Published in final edited form as:

J Speech Lang Hear Res. 2014 April 01; 57(2): 532–554. doi:10.1044/2014_JSLHR-H-12-0404.

Perception of speech produced by native and nonnative talkers by listeners with normal hearing and listeners with cochlear implants

Caili Ji^{1,2}, John J. Galvin², Yi-ping Chang², Anting Xu¹, and Qian-Jie Fu²

Caili Ji: caili_ji@yahoo.com; John J. Galvin: jgalvin@hei.org; Yi-ping Chang: yiping.chang@gmail.com; Anting Xu: xu11110000@yahoo.com.cn

¹Department of Otolaryngology, Head and Neck Surgery, Qilu Hospital, Shandong University

²Communication and Neuroscience Division, House Ear Institute, 2100 W. 3rd St., Los Angeles, CA 90057

Abstract

Purpose—The aim of this study was to evaluate normal-hearing (NH) and cochlear implant (CI) listeners' understanding of English sentences produced by native (English) and non-native (Spanish) talkers.

Method—Sentence recognition in noise was measured in adult CI and NH subjects who were native talkers of American English. Test sentences were from the Hearing-in-Noise Test (HINT) database and were produced in English by 4 native and 8 non-native talkers. Subjects also rated the intelligibility and accent for each talker.

Results—CI and NH subjects' speech recognition thresholds in noise were 4.23-dB and 1.32-dB poorer with non-native talkers than with native talkers, respectively. Performance was significantly correlated with talker intelligibility and accent ratings for CI subjects but only correlated with talker intelligibility ratings for NH subjects. For all subjects, performance with individual non-native talkers was significantly correlated with talkers' number of years of residence in the United States.

Conclusion—CI users exhibited a larger deficit in speech understanding with non-native talkers than did NH subjects, relative to native talkers. Non-native talkers' experience with native culture contributed strongly to CI and NH listeners' speech understanding in noise, intelligibility ratings, and accent ratings.

Introduction

While speech understanding with cochlear implants (CIs) can be quite good under quiet, optimal listening conditions, CI users generally have great difficulty understanding speech in noise, identifying talkers, perceiving musical melody, and recognizing vocal emotion and prosody (Shannon et al., 2004; Vongpoe and Zeng, 2005; Galvin et al., 2007; Luo and Fu, 2005; Chatterjee and Peng, 2008). Performance in these challenging listening tasks is most

likely limited by the poor spectro-temporal resolution of the CI device. CI users also report greater difficulty understanding “conversational” speech in noise (Zhao et al., 1997), compared to the “clear” speech that is typically used to assess speech performance in the clinic. Clear speech is typically produced at a slower speaking rate, with an expanded vowel space, deeper temporal modulation, and greater F0 variability (Payton et al. 1994; Picheny et al., 1985; Krause and Braida, 2002). Liu et al. (2004) found significantly poorer speech understanding in noise with conversational speech, relative to clear speech, for CI users and for normal-hearing (NH) subjects listening to unprocessed speech or to acoustic CI simulations. Ji et al. (2012) found significantly poorer CI performance with synthetic speech, compared to normally produced speech, especially at fast speaking rates. These previous studies suggest that CI users have difficulty accommodating the acoustic variability found in everyday listening conditions.

Speech from native and non-native talkers may contain acoustic variability. Among native talkers, phonetic or phonological differences associated with different areas of the country (i.e., “dialect” or “idiolect”) may make speech understanding difficult for both NH and CI listeners. Watanabe et al. (1997) tested CI users with two nouns with dialect accents and found that CI users had difficulty in discriminating the word accent even when their dialect was taken into consideration. The acoustic variability found in non-native speech can be considered as an extreme form of native accent with different dialects (Nygaard and Pisoni, 1998). Understanding of non-native, “accented” speech is poorer than native speech (Lane, 1963; Munro and Derwing, 1995a). Previous cross-language perception studies have shown that adult listeners have difficulty in discriminating vowel and consonant contrasts, as well as recognizing words or sentences produced by non-native talkers, due to phonemic and phonological restrictions of their native language (Best et al., 2001; James, 2001; Hojen and Flege, 2006; Imasi et al., 2005; Munro, 1995). Phonemic and allophonic variations may explain some of the perceptual difficulties associated with non-native speech (Harnsberger, 2001; Levy, 2009).

As noted above, CI users have great difficulty understanding speech in noise. Understanding of non-native speech in noise would likely be even more difficult. Van Wijngaarden et al. (2002) investigated native NH listeners' understanding of speech produced by native and non-native talkers. Results showed poorer performance with non-native talkers, especially in noise. Similar findings for NH listeners have been reported by Cutler (2008) and García Lecumberri (2006). However, the effect of non-native talkers on CI users' speech perception is not well understood, especially in noise, the dominant “real-world” listening condition. In this study, sentence recognition in noise was measured in NH and CI listeners who were adult, native speakers of American English. All sentences were produced by adult native (English) talkers or non-native (Spanish) talkers. Subjective ratings of intelligibility and accent were obtained for each talker using speech materials presented in quiet. Acoustic characteristics (e.g., speaking rate, fundamental frequency, formant frequencies, etc.) were analyzed for the experimental talkers and compared to speech intelligibility. Finally, talker and subject demographics were compared to speech recognition measures.

