contralateral acoustic hearing: word and sentence recognition with bimodal hearing," *J. Speech Lang. Hear. Res.* **50**(4), 835–843.
- Grant, K., Ardell, L., Kuhl, P., and Sparks, D. (1984). "The contribution of fundamental frequency, amplitude envelope and voicing duration cues to speech-reading in normal-hearing subjects," *J. Acoust. Soc. Am.* **77**(2), 671–677.
- Grant, K. W., Braida, L. D., Renn, R. J. (1991). "Single band amplitude envelope cues as an aid to speechreading," *Q. J. Exp. Psychol. A* **43**, 621–645.
- Green, T., Faulkner, A., and Rosen, S. (2002). "Spectral and temporal cues to pitch in noise-excited vocoder simulations of continuous-interleaved-sampling cochlear implants," *J. Acoust. Soc. Am.* **112**(5), 2155–2164.
- Green, T., Faulkner, A., and Rosen, S. (2004). "Enhancing temporal cues to voice pitch in continuous interleaved sampling cochlear implants," *J. Acoust. Soc. Am.* **116**(4), 2298–2310.
- Green, T., Faulkner, A., Rosen, S., and Macherey, O. (2005). "Enhancement of temporal periodicity cues in cochlear implants: Effects on prosodic perception and vowel identification," *J. Acoust. Soc. Am.* **118**(1), 375–385.
- Greenwood, D. D. (1990). "A cochlear frequency-position function for several species—29 years later," *J. Acoust. Soc. Am.* **87**(6), 2592–2605.
- Jin, S., and Nelson, P. (2006). "Speech perception in gated noise: the effects of temporal resolution," *J. Acoust. Soc. Am.* **119**(5), 3097–3108.
- Jones, P., McDermott, H., Seligman, P., and Millar, J. (1995). "Coding of voice source information in the Nucleus cochlear implant system," *Ann. Otol. Rhinol. Laryngol. Suppl.* **166**, 363–365.
- Kawahara, H., Masuda-Katsuse, I., and de Cheveigne, A. (1999). "Restructuring speech representations using a pitch-adaptive time-frequency smoothing and an instantaneous-frequency-based F0 extraction: possible role of a repetitive structure in sounds," *Speech Commun.* **27**, 187–207.
- Kong, Y. Y., Stickney, G. S., and Zeng, F. G. (2005). "Speech and melody recognition in binaurally combined acoustic and electric hearing," *J. Acoust. Soc. Am.* **117**(3), 1351–1361.
- Kong, Y. Y., and Carlyon, R. P. (2007). "Improved speech recognition in noise in simulated binaurally combined acoustic and electric stimulation," *J. Acoust. Soc. Am.* **121**(6), 3717–3727.
- Laneau, J., Moonen, M., and Wouters, J. (2006). "Factors affecting the use of noise-band vocoders as acoustic models for pitch perception in cochlear implants," *J. Acoust. Soc. Am.* **119**(1), 491–506.
- Levitt, H. (1971). "Transformed up-down methods in psychoacoustics," *J. Acoust. Soc. Am.* **49**(2), 467–477.
- Luo, X., and Fu, Q.-J. (2006). "Contribution of low-frequency acoustic information to Chinese speech recognition in cochlear implant simulations," *J. Acoust. Soc. Am.* **120**(4), 2260–2266.
- Marin, C. M., and McAdams, S. (1991). "Segregation of concurrent sounds. II: Effects of spectral envelope tracing, frequency modulation coherence, and frequency modulation width," *J. Acoust. Soc. Am.* **89**(1), 341–351.
- McAdams, S. (1989). "Segregation of concurrent sounds. I: Effects of frequency modulation coherence," *J. Acoust. Soc. Am.* **86**(6), 2148–2159.
- McDermott, H. J., and McKay, C. M. (1994). "Pitch ranking with non-simultaneous dual-electrode electrical stimulation of the cochlea," *J. Acoust. Soc. Am.* **96**(1), 155–162.
- Qin, M. K., and Oxenham, A. J. (2006). "Effects of introducing unprocessed low-frequency information on the reception of envelope-vocoder processed speech," *J. Acoust. Soc. Am.* **119**(4), 2417–2426.
