Recognition of interrupted sentences under conditions of spectral degradation

Monita Chatterjee[a] **and Fabiola Peredo**
Cochlear Implants and Psychophysics Laboratory, Department of Hearing and Speech Sciences, The University of Maryland, College Park, Maryland 20742
mchatterjee@hesp.umd.edu, fperedo@hesp.umd.edu

Desirae Nelson
Molecular Biology, Biochemistry, and Bioinformatics Program, Towson University, Towson, Maryland 21252
dnelso2@students.towson.edu

Deniz Başkent
Department of Otorhinolaryngology/Head and Neck Surgery, University Medical Center Groningen, Groningen, The Netherlands and School of Behavioral and Cognitive Neuroscience, University of Groningen, Groningen, The Netherlands
d.baskent@med.umcg.nl

Abstract: Cochlear implant (CI) and normally hearing (NH) listeners' recognition of periodically interrupted sentences was investigated. CI listeners' scores declined drastically when the sentences were interrupted. The NH listeners showed a significant decline in performance with increasing spectral degradation using CI-simulated, noise-band-vocoded speech. It is inferred that the success of top-down processes necessary for the perceptual reconstruction of interrupted speech is limited by even mild degradations of the bottom-up information stream (16 and 24 band processing). A hypothesis that the natural voice-pitch variations in speech would help in the perceptual reconstruction of the sentences was not supported by experimental results.

PACS numbers: 43.66.Ts, 43.71.Es, 43.71.Ky [JH]
Date Received: October 6, 2009 **Date Accepted:** December 4, 2009

1. Introduction

When normally hearing (NH) individuals listen to speech that has been periodically interrupted by silence, their performance declines relative to that with intact speech; however, speech recognition can sometimes still be remarkably good under these conditions (Miller and Licklider, 1950). If noise, or some other *plausible* masker, is now placed in the silent gaps, an illusory continuity in the target speech is perceived, and performance may actually improve (Miller and Licklider, 1950; Warren and Obusek, 1971; Powers and Wilcox, 1977; Bashford and Warren, 1987). This phenomenon is termed "phonemic restoration." The ability of the auditory system to recognize interrupted speech, the illusion of continuity in the presence of noise, and the ability to correctly reconstruct the interrupted speech clearly involve top-down processing, where the brain uses *a priori* knowledge to "fill in" the missing pieces.

It appears that this process of top-down restoration is more difficult when the target speech has been spectrally degraded or when the listener is hearing-impaired or uses a cochlear implant (CI). Thus, Nelson and Jin (2004) reported that CI listeners had great difficulty recognizing periodically interrupted sentences with gating frequencies from 1 to 32 Hz. The NH listeners attending to noise-band-vocoded speech also had difficulty with the task, although increasing the spectral resolution from 4 to 12 bands improved performance considerably. Nelson and Jin

[a]Author to whom correspondence should be addressed.

speculated that the lack of F0 resolution in spectrally degraded speech might contribute to the difficulty. Başkent (2007) also observed that CI listeners, in particular, had great difficulty in recognizing interrupted sentences, even when the interruptions were very brief. Additionally, Başkent *et al.* (in press) reported that hearing-impaired listeners had greater difficulty than NH listeners listening to interrupted sentences. Even with moderate levels of impairment, listeners were no longer able to benefit from phonemic restoration. These results suggest that the process of "filling in" is more successful when the bottom-up information is spectrally intact. In the present study, a particular focus of interest was in NH listeners' performance under conditions of *mild* spectral degradation [16 and 24 channels of noise-band-vocoded (NBV) speech, which normally allow near-perfect performance in sentence recognition tasks in quiet]. A group of CI listeners were also tested to confirm previous findings: only high-performing CI users were recruited for participation. As previous studies (e.g., Friesen *et al.*, 2001) have shown that the best CI listeners do not receive more than eight channels of spectral information, the NH listeners were also tested with eight-channel NBV speech to allow for a best-case comparison. In addition, the role of the fundamental frequency (F0) contour in listeners' ability to reconstruct the missing pieces of speech was investigated. Specifically, we hypothesized that speech intonation cues might contribute to auditory processes with which segments of interrupted speech are "strung together." Flattening the pitch contour has been shown to reduce accuracy in sentence recognition (Laures and Weismer, 1999). Here, it was hypothesized that the pitch contour helps NH listeners to "connect the dots," as it were, at either end of the intervening silence, and thus contribute to a sense of continuity in the sentence.

2. Methods

2.1 Participants

Two groups of listeners participated in this study. The first group consisted of 12 NH individuals with pure-tone thresholds less than or equal to 20 dB HL (hearing level) at audiometric frequencies of 250–8000 Hz for both ears. Their ages ranged from 20 to 30 years (mean=22). The second group consisted of six post-lingually deafened CI listeners (all users of the Cochlear Freedom or N24 devices). These participants were between ages 21–71 years (mean=60). All participants were monolingual native speakers of English.