Methods

Subjects

Fourteen post-lingually deafened adult CI users (6 men and 8 women, aged 24-80 years old) and 6 NH listeners (3 men and 3 women, aged 20-37 years old) participated in the study. All subjects were native speakers of American English. The mean age at testing was 59.6 years old for CI subjects and 29.7 years for NH subjects. CI subject demographics are shown in Table 1. CI subjects were tested with their clinical CI processors and settings; subjects were instructed to use these settings throughout the experiments. If subjects used 2 CI devices for everyday listening, subjects were tested while listening with both devices. NH pure-tone thresholds were <15 dB HL at octave frequencies from 0.25 kHz to 8 kHz, for both ears. All subjects provided informed consent in accordance with the local Institutional Review Board, and all were compensated for their participation.

Materials

Test materials consisted of the 260 sentences used in Hearing in Noise Test (HINT; Nilsson et al., 1994). HINT sentences were produced by 4 (2 male and 2 female) native (English) talkers and 8 (4 male and 4 female) non-native (Spanish) talkers. The age of English acquisition for non-native talkers ranged from 4 to 37 years old, and their duration of residence in the United States ranged from 3 to 24 years. Table 2 shows talker demographics. Throughout the paper, in reference to specific talkers, “En” = English, “Sp” = Spanish, “M” = male, and “F” = female.

Recordings were made in a sound-treated booth. The microphone (AKG 414) was placed approximately 12 inches directly in front of the talker. The microphone was connected to a USB audio interface (Edirol UA-25), which was connected to a computer. Speech materials were recorded with 16-bit resolution and 44.1 kHz sampling frequency. Sentences were normalized to have the same long-term root-mean-square (RMS) amplitude (65 dB).

Procedures

All testing was conducted in sound field. Testing was performed using custom software (i-STAR, developed by Qian-Jie Fu at House Research Institute). Subjects were seated in a sound-treated booth directly facing a single loudspeaker (Tannoy Reveal) 1 m away. Sentence recognition thresholds (SRTs) in noise – defined as the signal-to-noise ratio (SNR) that produces 50% correct whole sentence recognition – were measured using an adaptive (1-up/1-down) procedure. Speech was presented at a fixed level (65 dBA) and the noise level was adjusted according to subject response. To achieve the target SNR, the speech and noise stimuli were digitally mixed within the experimental software. The speech level was fixed and the noise level was adjusted to avoid the clipping of speech or noise signal since it is unlikely for both NH listeners and CI users that SNR level would be -20 dB or lower during SRT testing. If the noise had been fixed and the speech level had been adjusted, at high SNRs (i.e. 20 dB or higher) the speech may have been peak-clipped and therefore distorted. During testing, a sentence was randomly selected (without replacement) from within the 260-sentence stimulus set and presented to the subject. The subject repeated as many words as possible, and the experimenter scored the number of words correctly identified. If the

subject repeated the entire sentence correctly, the SNR was reduced by 2 dB and a new test sentence was presented. If the subject did not repeat the entire sentence correctly, the SNR was increased by 2 dB. A total of 8 reversals in SNR were recorded, and the final six reversals were calculated as the SRT. For each subject, a minimum of 3 SRTs was averaged to calculate the SRT for each talker.

Subjective quality ratings were obtained for each talker in terms of intelligibility and accent using the test materials presented in quiet (10 sentences from each talker). For intelligibility, subjects were asked to rate each talker a scale from 1 to 10, with 1 representing “poor” intelligibility and 10 representing “high” intelligibility. Similarly, for accent, subjects were asked to rate each talker a scale from 1 to 10, with 1 representing “light” and 10 representing “heavy” accent.

Results

Table 3 shows acoustic characteristics of the recorded sentence materials. For each talker, the mean number of words-per-second (wps) was calculated across all sentences according by dividing the number of words produced by the sentence duration. Mean F0, F1, and F2 values were calculated using Pratt software (Boersma, 2001; Boersma & Weenink, 2012). For each talker, mean F0 values were calculated across all sentences. For each talker, mean F1 and F2 values were calculated for phonemes extracted from select sentences; the same excerpts were selected for each talker, so that consistent transitions between speech segments were analyzed across talkers.

Speech production was generally slower for the recorded sentences than for the HINT materials used by Nilsson et al (1994). The mean production rate for the male talker in Nilsson et al. (1994) was 3.12 wps. The mean production rate was 2.90 wps for native talkers and 2.44 wps for non-native talkers. A one-way analysis of variance (ANOVA) with Bonferroni correction showed no significant difference in speaking rate between native and non-native talkers; however, the speaking rate was significantly slower for non-native talkers compared to the male talker from Nilsson et al. (2004) ($p < 0.05$).

F0, F1 and F2 values from the experimental talkers were compared to those for isolated phonemes from Hillenbrand et al. (1995). A one-way ANOVA showed a significant difference in F0 between male and female talkers ($p < 0.05$), but not among the native, non-native and Hillenbrand talkers. One-way ANOVAs showed no significant differences in F1 values among the native, non-native and Hillenbrand talkers. A one-way ANOVA (with Bonferroni correction) showed significant effect for talker on F2 values for /æ/ ($p = 0.009$), with higher values for the Hillenbrand talkers than for the native or non-native talkers. One-way ANOVAs for the remaining phonemes showed no significant difference in F2 values among the native, non-native and Hillenbrand talkers.

