Published in final edited form as:

Hear Res. 2010 December 1; 270(1-2): 81–88. doi:10.1016/j.heares.2010.09.005.

Effects of Spectral Shifting on Speech Perception in Noise

Tianhao Li and Qian-Jie Fu

Division of Communication and Auditory Neuroscience, House Ear Institute, 2100 W. 3rd Street, Los Angeles, CA 90057

Tianhao Li: tianhaol@gmail.com; Qian-Jie Fu: qfu@hei.org

Abstract

The present study used eight normal-hearing (NH) subjects, listening to acoustic cochlear implant (CI) simulations, to examine the effects of spectral shifting on speech recognition in noise. Speech recognition was measured using spectrally matched and shifted speech (vowels, consonants, and IEEE sentences), generated by 8-channel, sine-wave vocoder. Measurements were made in quiet and in noise (speech-shaped static noise and speech-babble at 5 dB signal-to-noise ratio). One spectral match condition and four spectral shift conditions were investigated: 2 mm, 3 mm, and 4 mm linear shift, and 3 mm shift with compression, in terms of cochlear distance. Results showed that speech recognition scores dropped because of noise and spectral shifting, and that the interactive effects of spectral shifting and background conditions depended on the degree/type of spectral shift, background conditions, and the speech test materials. There was no significant interaction between spectral shifting and two noise conditions for all speech test materials. However, significant interactions between linear spectral shifts and all background conditions were found in sentence recognition; significant interactions between spectral shift types and all background conditions were found in vowel recognition. Overall, the results suggest that tonotopic mismatch may affect performance of CI users in complex listening environments.

Keywords

Speech perception in noise; Spectral shifting

1. Introduction

Most currently available cochlear implants (CIs) make use of tonotopic representation to deliver acoustic frequency information. Acoustic signals are filtered into a series of contiguous frequency bands. In each band, the temporal envelope is extracted with a low-pass filter and used to modulate electrical pulses. The electrical pulses are then mapped and sent to different electrodes approximating the tonotopic representation found in normal hearing. Electrical pulses from low-frequency bands are delivered to electrodes toward the apex of the cochlea, and electrical pulses from high-frequency bands are delivered to electrodes toward the base of the cochlea. However, CI users generally receive tonotopically shifted and/or distorted signals, due to device- and patient-related factors. Speech recognition for tonotopically shifted and distorted speech has been investigated with both

Send Correspondence to: Tianhao Li, Ph.D., Division of Communication and Auditory Neuroscience, House Ear Institute, 2100 W. 3rd Street, Los Angeles, CA 90057, Phone: 1-213-989-6766, FAX: 1-213-413-0950, tianhaol@gmail.com.

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

acute studies and perceptual adaptation studies (Dorman et al., 1997; Shannon et al., 1998; Fu and Shannon, 1999; Rosen et al., 1999; Baskent and Shannon, 2003; Svirsky et al., 2004; Faulkner et al., 2006; Stacey and Summerfield, 2008; Li et al., 2009). The results suggest that human listeners can tolerate tonotopic mismatch to some extent under ideal listening conditions, especially after some experience.

Although speech recognition in quiet is relatively tolerant of tonotopic mismatch, speech recognition in noise may be more susceptible for several reasons. First, the demands for acoustic cues for speech recognition are different in quiet than in noise. While the spectral-temporal envelope is sufficient for speech recognition in quiet, more spectral-temporal fine structure is necessary for speech recognition in noise (Dudley, 1939; Remez et al., 1981; Shannon et al., 1995; Smith et al., 2002; Zeng et al., 2005). Furthermore, different peripheral auditory representation and central neural mechanisms may be involved in speech perception in quiet than in noise. In CI users, speech recognition scores in quiet are equivalent to normal-hearing (NH) subjects listening to an 8-channel acoustic CI simulation, while speech recognition scores in noise are equivalent to a 4-channel acoustic CI simulation or less (Zeng et al., 2005; Fu and Nogaki, 2005). Thus, some factors affecting speech recognition in noise may not affect speech recognition in quiet. For example, spectral smearing has stronger effects on speech recognition in noise than in quiet (Baer and Moore, 1993, 1994). Tonotopic mismatch may be another factor that specifically affects speech recognition in noise.

The second reason that speech recognition may be more susceptible to tonotopic mismatch in noise than in quiet is that human listeners glimpse the speech signals in the presence of background noise for speech recognition (Miller and Lickider, 1950; Howard-Jones and Rosen, 1993; Cooke, 2006). According to the definition given by Cooke, a glimpse is a spectral-temporal region that contains a reasonably undistorted view of local signal properties. Detection of these glimpses and integration of these glimpses are important parts of the glimpsing process. To effectively separate speech signals from various background sounds, more tonotopic information may be needed in the glimpses. For effective integration of the glimpses, the spectral-temporal relationships among these glimpses must remain fixed to some extent. Tonotopic mismatch may change the absolute and relative locations of these glimpses in the spectral-temporal domain. Furthermore, the frequency location of glimpses has significant effects on speech recognition in noise (Li and Loizou, 2007), and the low-frequency (<1 kHz) glimpses provide better speech recognition performance.