2.2 Stimuli

Speech stimuli consisted of Hearing in Noise Test (HINT) sentence materials spoken by a single male talker (Nilsson *et al.*, 1994). PRAAT (Boersma, 2001) was used to create flat (100 Hz) F0 contours. The original HINT sentences and flattened F0-contour sentences were then noise-band-vocoded (NBV) offline into 8-, 16-, and 24-channel noise-bands using the TIGERCIS program developed by Qian-Jie Fu, Tigerspeech Technology, House Ear Institute (Los Angeles, CA; Fu, 2006). Noise-band vocoding methods were similar to those used by Shannon *et al.* (1995). The lowpass filter cutoff for temporal envelope extraction was 400 Hz (24 dB/oct) to preserve usable F0-based intonation cues from the male talker (Chatterjee and Peng, 2008; Peng *et al.*, 2009). Signal gating was applied (using MATLAB) to the NBV sentences (50% duty cycle, 100 ms on, 100 ms off, or 5 Hz gating, and 10 ms cosine-tapered rise/fall times). Preliminary work showed that the 5 Hz gating provides reasonable baseline performance with full-spectrum speech while avoiding floor and ceiling effects. Stimuli were routed from the computer sound card through a mixer and an amplifier to a single loudspeaker within the sound-attenuating booth.

2.3 Procedure

The 12 NH listeners were randomly assigned to two equal-sized groups. One group heard the original HINT sentences, while the others heard the HINT sentences with the flattened F0 contour. The task consisted of keyword-in-sentence recognition and was controlled using the I-STAR program developed by Qian-Jie Fu (Tigerspeech Technology, House Ear Institute). Sentences were presented at 65 dBA via a single loudspeaker located directly in front of the listener at a distance of 1 m. The participants were instructed to repeat back what they heard after hearing each

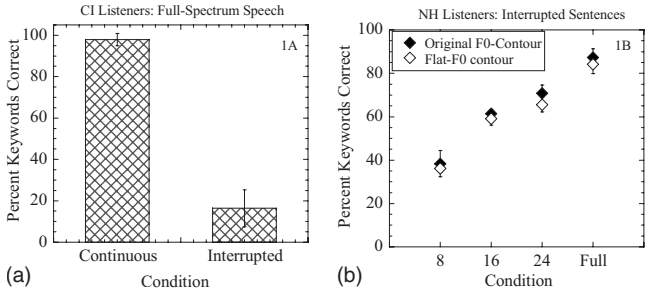

Fig. 1. (a) Sentence recognition scores (percent keyword correct) obtained by the CI listeners attending to continuous (left hand bar) and interrupted (right hand bar) sentences. Error bars show ±1 s.d. from the mean. (b) Recognition of interrupted sentences by NH listeners as a function of the degree of spectral resolution. Open and filled symbols represent scores obtained with the original and flattened F0 contours, respectively. Error bars show ±1 s.d. from the mean.

sentence. The experimenter sat outside the soundproof booth and manually marked the correct words on the computer screen. The percentage of key words correctly understood by the listener was automatically calculated by the computer program and written into a log file to be analyzed offline.

Normal-hearing listeners completed practice runs of two lists of interrupted sentences (ten sentences in each list) in each of the following four conditions: full-spectrum and 8, 16, and 24 channels of NBV speech. After the practice trial was completed, participants listened to one of the two sets of interrupted sentences (flattened F0-contour HINT or original HINT), again processed to have four levels of spectral degradation (8, 16, and 24 channels and full-spectrum). In each spectral processing condition, four lists were presented to the listener. Thus, NH participants listened to a total of 24 lists of sentences (including the 8 lists for practice trials). The order of presentation of stimulus type and condition was randomized for each participant. Care was taken to ensure that no listener heard the same sentence twice.

The CI users were tested with the original (full-spectrum) HINT sentences, either in the continuous or interrupted form, while wearing their everyday, clinically assigned speech processors. These listeners had the same practice as the NH participants and they also completed four trials of each stimulus type.

3. Results

Figure 1(a) shows the mean performance of six CI listeners attending to continuous and interrupted sentences with no F0 manipulation. It is apparent that (i) these listeners are excellent performers in open set sentence recognition and (ii) interrupting the sentences has a drastic effect on their performance. No notable differences in performance were observed between the youngest CI listeners (21 years old) and the others.