Figure 1 shows mean SRTs for individual CI subjects; subjects are ordered according to performance with native talkers. There was a wide range in SRTs across CI subjects. Mean SRTs with native talkers ranged from -3.06 dB (CI-1) to 9.06 dB (CI-14), and mean SRTs with non-native talkers ranged from -1.66 dB (CI-1) to 17.31 dB (CI-14). There was greater

across-subject variability in performance with the non-native talkers (gray bars) than with the native talkers (black bars). However, linear regression analysis showed that CI performance with non-native listeners was strongly correlated with their performance with native listeners ($r^2=0.97$, $p<0.001$).

Figure 2 shows SRT data for CI and NH subjects, as a function of talker gender and language; SRTs were averaged across talkers within each category. In general, SRTs were much better for NH subjects than for CI subjects, better with the native (English) than the non-native (Spanish talkers), and better with the female than the male talkers. There was also much greater inter-subject variability in performance for CI subjects than for NH subjects. Mean CI SRTs (across subjects and all talkers) were 1.41 dB for native talkers and 5.64 dB for non-native talkers; mean NH SRTs were -7.64 dB for native talkers and -6.32 dB for non-native talkers. Thus, the deficit with non-native speech, relative to native speech was greater for CI subjects than for NH subjects. It is possible but unlikely that the difference in noise level (the noise level in NH listeners is much higher than that in CI users) contributes to such difference in the deficits since the overall noise level is still well under the clipping level even for NH listeners. A split-plot ANOVA was performed on the data shown in Figure 1, with within-subject factors of gender (male, female) and language (native, non-native) and a between-subject factor of subject group (CI, NH). Results showed significant within-subject effects for talker gender [$F(1,18)=47.99$, $p<0.001$] and language [$F(1,18)=20.23$, $p<0.001$]; there was a significant interaction among gender, language, and subject group [$F(1,18)=13.19$, $p=0.002$]. NH performance was significantly better than CI performance [$F(1,18)=40.80$, $p<0.001$].

Figure 3 shows SRT data for CI and NH subjects, as a function of talker; mean SRTs for the native and non-native talkers is shown at far right. For both CI and NH subjects, performance was variable across talkers. A split-plot ANOVA was performed on the data shown in Figure 2, with a within-subject factor of talker and a between-subject factor of subject group (CI, NH). Results showed significant effect for talker [$F(11,198)=27.87$, $p<0.001$], and a significant interaction between talker and group [$F(11,198)=7.03$, $p=0.001$]. NH performance was significantly better than CI performance [$F(1,18)=38.76$, $p<0.001$]. The results from post-hoc Bonferroni pairwise comparisons are shown as a matrix in Table 4. Correlational analyses showed no significant relationship between SRTs and non-native talkers' age at English language acquisition ($p>0.05$ for both CI and NH subjects). However, a significant correlation was observed between SRTs and non-native talkers years of residence in the US (CI subjects: $r^2=0.68$, $p<0.001$; NH subjects: $r^2=0.62$, $p=0.035$).

Figure 4 shows CI and NH subjects' mean SRTs as a function of talker intelligibility (bottom x-axis) and accent ratings (top x-axis). For CI subjects, SRTs were strongly correlated with talker intelligibility ($r^2=0.76$, $p<0.001$) and accent ratings ($r^2=0.81$, $p<0.001$). For NH subjects, correlations were less strong, and significant only between SRT and intelligibility ($r^2=0.34$, $p<0.045$). Correlational analyses showed no significant relationships between CI users' SRTs and age at testing, age at onset of profound hearing loss, or CI experience ($p>0.05$ in all cases). Performance with different talkers was significantly correlated between CI and NH subjects ($r^2=0.53$, $p=0.007$), suggesting that the acoustic variability across talkers similarly affected both subject groups.

Discussion

The present data show that speech understanding was poorer with non-native than with native talkers, for both CI and NH subjects. The deficit in SRTs with non-native talkers relative to native talkers was nearly 3 dB greater for CI subjects than for NH subjects. CI users seem to have experienced greater difficulty with the acoustic variability within these non-native speech samples than experienced by NH listeners. Below, we discuss the results in greater detail.