The third reason that speech recognition may be more susceptible to tonotopic mismatch in noise than in quiet is that sound segregation and streaming are affected by pitch perception and changes (Bregman, 1990; Brox and Nootboom, 1982; Scheffer, 1983; Assmann and Summerfield, 1990; Bird and Darwin, 1998; Oxenham, 2008). Concurrent vowel or sentence recognition can be improved when there are differences between the fundamental frequencies of the two vowels or sentences. While the studies using CI users or NH subjects listening to CI simulations showed few benefits from pitch differences in concurrent speech recognition (Qin and Oxenham, 2005; Stickney et al., 2007; Luo et al., 2009), pitch perception may still be helpful for grouping sounds from the target speech into auditory objects in CI users; pitch variation may facilitate lexical segmentation of degraded speech (Spitzer et al., 2009). Due to limited spectral and temporal resolution, CI users rely on tonotopic information and temporal envelope cues for pitch-related perception. Carlyon (1996) suggested that listeners were not able to extract two pitches from a single temporal envelope by using unresolved harmonics overlapped in the spectral domain. Fu et al. (2005) found that voice gender discrimination performance in a CI simulation dropped after a one-octave upward spectral shift, suggesting that tonotopic mismatch may affect CI users' pitch-

related perception. Therefore, both correct tonotopic information and temporal envelope cues might be important for speech perception in complex environments.

Few previous studies have examined the interactive effects of spectral shifting and background conditions on speech perception. Understanding these interactions is not only important for improving the performance of CI users in complex environments, but also maybe useful for discovering the mechanisms underlying speech perception in noise. The present study used NH subjects listening to acoustic CI simulations to test the hypothesis that tonotopic shift has a different and stronger effect on speech (vowels, consonants, and sentences) recognition in noise (speech-shaped static noise and 6-talker speech-babble) than in quiet. The results showed that speech recognition performance dropped because of noise and spectral shifting, and that the interactions between spectral shifting and background conditions depended on the degree/type of spectral shift, background conditions, and the speech test materials.

2. Methods

2.1 Subjects

The eight NH subjects included three females and five males, aged 19 to 38 years. All were native speakers of American English. The subjects had pure tone thresholds better than 20 dB HL at octave frequencies from 125 to 8000 Hz. Five had no experience with acoustic CI simulations, and three had prior experience adapting to spectrally shifted vowels for voice gender discrimination. The Institutional Review Board of St. Vincent Medical Center approved the involvement of human subjects in the present project. All subjects gave informed consent and were paid for their participation.

2.2 Speech materials

Speech test materials included vowels, consonants, and sentences. The vowel stimuli consisted of twelve medial vowels presented in the context of /h-V-d/ (i.e., had, hod, hawed, head, hayed, heard, hid, heed, hoed, hood, hud, who'd). Eight talkers, four males and four females, were randomly chosen from the large vowel set recorded by Hillenbrand et al. (1995). In total, the vowel stimulus set used in the present study had 96 tokens (12 vowels $\times$ 8 talkers) in each trial. The consonant stimuli consisted of twenty consonants presented in the format of /a-C-a/ (i.e., aba, aga, ada, aza, asha, aya, ala, ata, aja, ava, awa, acha, atha, afa, apa, aka, ara, ama, ana, asa). All consonant test tokens were originally from recordings made by Shannon et al. (1999), and were produced by three female and two male talkers. The total consonant stimulus set had 100 tokens (20 consonants $\times$ 5 talkers) in each trial. The sentence stimulus set was taken from the IEEE (1969) database, which consists of 720 low-context sentences from each talker. The recordings of one male talker were used in the present study. The 720 sentences were divided into 72 subsets, with 10 sentences in each subset for each trial. The sentences in each subset were balanced with respect to difficulty.

2.3 Signal processing and test conditions

An 8-channel sine-wave vocoder was used to generate spectrally matched or shifted speech stimuli. The input speech signal was filtered into eight frequency analysis bands (4th-order Butterworth filters with a slope of 24 dB/octave). The temporal envelope was extracted from each analysis band by half-wave rectification and low-pass filtering (4th-order Butterworth filter with corner frequency at 160 Hz). The temporal envelope from each channel was used to modulate a corresponding sine wave carrier; the sine wave carrier frequencies were varied according to the experimental conditions (Table 1). The modulated sine waves were summed and the output was adjusted to have the same long-term root-mean-square (RMS) energy as the input speech signal (65 dB).

It is generally thought that CI users received upwardly shifted spectral information because of the shallow insertions of the electrode array. Therefore, one spectral match condition and four upward spectral shift conditions were used in the present study. The overall input frequency range was 200–7000 Hz for all conditions (Table 1). The spatial distribution of center frequencies was calculated according to Greenwood's (1990) formula, assuming a 35 mm-long cochlea. Center frequencies of analysis bands were separated by 2.7 mm. For the spectral match condition, the sine-wave carriers matched the center frequencies of the analysis bands. The upward shift conditions included three linear spectral shifts. For the three linear shifts, the sine-wave carriers were higher in frequency than the center frequencies of the analysis bands; and the center frequencies of all carrier bands were shifted a constant 2 mm, 3 mm or 4 mm, in terms of cochlear distance, toward the base of the cochlea.

In an actual CI, the insertion depth of the electrode array determines the location of the most apical electrode. However, the clinical mapping will also affect the degree of spectral shift. Therefore, we also investigated an upward shift condition with compression, to compare effects of different spectral shift types. For the upward shift with compression, the overall output frequency range simulated a 16 mm-long electrode array, with 2 mm electrode spacing. The sine-wave carriers for the low-frequency channels were higher in frequency than the center frequencies of the analysis bands, and the sine-wave carriers for the high-frequency channels were lower in frequency than the center frequencies of the analysis bands. Thus, the degree of spectral mismatch, in terms of cochlear distance, gradually decreased from 3 mm for the lowest frequency channel to −2.2 mm for the highest frequency channel (Table 1).

For all spectral match and shift conditions, speech recognition was measured in quiet and in two kinds of background noise: speech-shaped static noise (constant spectrum level below 800 Hz and 10 dB/octave roll-off above 800 Hz, Fu et al. 1998) and a segment of 6-talker speech-babble, which was randomly selected from 90 s-long 6-talker babble during each trial. The speech signal and the noise were mixed at a 5 dB signal-to-noise ratio (SNR) prior to processing for the acoustic simulation. The SNR was the difference, in decibels, between the long-term RMS levels of the speech signal and the noise. After mixing the speech signal and the noise, the overall presentation level was adjusted to 65 dB SPL. The background noise commenced 500 ms before and ended 500 ms after the speech signal. The combined speech and noise were then processed by the sine-wave vocoder to generate spectrally matched and shifted speech stimuli.