- Rosen, S., Fourcin, A. and Moore, B. (1981). "Voice pitch as an aid to lip-reading," *Nature* **291**(5811), 150–152.
- Rothausler, E. H., Chapman, W. D., Guttman, N., Hecker, M. H. L., Norby, K. S., Silbiger, H. R., Urbanek, G. E., and Weinstock, M. (1969). "IEEE recommended practice for speech quality measurements," *IEEE Trans. Audio Electroacoust.* **AU-17**(3), 225–246.
- Seebeck, A. (1841). "Beobachtungen über einige bedingungen der ent-scheidung von tonen (Observations about some conditions of the origin of tones)," *Ann. Phys. Chem.* **53**, 417–436.
- Shannon, R. V. (1983). "Multichannel electrical stimulation of the auditory nerve in man. I. Basic psychophysics," *Hear Res* **11**(2), 157–189.
- Shannon, R. V., Zeng, F.-G., Kamath, V., Wygonski, J., and Ekelid, M. (1995). "Speech recognition with primarily temporal cues," *Science* **270**(5234), 303–304.
- Shannon, R. V., Fu, Q.-J., Galvin, J. 3rd. (2004). "The number of spectral channels required for speech recognition depends on the difficulty of the listening situation," *Acta Otolaryngol. Suppl.* **552**, 50–54.
- Stickney, G., Zeng, F.-G., Litovsky, R., and Assmann, P. (2004). "Cochlear implant speech recognition with speech maskers," *J. Acoust. Soc. Am.* **116**(2), 1081–1091.
- Summerfield, Q., and Culling, J. F. (1992). "Auditory segregation of competing voices: absence of effects of FM or AM coherence," *Philos. Trans. R. Soc. London, Ser. B* **336**(1278), 357–65; **336**(1278), 365–366.
- Townshend, B., Cotter, N., Van Compernelle, D., and White, R. L. (1987). "Pitch perception by cochlear implant subjects," *J. Acoust. Soc. Am.* **82**(1), 106–115.
- Turner, C. W., Gantz, B. J., Vidal, C., Behrens, A., and Henry, B. A. (2004). "Speech recognition in noise for cochlear implant listeners: benefits of residual acoustic hearing," *J. Acoust. Soc. Am.* **115**(4), 1729–1735.
- Van Tasell, D., Soli, S., Kirby, V., and Widin, G. (1987). "Speech waveform envelope cues for consonant recognition," *J. Acoust. Soc. Am.* **82**(4), 1152–1161.
- Vandali, A. E., Sucher, C., Tsang, D. J., McKay, C. M., Chew, J. W., and McDermott, H. J. (2005). "Pitch ranking ability of cochlear implant recipients: a comparison of sound-processing strategies," *J. Acoust. Soc. Am.* **117**(5), 3126–3138.
- von Békésy, G. (1960). "Wave motion in the cochlea," in *Experiments in Hearing* (McGraw-Hill, New York), pp. 485–510.
- von Ilberg, C., Kiefer, J., Tillein, J., Pfenningdorff, T., Hartmann, R., Sturzebecher, E., and Klinke, R. (1999). "Electric-acoustic stimulation of the auditory system. New technology for severe hearing loss," *ORL J. Otorhinolaryngol. Relat. Spectrosc.* **61**(6), 334–340.
- Vongphoe, M., and Zeng, F.-G. (2005). "Speaker recognition with temporal cues in acoustic and electric hearing," *J. Acoust. Soc. Am.* **118**(2), 1055–1061.
- Zhang, T., Dorman, M., and Spahr, A. (2010a). "Information from the voice fundamental frequency (F0) region accounts for the majority of the benefit when acoustic stimulation is added to electric stimulation," *Ear Hear.* **31**(1), 63–69.
- Zhang, T., Spahr, A., and Dorman, M. (2010b). "Frequency overlap between electric and acoustic stimulation and speech-perception benefit in patients with combined electric and acoustic stimulation," *Ear Hear.* **31**(2), 195–201.