Mean SRTs (across subjects and all talkers) with native talkers were 9.05 dB better for NH subjects than for CI subjects; with non-native talkers, mean SRTs were 11.96 dB better for NH subjects than for CI subjects. Thus, absolute performance was poorer for CI subjects, and CI subjects had greater difficulty with the acoustic variability in non-native speech samples. It should be noted that the mean age was 59.6 years old for CI subjects and 29.7 years for NH subjects. Age has been shown to be a factor in understanding spectrally degraded speech (as experienced by CI users) in noise (Schvartz & Chatterjee, 2008). Correlational analyses were performed between age-related demographics (age at implantation, age at profound hearing loss, and duration of CI use, all shown in Table 1) and mean performance with native and non-native performance (individual subject data shown in Fig. 1). Results showed no significant correlation between any of the age-related demographics and speech performance. Thus, while subject age may be a factor that contributed to differences in absolute performance between NH and CI subjects, subject age does not appear to be a factor within the CI group. It was difficult to observe clear differences between native and non-native speech in terms of speech segment formant frequencies, voice pitch, or speaking rate. However, as the talker accent ratings suggest, non-native speech was audibly different in terms of speaking style, relative to native speech. Differences in phonetic inventories between English and Spanish may have caused the present non-native talkers to experience phonological constraints from their native language onto speech production of the non-native language. Best et al. (1988; 1995) developed a perceptual assimilation model (PAM) to account for variability in native speech understanding by non-native listeners. According to PAM, a phoneme that is not in the listener's native language may be perceptually assimilated into the native system of phonemes in one three ways: 1) it may (incorrectly) be perceived as one of the phonemes categorized within the native language, 2) it may be perceived as an uncategorized phoneme similar to two or more phonemes within the native phonetic inventory, or 3) it may be perceived as a "non-speech" sound that has no similarity to any native phonemes. Among the present non-native Spanish-speaking talkers, the phoneme /r/ in the second syllable of a word was often poorly articulated or not articulated at all. In some cases, an additional phoneme /t/ was appended to the final fricative /s/. Consonants such as /z/ were often pronounced as /s/ (e.g., Zara). Talkers sometimes had difficulty with consonant /j/, which according to the letterform, is pronounced like a /y/ or /h/ in Spanish (e.g., Julio). In the present study, perceptual assimilation may have affected native listeners' perception of non-native speech.

Interestingly, there was a strong correlation between SRTs and non-native talkers' length of residence in the United States, but not with talkers' age at English language acquisition. Among these talkers, immersion in US culture (rather than learning English in their native

country) seemed to benefit English language production. Hogen and Flege (2006) found that most early second-language (L2) learners were able to discriminate difficult English language vowel contrasts at near native listener levels, compared with near-chance level performance by Spanish monolingual listeners. Heeren and Schouten (2008) studied the effects of short- and long-term training on L2 learning when native speakers of Dutch listened to Finnish phoneme contrasts under a pretest-training-post-test design. They found that short-term training with non-native phoneme contrasts produces minimal changes in perceptual sensitivity, while with long-term training, advanced language learners can display perceptual sensitivity comparable to that in native listeners.

As shown in Figure 3 and Table 4, there was significant variability in performance within and across native and non-native talkers. Across all talkers, performance was significantly better with the female talkers. For CI subjects, there was a 3 dB range in SRTs among the 4 native talkers. Speaking rate may explain some of the variability in intelligibility across native talkers, as performance was best with the slowest talker (EnF1) and poorest with the fastest talker (EnM1). However, CI performance with non-native talkers was poorest with SpF1, the slowest talker and best with SpM4, one of the faster non-native talkers. In this case, slower speaking rates may reflect hesitancy in production that may reduce intelligibility. While speaking rate may explain some of the variability, talker variability effects have been observed in many CI studies (Kaiser et al., 2003; Chang & Fu, 2006). Talker variability may impact cognitive resources; speaker normalization may interfere with other perceptual processes when listening to multi-talker speech (Mullennix & Pisoni, 1990). Because of the reduced spectral resolution some cross-talker acoustic differences may be interpreted as phonemic differences (Chang & Fu, 2006). However, given enough exposure (e.g., full sentences rather than isolated words or phonemes), some degree of speaker normalization is possible (Chang & Fu, 2006). In this study, sentences were presented in noise, which may have made speaker normalization difficult. CI users have been shown to perform better with one talker than another (Liu et al., 2008; Luo et al., 2005). It is possible that speech patterns from some talkers (especially females) were better than others, after CI signal processing. CI users' signal processing interacts with patient-related factors such as electrode placement, patterns of neural survival, etc. Thus, some talkers may elicit a better representation than others, and may vary across CI users. For example, there was a wide range in CI performance with talker SpF1m with SRTs ranging from -0.34 to 27.28 dB. Similarly, SRTs ranged from -1.43 to 23.55 for talker SpM2 and from -4.34 to 17.44 dB for SpF3.

Pre-processing the acoustic input (e.g., optimal spectral shifting and/warping), as in Liu et al. (2008) might improve intelligibility for these talkers, at least for some subjects. Patterns of nerve survival will ultimately determine the optimal frequency allocation for each CI patient. Focused stimulation such as tripolar stimulation may better reveal patterns of neural survival than monopolar stimulation (Bierer, 2007), which is the recommended stimulation mode for clinical fitting of CI patients' speech processors. Tripolar stimulation may also reduce channel interaction, allowing for better mapping of acoustic frequency information to independent neural populations (Landsberger et al., 2012). Even if tripolar stimulation is not feasible in clinical speech processors, focused stimulation may be used to guide the optimization of acoustic frequency allocation for monopolar stimulation.

Increased exposure would most likely benefit understanding of non-native speech. Just as non-native talkers' intelligibility was correlated with their years of residence in the US, listeners' understanding may be correlated with the amount of exposure to non-native speech. Clarke and Garrett (2004) explored the effect of exposure to non-native speech on perceptual processing speed. Results showed that processing speed was initially slower for non-native speech than for native speech, but that this deficit diminished within one minute of exposure. Weil (2001) found better understanding of Marathi-accented speech for trained listeners than for untrained listeners. For NH listeners, such training effects are not surprising. Such training may also benefit CI users. However, the limited spectral resolution in CIs may limit the tolerance for acoustic variability in speech patterns. Training in combination with optimized signal processing may help CI users better accommodate acoustic variability in speech patterns, whether from native or non-native talkers.