2.4 Procedures

Speech recognition was measured with vocoder-processed, spectrally matched and shifted vowels, consonants, and sentences, in quiet and in noise. Subjects were seated in a double-walled, sound-treated booth and listened to the speech tokens via a loudspeaker. Because human listeners are able to adapt to spectrally shifted speech in quiet, given short-term experience, the present study reduced the variability brought by perceptual adaptation by giving subjects practice before the formal test, and by allowing them to repeat the stimulus during the test. All eight subjects took part in the vowel and consonant recognition tests. Subjects were given ~1 hour practice on spectrally matched, vocoder-processed vowels and consonants in quiet before the formal test. Each subject was tested with 15 vowel and 15 consonant test conditions (5 spectral shift conditions $\times$ 3 background conditions for each). Vowel recognition was measured in a 12 alternative forced-choice (AFC) paradigm; consonant recognition was measured in a 20 AFC paradigm. For each trial, a stimulus token was chosen randomly, without repetition, from the 96 vowel test tokens, or the 100 consonant test tokens. The presentation order of all vowel and consonant test conditions was randomized for each subject and across subjects. Following the presentation of each token,

the subject responded by pressing one of 12 buttons labeled with the possible vowel responses, or one of 20 buttons labeled with the possible consonant responses. No feedback was provided. Subjects were allowed to listen to each token up to three times and to guess if they were not sure. Subjects' speech performance was measured 2~3 times for each condition.

All eight subjects took part in sentence recognition tests. Subjects received a preview of 10 sentences for all test conditions, before the formal test. Each subject was tested with 15 sentence test conditions (5 spectral shift conditions $\times$ 3 background conditions). For each trial, a sentence subset (10 sentences) was randomly chosen from 72 subsets. The presentation order was randomized for each subject and across subjects, and no sentence was repeated. Following the presentation of each sentence, subjects were asked to repeat each sentence as accurately as possible. No feedback was provided. Subjects were allowed to listen to each sentence up to two times. Performance was quantified as the percentage of all correctly recognized words across ten sentences. Subjects' speech performance was tested with 2~3 times for each test condition.

2.4 Statistical Analyses

Two-way repeated-measures (RM) analyses of variance (ANOVAs) were used to test the effects of linear spectral shifts and background conditions, or spectral shift types and background conditions; and interactions between the two factors on speech recognition. *Post-hoc* Bonferroni t-tests were used to compare the effect of linear spectral shifts, or spectral shift types, on speech recognition in each background condition.

3. Results

3.1 Effects of linear spectral shifts on speech recognition in noise

Vowel recognition performance was consistently lower in speech-shaped noise than in quiet, and lowest in speech-babble (Fig. 1). Performance in quiet was not significantly affected by a linear spectral shift up to 3 mm, but dropped significantly under the 4 mm shift condition (Table 2). Performance in both speech-shaped noise and speech-babble declined significantly between the spectral match condition and the 3 mm linear shift condition (Table 2). Background conditions/noise conditions and linear spectral shifts had significant effects on vowel recognition performance, but there was no significant interaction between the two factors (Table 2).

Consonant recognition performance for the spectral match condition declined ~16% in speech-shaped noise and ~20% in speech-babble, compared to in quiet (Fig. 1). Under all background conditions, performances dropped significantly under the 4 mm spectral shift compared to the spectral match condition, but not under smaller spectral shifts (Table 2). Background conditions/noise conditions and linear spectral shifts had significant effects on consonant recognition performance, but there was no significant interaction between the two factors (Table 2).

As with vowel and consonant recognition, sentence recognition performance was highest in quiet and lowest with a speech-babble background, and performance in quiet was significantly affected by a 4 mm linear spectral shift (Fig. 1 and Table 2). Unlike vowel and consonant performance declining in a somewhat parallel manner across background conditions, sentence recognition in speech-shaped noise was relatively high for the spectral match condition, and dropped sharply with increasing spectral shift. In speech-babble noise, speech recognition was relatively low for the spectral match condition and showed an accelerated decline with increasing spectral shift, although the difference between the 3 and 4 mm shift conditions was not significant. Effects of background conditions/noise conditions

and linear spectral shifts on sentence recognition performance were significant, and there was a significant interaction between all background conditions and linear spectral shifts, however, the interaction between two noise conditions and linear spectral shifts was not significant (Table 2).

3.2 Effects of the spectral shift type on speech recognition in noise

Effects of linear and compressed spectral shifts on vowel recognition were not significantly different across any of the background conditions, but there were significant interactions between the type of spectral shift and background conditions (Fig. 2, Table 3). Performance in quiet was significantly better for the 3 mm linear spectral shift than for the 3 mm shift with compression, but the difference vanished in both types of background noise. Furthermore, there was no significant interaction between the type of spectral shift and noise conditions (Table 3).

Although consonant recognition performance was consistently lower for the compressed shift than for the linear shift condition, the difference was significant only in speech-shaped noise (Fig. 2, Table 3). Effects of spectral shift types were significant across all background conditions, but there was no significant interaction between spectral shift types and background conditions/noise conditions.

The types of 3 mm spectral shift on sentence recognition performance were not consistent for different background conditions (Fig. 2). Although performance with the compressed shift was slightly lower than for the linear shift in speech-shaped noise, and slightly higher in speech-babble, none of the effects were significant (Table 3). Effects of different types of spectral shift on sentence recognition were not significantly different across background conditions/noise conditions, and there was no significant interaction between the two factors.

4. Discussion

In the quiet condition, vowel and consonant recognition results of the present study corroborated results of previous studies in instances of comparable observations, which found that vowel and consonant recognition declined with linear spectral shifts >3 mm, and with a 3 mm shift with compression compared to a linear shift (Dorman et al., 1997; Fu and Shannon, 1999; Baskent and Shannon, 2003, 2007). The results were also in agreement with results of previous studies about speech perception in noise (Pickett 1957; Fu et al. 1998; Parikh and Loizou, 2005; Phatak and Allen, 2007), and vowel and consonant recognition performance in background noise was significantly worse than in quiet, and performance in speech-babble was worse than in speech-shaped noise in CI simulations.

We have hypothesized that tonotopic mismatch might have a different and stronger effect on speech recognition in noise than in quiet, due to the different demands for acoustic cues for speech perception in noise than in quiet; the glimpsing process; and the role of pitch cues. However, a notable result of the present study was that linear spectral shifts and background conditions had no significant interaction on vowel and consonant recognition, i.e., effects of linear spectral shifts on vowel and consonant recognition were not particularly important in speech-shaped noise or speech-babble. It is possible that vowel and consonant recognition in noise require relatively simple speech glimpses, which depend on one-dimensional features in the spectral-temporal domain and are not strongly affected by linear spectral shifting. For example, vowel recognition depends largely on spectral cues, and consonant recognition on temporal envelope cues (van Tasell et al., 1987; Nie et al., 2006). Moreover, the integration of the glimpses might be relatively easy, because vowel and consonant recognition occur over relatively short time scales. Another possible explanation for the lack of an interactive

effect on vowel and consonant recognition is that there were no harmonics available with the acoustic CI simulations. Pitch perception was greatly weakened and there were not enough pitch cues to support sound segregation, even for the spectral match condition, as found in previous studies (Qin and Oxenham, 2005; Luo and Fu, 2009).

Another notable result of the present study was the significant interaction of spectral shift types and background conditions on vowel perception. In quiet, vowel recognition performance for the 3 mm shift with compression was about 9% poorer than for the 3 mm linear shift, but differences between the two types of spectral shift were significantly smaller under both noise conditions (Fig. 2). It is possible that features based on absolute tonotopic information are more important for vowel recognition in noise than in quiet. For some frequency channels, especially the frequency range for second formants (F2), the degree of spectral mismatch was much smaller in the spectral shift with compression than in the linear shift, so there might have been more favorable glimpses for vowel recognition in the compressed shift. On the other hand, there was more spectral envelope distortion in the spectral shift with compression, due to the scale-transformation, possibly resulting in a roughly equivalent trade-off between the two types of spectral shift. Indeed, this possibility was supported by the percent of information transmitted (calculated with the methods introduced by Miller and Nicely, 1955) for first and second formants (F1 and F2). Table 4 shows F1 and F2 feature categories for the 12 vowels used in the information transfer analysis. In quiet, F2 information transmitted for the spectral shift with compression (77.2%) was slightly better than that for the linear shift (72.3%), and F1 information transmitted (50.2%) was significantly poorer than that for the linear shift (68.3%). However, in speech-babble, F2 information transmitted for the spectral shift with compression (54.7%) were significantly better than that for the linear shift (43.8%), and F1 information transmitted (38.8%) was not significantly different than that for the linear shift (37.1%). Because information transfer analysis based on small sample confusion matrixes (i.e., individual subjects' confusion matrixes) has instinct bias (Sagi and Svirsky, 2008), F1 and F2 information transmitted were also estimated by pooling confusion matrixes across subjects. The results showed the similar pattern. In contrast to vowel recognition, consonant recognition showed no significant interaction between spectral shift types and background conditions, probably because consonant recognition is primarily related to temporal envelope cues.

Results of the present study showed that a 4 mm linear shift significantly reduced sentence recognition performance in quiet. Similarly, Baskent and Shannon (2003) found significant effects of spectral expansion and compression up to 5 mm on sentence recognition in quiet. However, the present study showed better sentence recognition performance for the spectral match condition in background noise than some previous studies (Zeng et al., 2005; Li and Loizou, 2007), possibly because we used sine-wave vocoders, gave subjects a preview, and allowed subjects to listen to each sentence up to twice. Sine-wave vocoders can provide more periodicity and intonation information than noise vocoders and benefit sentence recognition (Souza and Rosen, 2009).

Similar with vowel and consonant recognition, there was no significant interaction between spectral shifting and two noise conditions on sentence recognition. The possible explanation for the lack of interactions between spectral shifting and noise conditions for all speech materials is that the two noise conditions might not be different enough in the spectral domain. A segment of 6-talker babble roughly approximated the speech-shaped noise in terms of the long-term spectrum and the major difference between the two noises lied in the temporal changes. In contrast to vowel and consonant recognition, linear spectral shifts and all background conditions did have a significant interactive effect on sentence recognition. However, different types of spectral shift and background conditions did not have a

significant interactive effect on sentence recognition, possibly because the shift degree of every spectral channel was not greater than 3 mm in both types of spectral shift and the spectral compression was moderate. With spectral shifts of less than 3 mm and moderate spectral compression, the drop of mean performance was less than 5% for consonant recognition and no more than 9% for vowel recognition within each background condition (Fig. 2 & 3). Consequently, the two spectral types might not be significantly different in affecting sentence recognition performance under each background condition.

Why did linear spectral shifts and background conditions have significant interactive effects on sentence recognition, but not on vowel and consonant recognition? First, spectral shifting and background noise affect both vowel and consonant recognition, so sentence recognition is correspondingly influenced. However, the combined effect of vowel and consonant recognition performance appears to be complex, rather than a simple additive function. For example, poorer vowel and consonant recognition may affect the determination of syllable boundary and lexical segmentation, and further reduce sentence recognition performance; better vowel and consonant recognition with the help of context cues may make sentence intelligibility even better.