Note that the use of noise in this study may have influenced findings. Noise was used to reduce ceiling performance effects; noise may have also masked some (but not all) of the acoustic variability across talkers. CI performance with native and non-native talkers was highly correlated, suggesting that the effect of noise was similar for both talker groups. CI subjects' SRTs in noise were strongly correlated and intelligibility and accent ratings obtained in quiet (Fig. 4), suggesting that CI subjects' perception of acoustic variability within non-native speech persisted in quiet and in noise.

General conclusions

Sentence recognition in noise for speech produced by native (English) and non-native (Spanish) talkers was measured in native NH and CI listeners. Results showed:

1. For NH and CI subjects, performance was significantly better with native talkers than with non-native talkers, and significantly with female than with male talkers.
2. There was great inter-subject variability in CI performance, with greater variability for non-native talkers than for native talkers. CI performance with native and non-native talkers was highly correlated. There was no clear advantage for bilateral CI use over unilateral CI use.
3. Compared to NH subjects, CI subjects had greater difficulty with the acoustic variability in non-native speech.
4. Intelligibility for each non-native talker was highly variable across CI subjects. Non-native talkers' intelligibility was significantly correlated with their years of residence in the US.
5. CI and NH subjects' ratings on non-native talkers' intelligibility and accent (measured in quiet) were significantly correlated with SRTs in noise.

Acknowledgments

The authors thank all subjects who participated in the study. The authors also thank two anonymous reviewers for their helpful comments. This study was supported by NIH grant 5R01DC004993.

References

- Chang Y, Fu QJ. Effects of talker variability on vowel recognition in cochlear implants. *Journal of Speech, Language, and Hearing Research*. 2006; 49:1331–1341.
- Cutler A. Consonant identification in noise by native and non-native listeners: Effect of local context. *Journal of Acoustic Society of America*. 2008; 124(2):1264–1268.
- Hogen A, Flege JE. Early learners' discrimination of second-language vowels. *Journal of Acoustic Society of America*. 2006; 119(5):3072–3084.
- Best CT, McRoberts GW, Gooddell E. Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listeners' native phonological system. *Journal of Acoustic Society of America*. 2001; 109(2):775–794.
- Best CT, McRobert S, et al. Divergent developmental patterns for infants' perception of two nonnative consonant contrasts. *Infant Behavior and Development*. 1995; 18(3):339–350.
- Best CT, McRoberts GW, Sithole NM. Examination of perceptual reorganization for nonnative speech contrasts: Zulu click discrimination by English-speaking adults and infants. *Journal of Experimental Psychology: Human Perception and Performance*. 1988; 14(3):345–360. [PubMed: 2971765]
- Bierer JA. Threshold and channel interaction in cochlear implant users: evaluation of the tripolar electrode configuration. *Journal of Acoustic Society of America*. 2007; 121(3):1642–1653.
- Boersma P. Praat, a system for doing phonetics by computer. *Glott International*. 2001; 5(9/10):341, 345.
- Boersma, P., Weenink, D. Praat: doing phonetics by computer (Version 5.3.32) [Computer program]. 2012. Retrieved October 19, 2012 from <http://www.praat.org/>
- Chatterjee M, Peng SC. Processing F0 with cochlear implants: Modulation frequency discrimination and speech intonation recognition. *Hearing Research*. 2008; 235(1-2):143–156. [PubMed: 18093766]
- Clarke CM, Garrett MF. Rapid adaptation to foreign-accented English. *Journal of Acoustic Society of America*. 2004; 116(6):3647–3658.
- Fetterman BL, Domico EH. Speech recognition in background of cochlear implant patients. *Otolaryngology Head and Neck Surgery*. 2002; 126(3):257–263. [PubMed: 11956533]
- Friesen LM, Shannon RV, Baskent D, et al. Speech recognition in noise as a function of the number of spectral channels: Comparison of acoustic hearing and cochlear implants. *Journal of Acoustic Society of America*. 2001; 110(2):1150–1163.
- Fu QJ, Nogaki G. Noise susceptibility of cochlear implant users: the role of spectral resolution and smearing. *Journal of the Association for Research in Otolaryngology*. 2005; 6(1):19–27. [PubMed: 15735937]
- Fu QJ, Shannon RV, Wang X. Effects of noise and spectral resolution on vowel and consonant recognition: acoustic and electric hearing. *Journal of Acoustic Society of America*. 1998; 104(6):3586–3596.
- Galvin JJ III, Fu QJ, Nogaki G. Melodic contour identification in cochlear implants. *Ear and Hearing*. 2007; 28(3):302–319. [PubMed: 17485980]
- García Lecumberri ML. Effect of masker type on native and non-native consonant perception in noise. *Journal of Acoustic Society of America*. 2006; 119(4):2445–2454.
- Gould J, Lane H, Vick J. Changes in speech intelligibility of postlingually deaf adults after cochlear implantation. *Ear and Hearing*. 2001; 22(6):453–460. [PubMed: 11770668]
- Heeren WFL, Schouten MEH. Perceptual development of phoneme contrasts: How sensitivity changes along acoustic dimensions that contrast phoneme categories. *Journal of Acoustic Society of America*. 2008; 124(4):2291–2302.
- Hamzavi J, Adunka O, Baumgartner WD, Gstöettner W. Speech discrimination in noise for patients with cochlear implants. *Wien Klin Wochenschr*. 2000; 112(11):498–504. [PubMed: 10890128]
- Hillenbrand J, Gayvert RT. Identification of steady-state vowels synthesized from the Peterson and Barney measurements. *Journal of Acoustic Society of America*. 1993; 94:668–674.
- Harnsberger JD. On the relationship between identification and discrimination of non-native nasal consonants. *Journal of Acoustic Society of America*. 2001; 110(1):489–502.