Another possible explanation for the significant interaction between spectral shifting and background conditions only on the sentence recognition task is that more speech glimpses integrated over longer time scales are needed to recognize sentences in background noise than to recognize vowels and consonants. Spectral shifting can change the distribution of speech glimpses, especially in speech-babble, and affect the detection of the glimpses. Specifically, according to the lexical-access model proposed by Stevens (2002), the first step is to identify peaks, valleys, and discontinuities in some frequency ranges of speech signals as acoustic landmarks. Spectral shifting may affect detection of these acoustic landmarks (i.e., F1 of vowels, the glide segments) in noise, thereby, affecting lexical access and segmentation for sentence recognition. The cumulative results of missing speech glimpses and acoustic landmarks might affect the integration of speech information over a relatively long time scale, and further reducing sentence recognition performance in noise.

A third possible explanation for the significant interaction between spectral shifting and background conditions only on the sentence recognition task is that pitch cues might be important for sound segregation, sound streaming, and speech recognition in noise (Bregman, 1990; Brokx and Nöteboom, 1982; Bird and Darwin, 1998; Assmann and Summerfield, 1990; Oxenham, 2008). Only weak pitch cues were available in the present study, due to the lack of harmonics in the acoustic CI simulations. Weak pitch perception and variations may facilitate only sound streaming and lexical segmentation for sentence recognition. Spectral shifting can further destroy weak pitch perception because of providing shifted tonotopic information.

Results of the present study suggested that tonotopic mismatch might have a different effect on speech recognition in noise than in quiet by affecting vowel recognition and lexical access/segmentation. On the other hand, the spectral resolution, noise types used, and the lack of perceptual adaptation experience might influence the general applicability of the results. The present study used 8-channel, sine-wave vocoder-processed speech. Subjects could almost perfectly recognize 8-channel, sine-wave vocoder-processed sentences in quiet. Therefore, there might be a ceiling effect in the quiet condition. However, fewer spectral channels would provide fewer speech glimpses for speech recognition, and spectral shifting might have a much stronger effect on speech recognition in noise. Subjects may have partly adapted to the relatively fixed background noise used in the present study; effects of spectral shifting might be enhanced by more fluctuation and random changes in the background

noise. Finally, perceptual adaptation experience improves speech recognition for spectrally mismatched speech in quiet, and might similarly reduce effects in noise.

Results of the present study have important implications for improving the performance of CI users in complex listening environments. CI users are capable of excellent speech recognition under optimal listening conditions, but have difficulty in background noise. While reduced spectral-temporal fine information may be the major factor causing reduced performance for complex listening tasks, results presented herein suggest that tonotopic mismatch may also affect speech perception by CI users in noise. It is challenging to provide fine spectral and temporal information for CI users within the current configuration of the CI device, but the optimal clinical mapping can minimize tonotopic mismatch. The minimization of binaural tonotopic mismatch is also critical for bilateral CI users (Siciliano et al., 2010). Future studies with CI users may provide further insight into the effects of tonotopic mismatch on speech recognition by CI users in noise.

5. Conclusions

Effects of spectral shifting on speech recognition in noise were observed to test the hypothesis that tonotopic mismatch has a different and stronger effect on speech recognition in noise than in quiet. Results showed that 1) both spectral shifting and background conditions had significant effects on speech recognition, as found in previous studies; 2) for vowel recognition, there was no significant interaction of linear spectral shifts and background conditions, but spectral shift types did interact with background conditions; 3) for consonant recognition, there was no significant interaction of spectral shifting and background conditions; 4) linear spectral shifting and background conditions did have significant interaction on sentence recognition, while linear spectral shifting did not interact with only noise conditions and spectral shift types did not interact with background conditions on sentence recognition. These results taken together suggest that tonotopic mismatch may be one factor affecting speech performance of CI users in complex listening environments, as tonotopic information may be more important for vowel recognition and lexical access/segmentation in noise than in quiet.

Acknowledgments

The authors would like to thank all subjects for their attention and time. We thank three anonymous reviewers for their valuable comments on the earlier version of the paper. NIH/NIDCD R01DC004972 supported the present project.

Abbreviations

CI	Cochlear implant
F1	First formant
F2	Second formant
IEEE	Institute of Electrical and Electronics Engineers
NH	Normal hearing
RMS	Root-mean-square
RM ANOVAs	Repeated measures Analyses of variance
SNR	Signal-to-noise ratio
SSN	Speech-shaped noise

References

- Assman PF, Summerfield Q. Modeling the perception of concurrent vowels: vowels with different fundamental frequencies. *J Acoust Soc Am* 1990;88(2):680–697. [PubMed: 2212292]
- Baer T, Moore BCJ. Effects of spectral smearing on the intelligibility of sentences in noise. *J Acoust Soc Am* 1993;94(3):1229–1241.
- Baer T, Moore BCJ. Effects of spectral smearing on the intelligibility of sentences in the presence of interfering speech. *J Acoust Soc Am* 1994;95(4):2277–2280. [PubMed: 8201124]
- Baskent D, Shannon RV. Speech recognition under conditions of frequency-place compression and expansion. *J Acoust Soc Am* 2003;113(4):2064–2076. [PubMed: 12703717]
- Baskent D, Shannon RV. Combined effects of frequency compression-expansion and shift on speech recognition. *Ear Hear* 2007;28(3):277–289. [PubMed: 17485977]
- Bird, J.; Darwin, CJ. *Psychophysical and Physiological Advances in Hearing*. Whurr; London: 1998. Effects of a difference in fundamental frequency in separating two sentences.
- Bregman, AL. *Auditory Scene Analysis*. MIT Press; Cambridge, MA: 1990.
- Brokx JPL, Nootboom SG. Intonation and the perceptual separation of simultaneous voices. *J Phonetics* 1982;10(1):23–36.
- Carlyon RP. Encoding the fundamental frequency of a complex tone in the presence of a spectrally overlapping masker. *J Acoust Soc Am* 1996;99(1):517–524. [PubMed: 8568039]
- Cooke MP. A glimpse model of speech perception in noise. *J Acoust Soc Am* 2006;119(3):1562–1573. [PubMed: 16583901]
- Dorman MF, Loizou PC, Rainey D. Simulating the effect of cochlear-implant electrode insertion-depth on speech understanding. *J Acoust Soc Am* 1997;102(5):2993–2996. [PubMed: 9373986]
- Dudley H. Remaking speech. *J Acoust Soc Am* 1939;11(2):169–177.
- Faulkner A, Rosen S, Norman C. The right information may matter more than frequency-place alignment: simulations of frequency-aligned and upward shifting Cochlear Implant processors for a shallow electrode array insertion. *Ear Hear* 2006;27(2):139–152. [PubMed: 16518142]
- Fu QJ, Shannon RV, Wang X. Effects of noise and spectral resolution on vowel and consonant recognition: acoustic and electric hearing. *J Acoust Soc Am* 1998;104(6):3586–3596. [PubMed: 9857517]
- Fu QJ, Shannon RV. Recognition of spectrally degraded and frequency-shifted vowels in acoustic and electric hearing. *J Acoust Soc Am* 1999;105(3):1889–1900. [PubMed: 10089611]
- Fu QJ, Nogaki G. Noise susceptibility of cochlear implant users: the role of spectral resolution and smearing. *J Assoc Res Otolaryngol* 2005;6(1):19–27. [PubMed: 15735937]
- Fu QJ, Chinchilla S, Nogaki G, Galvin JJ. Voice gender identification by cochlear implant users: the role of spectral and temporal resolution. *J Acoust Soc Am* 2005;118(3):1711–1718. [PubMed: 16240829]
- Greenwood DD. A cochlear frequency-position function for several species –29 years later. *J Acoust Soc Am* 1990;87(6):2592–2605. [PubMed: 2373794]
- Hillenbrand J, Getty LA, Clark MJ, Wheeler K. Acoustic characteristics of American English vowels. *J Acoust Soc Am* 1995;97(5 Pt 1):3099–3111. [PubMed: 7759650]
- Howard-Jones PA, Rosen S. Unmodulated glimpsing in “checkboard” noise. *J Acoust Soc Am* 1993;93(5):2915–2922. [PubMed: 8315155]
- IEEE. *IEEE Trans Audio Electroacoust*. Vol. 17. 1969. IEEE recommended practice for speech quality measurements; p. 225–246.
- Li N, Loizou P. Factors influencing glimpsing of speech in noise. *J Acoust Soc Am* 2007;122(2):1165–1172. [PubMed: 17672662]
- Li T, Galvin JJ 3rd, Fu QJ. Interactions between unsupervised learning and the degree of spectral mismatch on short-term perceptual adaptation to spectrally shifted speech. *Ear Hear* 2009;30(2):238–249. [PubMed: 19194293]
- Luo X, Fu QJ. Concurrent-vowel and tone recognitions in acoustic and simulated electric hearing. *J Acoust Soc Am* 2009;125(5):3223–3233. [PubMed: 19425665]

- Luo X, Fu QJ, Wu HP, Hsu CJ. Concurrent-vowel and tone recognition in Mandarin-speaking cochlear implant users. *Hear Res* 2009;256(1-2):75–84. [PubMed: 19595753]
- Miller GA, Licklider JCR. The intelligibility of interrupted speech. *J Acoust Soc Am* 1950;22(2):167–173.
- Miller GA, Nicely PE. An analysis of perceptual confusions among some English consonants. *J Acoust Soc Am* 1955;27(2):338–352.
- Nie K, Barco A, Zeng FG. Spectral and temporal cues in cochlear implant speech perception. *Ear Hear* 2006;27(2):208–217. [PubMed: 16518146]
- Oxenham AJ. Pitch perception and auditory stream segregation: Implications for hearing loss and cochlear implants. *Trends in Amplification* 2008;12(4):316–331. [PubMed: 18974203]
- Parikh G, Loizou PC. The influence of noise on vowel and consonant cues. *J Acoust Soc Am* 2005;118(6):3874–3888. [PubMed: 16419830]
- Phatak S, Allen J. Consonant and vowel confusions in speech-weighted noise. *J Acoust Soc Am* 2007;121(4):2312–2326. [PubMed: 17471744]
- Pickett JM. Perception of vowels heard in noises of various spectra. *J Acoust Soc Am* 1957;29(5):613–620.
- Qin M, Oxenham AJ. Effects of envelope-vocoder processing on F0 discrimination and concurrent-vowel identification. *Ear Hear* 2005;26(5):451–460. [PubMed: 16230895]
- Remez RE, Rubin PE, Pisoni DB, Carrell TD. Speech perception without traditional speech cues. *Science* 1981;212(4497):947–949. [PubMed: 7233191]
- Rosen S, Faulkner A, Wilkinson L. Adaptation by normal listeners to upward spectral shifts of speech: implications for cochlear implants. *J Acoust Soc Am* 1999;106(6):3629–3636. [PubMed: 10615701]
- Sagi E, Svirsky MA. Information transfer analysis: A first look at estimation bias. *J Acoust Soc Am* 2008;123(5):2848–2857. [PubMed: 18529200]
- Scheffers, MTM. PhD thesis. Groningen University; 1983. Sifting vowels: Auditory pitch analysis and sound segregation.
- Shannon RV, Zeng FG, Wygonski J, Kamath V, Ekelid M. Speech recognition with primarily temporal cues. *Science* 1995;270(5234):303–304. [PubMed: 7569981]
- Shannon RV, Zeng FG, Wygonski J. Speech recognition with altered spectral distribution of envelope cues. *J Acoust Soc Am* 1998;104(4):2467–2476. [PubMed: 10491708]
- Shannon RV, Jensvold A, Padilla M, Robert ME, Wang X. Consonant recordings for speech testing. *J Acoust Soc Am* 1999;106(6):L71–74. [PubMed: 10615713]
- Siciliano CM, Faulkner A, Rosen S, Mair K. Resistance to learning binaurally mismatched frequency-to-place maps: Implications for bilateral stimulation with cochlear implants. *J Acoust Soc Am* 2010;127(3):1645–1660. [PubMed: 20329863]
- Smith ZM, Delgutte B, Oxenham AJ. Chimaeric sounds reveal dichotomies in auditory perception. *Nature* 2002;416(6876):87–90. [PubMed: 11882898]
- Smooenburg GF. Pitch perception of two-frequency stimuli. *J Acoust Soc Am* 1970;48(4):924–942. [PubMed: 5480388]
- Souza P, Rosen S. Effects of envelope bandwidth on the intelligibility of sine- and noise-vocoded speech. *J Acoust Soc Am* 2009;126(2):792–805. [PubMed: 19640044]
- Spitzer S, Liss J, Spahr T, Dorman M, Lansford K. The use of fundamental frequency for lexical segmentation in listeners with cochlear implants. *J Acoust Soc Am* 2009;125(6):EL236–241. [PubMed: 19507928]
- Stacey PC, Summerfield AQ. Comparison of word-, sentence-, and phoneme-based training strategies in improving the perception of spectrally distorted speech. *J Speech Lang Hear Res* 2008;51(2):526–528. [PubMed: 18367694]
- Stevens KN. Toward a model for lexical access based on acoustic landmarks and distinctive features. *J Acoust Soc Am* 2002;111(4):1872–1891. [PubMed: 12002871]
- Stickney G, Assmann P, Chang J, Zeng FG. Effects of implant processing and fundamental frequency on the intelligibility of competing sentences. *J Acoust Soc Am* 2007;122(2):1069–1078. [PubMed: 17672654]