- Imai S, Walley AC, Flege JE. Lexical frequency and neighborhood density effects on the recognition of native and Spanish-accented words by native English and Spanish listeners. *Journal of Acoustic Society of America*. 2005; 117(2):896–907.
- Ji C, Galvin J 3rd, Xu A, Fu QJ. Effect of speaking rate on recognition of synthetic and natural speech by normal-hearing and cochlear implant listeners. *Ear and Hearing*. 2012; 34(3):313–323.
- Kaiser AR, Kirk KI, Lachs L, Pisoni DB. Talker and lexical effects on audiovisual word recognition by adults with cochlear implants. *Journal of Speech, Language, and Hearing Research*. 2003; 46:390–404.
- Kelsall DC, Shallop JK, Burnelli T. Cochlear implantation in the elderly. *American Journal of Otolaryngology*. 1995; 16:609–615.
- Kwon BJ. Effects of electrode separation between speech and noise signals on consonant identification in cochlear implants. *Journal of Acoustic Society of America*. 2009; 126(6):3258–3267.
- Krause JC, Braida LD. Investigating alternative forms of clear speech: The effects of speaking rate and speaking mode on intelligibility. *Journal of Acoustic Society of America*. 2002; 112:2165–2173.
- Landsberger DM, Padilla M, Srinivasan AG. Reducing current spread using current focusing in cochlear implant users. *Hearing Research*. 2012; 284(1-2):16–24. [PubMed: 22230370]
- Lane H. Foreign accent and speech distortion. *Journal of Acoustic Society of America*. 1963; 35:451–453.
- Liu S, Del Rio E, Bradlow AR, et al. Clear speech perception in acoustic and electric hearing. *Journal of Acoustic Society of America*. 2004; 116(4 Pt 1):2374–2383.
- Liu C, Galvin JJ III, Fu QJ, et al. Effect of spectral normalization on different talker speech recognition by cochlear implant users. *Journal of Acoustic Society of America*. 2008; 123:2836–2847.
- Levy ES. Language experience and consonantal context effects on perceptual assimilation of French vowels by American-English learners of French. *Journal of Acoustic Society of America*. 2009; 125(2):1138–1152.
- Luo X, Fu QJ. Speaker normalization for Chinese vowel recognition in cochlear implants. *IEEE Transactions on Biomedical Engineering*. 2005; 52(7):1358–1361. [PubMed: 16042003]
- Maillet CJ, Tyler RS, Jordan HN. Change in the quality of life of adult cochlear implant patients. *Annals of Otolaryngology, Rhinology and Laryngology*. 1996; (165):31–48.
- Mullennix JW, Pisoni DB. Stimulus variability and processing dependencies in speech perception. *Perception and Psychophysics*. 1990; 47:379–390. [PubMed: 2345691]
- Munson B, Nelson PB. Phonetic identification in quiet and in noise by listeners with cochlear implants. *Journal of Acoustic Society of America*. 2005; 118(4):2607–2617.
- Munro MJ, Derwing TM. Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language Learning*. 1995; 45:73–97.
- Munro MJ, Derwing TM. Processing time, accent, and comprehensibility in the perception of native and foreign-accented speech. *Language and Speech*. 1995; 38(3):289–306. [PubMed: 8816082]
- Nelson PB, Jin SH, Carney AE, Nelson DA. Understanding speech in modulated interference: cochlear implant users and normal-hearing listeners. *Journal of Acoustic Society of America*. 2003; 113:961–968.
- Nilsson M, Soli SD, Sullivan JA. Development of the Hearing in Noise Test for the measurement of speech reception thresholds in quiet and in noise. *Journal of Acoustic Society of America*. 1994; 95:1085–1099.
- Nygaard LC, Pisoni DB. Talker-specific learning in speech perception. *Perception & Psychophys*. 1998; 60:355–376.
- Payton KL, Uchanski RM, Braida LD. Intelligibility of conversational and clear speech in noise and reverberation for listeners with normal and impaired hearing. *Journal of Acoustic Society of America*. 1994; 95:1581–1592.
- Picheny MA, Durlach NI, Braida LD. Speaking clearly for the hard of hearing: Intelligibility differences between clear and conversational speech. *Journal of Speech and Hearing Research*. 1985; 28:96–103. [PubMed: 3982003]
- Shannon RV, Fu QJ, Galvin JJ III. The number of spectral channels required for speech recognition depends on the difficulty of the listening situation. *Acta Oto-laryngologica*. 2004; (552):50–54.

- Schum, DJ, Valente M Hearing aids: standards, options, and limitations. New York: Thieme; 1996. Speech understanding in background noise; p. 368-406.
- Schvartz KC, Chatterjee M, Gordon-Salant S. Recognition of spectrally degraded phonemes by younger, middle-aged, and older normal-hearing listeners. *Journal of Acoustic Society of America*. 2008; 124(6):3972–3988.
- Uchanski RM, Choi SS, Braida LD, et al. Speaking clearly for the hard of hearing IV: Further studies of the role of speaking rate. *Journal of Speech and Hearing Research*. 1996; 39(3):494–509. [PubMed: 8783129]
- Van Wijngaarden SJ, Steeneken HJ, Houtgast T. Quantifying the intelligibility of speech in noise for non-native talkers. *Journal of Acoustic Society of America*. 2002; 112(6):3004–3013.
- Vongphoe M, Zeng FG. Speaker recognition with temporal cues in acoustic and electric hearing. *Journal of Acoustic Society of America*. 2005; 118(2):1055–1061.
- Wei C, Cao K, Jin X, Chen X, Zeng FG. Psychophysical performance and Mandarin tone recognition in noise by cochlear implant users. *Ear and Hearing*. 2007; 28(2 Suppl):62S–65S. [PubMed: 17496650]
- Weil SA. Foreign accented speech: Encoding and generalization. *Journal of Acoustic Society of America*. 2001; 109:2473(A).
- Whitmal NA, Poissant SF, Freyman RL, Helfer KS. Speech intelligibility in cochlear implant simulations: Effects of carrier type, interfering noise, and subject experience. *Journal of Acoustic Society of America*. 2007; 122(4):2376–2388.
- Watanabe Y, Miura I, Iwaki T, Kubo T. Discrimination of Word Accent and Timing by Cochlear Implant Patients. *Advances in Oto-Rhino-Laryngology*. 1997; 52:233–235. [PubMed: 9042494]
- Zhao F, Stephens SDG, Sim SW, Meredith R. The use of qualitative questionnaires in patients having and being considered for cochlear implants. *Clinical Otolaryngology*. 1997; 22:254–259. [PubMed: 9222632]

Figure 1.

Mean SRTs (across talkers) for individual CI subjects. The black bars show data for native talkers, the gray bars show data for non-native talkers. There error bars show the standard error.

Figure 2. SRTs (across subjects) for CI (left panel) and NH subjects (right panel). Data is shown for English (male and female) and Spanish (male and female) talkers. The solid line shows median performance, the dashed line shown mean performance; the error bars show the 10th and 90th percentile, and the circles show outliers.

Figure 3.

Mean SRTs (across subjects) for each talker, for CI (top panel) and NH subject (right panel); mean performance for the English and Spanish talkers is shown at far right. The solid line shows median performance, the dashed line shown mean performance; the error bars show the 10th and 90th percentile, and the circles show outliers.

Figure 4.

Mean SRTs as a function of mean talker intelligibility rating (black circles; bottom x-axis) and mean talker accent rating (white circles; top x-axis), for CI (left panel) and NH subjects (right panel). The solid lines show linear regressions between intelligibility ratings and SRTs, and the dashed lines show linear regressions between accent ratings and SRTs. r^2 and p values are shown for each correlation in the lower right corner of each plot.

Table 1

CI subject demographic information. Nuc = Nucleus; AB = Advanced Bionics, ACE = Advanced combination encoder, SPEAK = Spectral peak processor, F120 = Fidelity 120

Subject	Gender	Age at testing (yrs)	Age at onset of profound hearing loss (yrs)	Duration of implant use (yrs)	Etiology	Device	Strategy
CI-1	F	26	23	3	AN	Nuc Freedom	ACE
CI-2	F	67	40	7	Genetic	Nuc 24	ACE
CI-3	M	59	30	R: 4	Genetic	Nuc Freedom	ACE
CI-4	F	60	10	R: 16 L: 4	Otoxic	R: AB CI L: AB CII	HiRes HiRes
CI-5	F	76	40	R: 0.1 L: 20	Cochlear otosclerosis	Nuc Freedom Nuc 22	ACE SPEAK
CI-6	M	72	67	1	Unknown	Nuc Freedom	ACE
CI-7	F	53	45	R: 1 L: 4	Genetic	R: AB L: AB	F120 F120
CI-8	F	73	9	7	Genetic	AB CII	HiRes
CI-9	M	24	2	0.3	Meningitis	Nuc 5	ACE
CI-10	F	63	57	4	Cochlear otosclerosis	Nuc Freedom	ACE
CI-11	F	63	28	19	Ototoxic drug	Nuc 24	ACE
CI-12	M	58	R:20 L:42	R: 2 L: 16	Viral infection	Nuc Freedom Nuc 22	ACE SPEAK
CI-13	M	80		15	Noise	Nuc 22	SPEAK
CI-14	F	61	42	18	Tubercular meningitis	Nuc 22	SPEAK

Table 2

Talker demographic information.