- Svirsky MA, Silveira A, Neuburger H, Teoh SW, Suarez H. Long-term auditory adaptation to a modified peripheral frequency map. *Acta Otolaryngol* 2004;124(4):381–386. [PubMed: 15224858]
- Van Tasell DJ, Soli SD, Kirby VM, Widin GP. Speech waveform envelope cues for consonant recognition. *J Acoust Soc Am* 1987;82(4):1152–1161. [PubMed: 3680774]
- Zeng FG, Nie K, Stickney GS, Kong YY, Vongphoe M, Bhargave A, Wei C, Cao K. Speech recognition with amplitude and frequency modulations. *Proc Natl Acad Sci USA* 2005;102(7):2293–2298. [PubMed: 15677723]

Figure 1. Vowel, consonant, and sentence recognition performance (mean $\pm$ S.E.) in quiet, in speech-shaped noise (SSN), and in speech-babble, as a function of the degree of linear spectral shift.

Figure 2.

Vowel, consonant, and sentence recognition performance (mean $\pm$ S.E.) for the 3 mm linear spectral shift (solid bars) and the 3 mm spectral shift with compression (open bars), in quiet, in speech-shaped noise (SSN), and in speech-babble.

Table 1

Analysis band and sine-wave carrier parameters used to generate spectrally matched and shifted speech stimuli

Analysis bands				Sine-wave carriers												
Number of Channel	Corner frequency (Hz)	Center frequency (Hz)	Greenwood Distances From apex (mm)	2 mm linear shift			3 mm linear shift			4 mm linear shift			3 mm shift with compression			Amount of shift toward base (mm)
				Center frequency (Hz)	Center frequency (Hz)	Center frequency (Hz)	Center frequency (Hz)	Center frequency (Hz)	Center frequency (Hz)	Center frequency (Hz)	Center frequency (Hz)	Center frequency (Hz)	Center frequency (Hz)	Center frequency (Hz)		
1	200-359	272	6.7	405	486	580	486	486	580	486	3					
2	359-591	464	9.4	658	777	914	777	777	914	688	2.3					
3	591-930	744	12.2	1028	1202	1401	1202	1202	1401	953	1.5					
4	930-1426	1155	14.9	1568	1822	2114	1822	1822	2114	1302	0.8					
5	1426-2149	1754	17.7	2358	2728	3156	2728	2728	3156	1763	0.0					
6	2149-3205	2629	20.4	3509	4051	4673	4051	4051	4673	2370	-0.7					
7	3205-4748	3905	23.2	5192	5983	6891	5983	5983	6891	3171	-1.4					
8	4748-7000	5769	26.9	7650	8805	10131	8805	8805	10131	4227	-2.2					

Table 2

Results of statistical analysis about linear spectral shifting effects on vowel, consonant, and sentence recognition under three background conditions.

	Vowels	Consonants	Sentences
Two-way RM ANOVAs (Factor 1: linear spectral shifts; Factor 2: three background conditions)			
Effects of linear spectral shifts	F (3,42) = 52.93 P < 0.001	F (3,42) = 32.16 P < 0.001	F (3,42) = 26.56 P < 0.001
Effects of background conditions	F (2,42) = 53.37 P < 0.001	F (2,42) = 109.98 P < 0.001	F (2,42) = 136.68 P < 0.001
Interactions between the two factors	F (6,42) = 1.94 P = 0.097	F (6,42) = 1.04 P = 0.417	F (6,42) = 8.83 P < 0.001
Two-way RM ANOVAs (Factor 1: linear spectral shifts; Factor 2: two noise conditions)			
Effects of linear spectral shifts	F (3,21) = 32.78 P < 0.001	F (3,21) = 25.4 P < 0.001	F (3,21) = 24.70 P < 0.001
Effects of noise conditions	F (1,21) = 60.35 P < 0.001	F (1,21) = 15.34 P = 0.006	F (1,21) = 294.62 P < 0.001
Interactions between the two factors	F (3,21) = 2.77 P = 0.067	F (3,21) = 0.824 P = 0.496	F (3,21) = 2.52 P = 0.086
Bonferroni t-tests for comparisons among linear spectral shift conditions in each background condition			
Quiet	0 mm vs. 2 mm: P = .559	0 mm vs. 2 mm: P = 1.00	0 mm vs. 2 mm: P = 1.00
	0 mm vs. 3 mm: P = .052	0 mm vs. 3 mm: P = .608	0 mm vs. 3 mm: P = 1.00
	0 mm vs. 4 mm: P < .001	0 mm vs. 4 mm: P < .001	0 mm vs. 4 mm: P < .001
	2 mm vs. 3 mm: P = 1.00	2 mm vs. 3 mm: P = 1.00	2 mm vs. 3 mm: P = 1.00
	2 mm vs. 4 mm: P < .001	2 mm vs. 4 mm: P < .001	2 mm vs. 4 mm: P < .001
	3 mm vs. 4 mm: P < .001	3 mm vs. 4 mm: P = .011	3 mm vs. 4 mm: P < .001
Speech-shaped noise	0 mm vs. 2 mm: P = .229	0 mm vs. 2 mm: P = .526	0 mm vs. 2 mm: P = 1.00
	0 mm vs. 3 mm: P = .02	0 mm vs. 3 mm: P = .526	0 mm vs. 3 mm: P = .207
	0 mm vs. 4 mm: P < .001	0 mm vs. 4 mm: P < .001	0 mm vs. 4 mm: P < .001
	2 mm vs. 3 mm: P = 1.00	2 mm vs. 3 mm: P = 1.00	2 mm vs. 3 mm: P = 1.00
	2 mm vs. 4 mm: P < .001	2 mm vs. 4 mm: P < .001	2 mm vs. 4 mm: P < .001
	3 mm vs. 4 mm: P < .001	3 mm vs. 4 mm: P < .001	3 mm vs. 4 mm: P = .003
Babble	0 mm vs. 2 mm: P = 1.00	0 mm vs. 2 mm: P = .128	0 mm vs. 2 mm: P = .102
	0 mm vs. 3 mm: P = .006	0 mm vs. 3 mm: P = .111	0 mm vs. 3 mm: P < .001
	0 mm vs. 4 mm: P < .001	0 mm vs. 4 mm: P < .001	0 mm vs. 4 mm: P < .001
	2 mm vs. 3 mm: P = .013	2 mm vs. 3 mm: P = 1.00	2 mm vs. 3 mm: P = .159
	2 mm vs. 4 mm: P < .001	2 mm vs. 4 mm: P < .001	2 mm vs. 4 mm: P < .001
	3 mm vs. 4 mm: P = .062	3 mm vs. 4 mm: P < .001	3 mm vs. 4 mm: P = .145

Table 3

Results of statistical analysis of effects of linear and compressed spectral shifts on vowel, consonant, and sentence recognition under three background conditions.

	Vowels	Consonants	Sentences
Two-way RM ANOVAs (Factor 1: the 3 mm linear shift and the 3 mm shift with compression; Factor 2: three background conditions)			
Effects of spectral shift types	$F(1,14) = 0.319$ $P = 0.59$	$F(1,14) = 7.17$ $P = 0.032$	$F(1,14) = 0.08$ $P = 0.789$
Effects of background conditions	$F(2,14) = 38.22$ $P < 0.001$	$F(2,14) = 64.61$ $P < 0.001$	$F(2,14) = 43.40$ $P < 0.001$
Interactions between the two factors	$F(2,14) = 12.82$ $P < 0.001$	$F(2,14) = 0.749$ $P = 0.491$	$F(2,14) = 2.13$ $P = 0.155$
Two-way RM ANOVAs (Factor 1: the 3 mm linear shift and the 3 mm shift with compression; Factor 2: two noise conditions)			
Effects of spectral shift types	$F(1,7) = 0.917$ $P = 0.37$	$F(1,7) = 7.79$ $P = 0.027$	$F(1,7) = 0.113$ $P = 0.746$
Effects of noise conditions	$F(1,7) = 45.35$ $P < 0.001$	$F(1,7) = 17.57$ $P = 0.004$	$F(1,7) = 25.51$ $P = 0.001$
Interactions between the two factors	$F(1,7) = 1.41$ $P = 0.273$	$F(1,7) = 0.99$ $P = 0.352$	$F(1,7) = 3.27$ $P = 0.113$
Bonferroni t-tests for comparisons between the 3 mm linear shift and the 3 mm shift with compression conditions in each background condition			
Quiet	$P = 0.008$	$P = 0.209$	$P = 0.916$
Speech-shaped noise	$P = 0.869$	$P = 0.013$	$P = 0.111$
Babble	$P = 0.154$	$P = 0.149$	$P = 0.317$

Table 4

F1 and F2 feature categories for the 12 vowels

	F1 (height)	F2 (place)
HAD	<i>Low</i>	<i>Front</i>
HOD	<i>Low</i>	<i>Back</i>
HAWED	<i>Middle</i>	<i>Back</i>
HEAD	<i>Middle</i>	<i>Front</i>
HAYED	<i>Middle</i>	<i>Front</i>
HEARD	<i>Middle</i>	<i>Central</i>
HID	<i>High</i>	<i>Front</i>
HEED	<i>High</i>	<i>Front</i>
HOED	<i>Middle</i>	<i>Back</i>
HOOD	<i>High</i>	<i>Back</i>
HUD	<i>Middle</i>	<i>Central</i>
WHO'D	<i>High</i>	<i>Back</i>