Talker	Gender	Age of acquisition of English (yr)	Time in US (yr)	Country of origin	Language(s) spoken at home	Other language(s) spoken
EnM01	M	Native	Native	US	English	None
EnM02	M	Native	Native	US	English	None
EnF01	F	Native	Native	US	English	None
EnF02	F	Native	Native	US	English	None
SpM01	M	17	10	Argentina	English	None
SpM02	M	28	12	Peru	Spanish	None
SpM03	M	37	20	Argentina	Spanish	None
SpM04	M	13	24	Mexico	Spanish	None
SpF01	F	22	14	Peru	Spanish/English	None
SpF02	F	6	10	Nicaragua/Mexico	English/Spanish	French
SpF03	F	4	3	Columbia	English/Spanish	None
SpF04	F	4	N/A	N/A	N/A	None

Table 3

Acoustic characteristics for different talkers. WPS = words-per-second, F0 = fundamental frequency (Hz), F1 = first formant (Hz), F2 = second formant (Hz), En = English, Sp = Spanish, M = male, F = female

Talker	wps		/æ/	/a/	/ə/	/e/	/ɛ/	/ɜr/	/ʊ/	/i/	/o/	/U/	/A/	/u/
		F0	F1	F1	F1	F1	F1	F1	F1	F1	F1	F1	F1	F1
			F2	F2	F2	F2	F2	F2	F2	F2	F2	F2	F2	F2
EnM01	3.35	102	535	787	679	511	384	520	388	307	487	479	606	357
			1691	1133	1169	1753	2264	1359	1817	2197	1048	985	1272	1410
EnM02	2.86	128	497	713	602	498	363	451	402	268	465	448	538	319
			1588	1372	1106	1613	2155	1429	1888	2422	1017	984	1331	1103
EnF01	2.54	180	746	873	776	649	486	433	386	379	512	550	744	273
			1524	1194	1305	2027	2698	1601	1979	2584	1153	1119	1435	1238
EnF02	2.85	188	729	917	823	693	470	467	555	380	510	657	631	230
			1736	1194	1275	2002	2747	1346	2149	2791	1138	1249	1561	1051
SpM01	2.78	103	556	687	567	460	391	440	352	331	447	423	725	376
			1748	1393	1212	1790	2314	1507	1884	2066	836	933	1338	1189
SpM02	2.63	124	412	744	468	455	319	511	324	197	492	256	715	221
			1181	1339	968	1778	2518	1176	2252	2417	821	780	1270	764
SpM03	2.56	99	613	601	538	462	361	475	300	280	509	381	687	341
			1361	1036	902	1751	2142	1488	2201	2127	818	882	1220	1022
SpM04	2.68	119	538	741	607	480	412	468	468	353	488	471	583	355
			1326	1196	1054	1597	2158	1463	1979	2150	1062	1131	1182	1036
SpF01	2.19	197	783	829	548	504	503	535	402	352	571	403	776	298
			1518	1556	1006	2004	2205	1818	2561	2622	1058	946	1435	1253
SpF02	2.41	177	824	762	678	543	485	384	419	249	477	484	607	269
			1599	1169	1181	1980	2589	1522	2225	2577	1150	1128	1287	1092
SpF03	2.45	199	744	773	667	562	485	553	407	317	571	454	741	382
			1630	1270	1136	1914	2311	1722	2497	2519	1087	919	1514	1296
SpF04	2.65	172	905	851	730	606	471	530	506	399	529	600	586	307
			1841	1381	1382	2026	2721	1710	2171	2909	1117	1396	1606	1141

Table 4

Significance values for Bonferroni pairwise comparisons from a split-plot ANOVA performed on the data shown in Figure 2. Statistically significant values ($p < 0.05$) are highlighted.

	EnM01	EnM02	EnF01	EnF02	SpM01	SpM02	SpM03	SpM04	SpF01	SpF02	SpF03	SpF04
EnM01	1.000	0.004	<0.001	0.014	<0.001	<0.001	1.000	0.001	<0.001	<0.001	<0.001	0.038
EnM02	0.004	1.000	1.000	1.000	0.004	<0.001	1.000	1.000	<0.001	0.002	<0.001	1.000
EnF01	<0.001	1.000	1.000	1.000	0.025	0.007	1.000	1.000	<0.001	0.006	<0.001	1.000
EnF02	0.014	1.000	1.000	1.000	0.025	0.001	1.000	1.000	<0.001	0.003	0.002	1.000
SpM01	<0.001	0.004	0.025	0.025	1.000	1.000	<0.001	0.003	1.000	1.000	1.000	0.011
SpM02	<0.001	0.000	0.007	0.001	1.000	1.000	<0.001	<0.001	1.000	1.000	1.000	<0.001
SpM03	1.000	1.000	1.000	1.000	<0.001	<0.001	1.000	1.000	<0.001	<0.001	<0.001	1.000
SpM04	0.001	1.000	1.000	1.000	0.003	<0.001	1.000	1.000	<0.001	0.001	0.000	1.000
SpF01	<0.001	<0.001	<0.001	0.000	1.000	1.000	<0.001	1.000	<0.001	1.000	1.000	<0.001
SpF02	<0.001	0.002	0.006	0.003	1.000	1.000	<0.001	0.001	1.000	1.000	1.000	<0.001
SpF03	<0.001	<0.001	<0.001	0.002	1.000	1.000	<0.001	<0.001	1.000	1.000	1.000	<0.001
SpF04	0.038	1.000	1.000	1.000	0.011	<0.001	1.000	1.000	<0.001	<0.001	<0.001	1.000