The filled symbols in Fig. 1(b) show mean results obtained with the six NH listeners who listened to the original-F0 HINT sentences, plotted as a function of the degree of spectral degradation. Note that even with eight channels of spectral information, the NH listeners' performance was better than that of the CI listeners. As the amount of spectral information increased, performance improved to about 87% correct with full-spectrum speech. However, even with 24 and 16 channels of spectral information, the NH listeners were only able to obtain on average 71% and 61% correct, respectively. With 16-channel uninterrupted speech, NH listeners are able to achieve near-perfect performance with ease (e.g., Friesen *et al.*, 2001). At best, if performance continued to improve at the same rate with increasing numbers of channels (i.e., 10% improvement per 8 channels or about 1.25% improvement per added channel), listeners would require about 37 channels to reach the same level of performance as with full-spectrum speech.

The open symbols in Fig. 1(b) show the results obtained with the listeners who attended to the flat-F0 HINT sentences. It is apparent that flattening the pitch contour had no appreciable impact on performance. A 4×2 repeated-measures analysis of variance was con-

ducted on the results obtained with the NH listeners, with the level of spectral information as a within-subjects factor and F0 contour as a between subjects factor. Not surprisingly, while the effect of spectral condition was highly significant [$F(3,30)=167.78$, $p<0.001$], no effect of F0 contour was observed, and no interaction between the two factors was found.

4. Discussion

The results presented here suggest that the top-down process that is presumably involved in filling in the missing portions of speech in interrupted-sentence recognition is significantly limited even with mild levels of spectral degradation. Thus, some minimal amount of bottom-up information is necessary for the process to succeed. It is somewhat surprising that even with 24 channels of spectral information, which would normally result in near-perfect speech recognition by NH listeners (e.g., Friesen *et al.*, 2001), the ability to restore the missing speech is impaired to the extent observed in these experiments. Even the relatively high degree of contextual cues available in HINT sentences did not appear to offset the negative effects of spectral degradation.

A second finding was that the intonation contour did not contribute to restoration of the interrupted speech under any level of spectral resolution. With 16–24 spectral channels and a 400 Hz envelope filter cutoff, sufficient F0 information is present in the speech signal to allow for good performance in tasks such as gender recognition and F0-based intonation recognition (e.g., Fu *et al.*, 2004; Gonzalez and Oliver, 2005; Chatterjee and Peng, 2008; Peng *et al.*, 2009). Thus, if the F0 contour is important in the process of reconstructing interrupted speech, we would minimally expect it to help in the perception of 16-channel, 24-channel, and full-spectrum speech and to some extent with 8-channel speech. The fact that flattening the contour did not influence performance in the task suggests that the continuity of the intonation pattern is not a contributing factor in the internal reconstruction of the interrupted speech. We note here that in unpublished preliminary results, we observed that introducing linearly increasing and decreasing F0 contours (one octave rise or fall from beginning to end) also did not impact performance in the task.

The results obtained with the CI listeners confirm the findings of Nelson and Jin (2004). The fact that even the high-performing CI patients recruited for the study found the task so difficult and underscores the difficulty of hearing with the prosthesis in everyday life. It is reasonable to speculate that part of the reason that CIs are successful is that top-down processes play an important role in reconstructing what is lost to poor spectral resolution and other potential degradations that occur in the sound transmission of the device. Further demands upon such processes, such as those that might be placed when attending to interrupted speech or speech in noise, may tax available cognitive resources excessively in most listeners.

The present results have important implications for HI listeners as well. Given the impact of even mild spectral degradation on performance with the task, it is to be expected that even listeners with mild or moderate hearing impairment would have difficulty with the perceptual restoration of interrupted sentences. Indeed, results obtained by Baskent *et al.* (in press) indicate that mildly and moderately HI listeners have increased difficulty in such tasks relative to NH individuals. Taken together, these results demonstrate clear limitations of the top-down process under conditions of mild spectral degradation and underscore the need for improved bottom-up information in auditory prostheses.

Acknowledgments

This work was supported by NIDCD Grant No. R01 DC004786 to M.C. We thank Qian-Jie Fu for the use of the software in the experiments. We are grateful to the participants for their listening time. The comments of the Associate Editor, two anonymous reviewers, and Kara C. Schvartz were very helpful in editing the manuscript.

References and links

Bashford, J. A., Jr., and Warren, R. M. (**1987**). "Multiple phonemic restorations follow the rules for auditory induction," Perception and Psychophyics **42**, 114–121.

Başkent, D. (**2007**). "Effects of amplitude ramps on phonemic restoration," in Proceedings of the 2007 Conference on Implantable Auditory Prostheses, July 16–20, Lake Tahoe, CA.

Başkent, D., Eiler, C. L., and Edwards, B. (in press, **2009**). "Perceptual restoration of speech by hearing-impaired listeners with mild to moderate sensorineural hearing loss," Hear. Res.
Boersma, P. (**2001**). "Praat, a system for doing phonetics by computer," Glot International **5**, 341–345.
Chatterjee, M., and Peng, S. C. (**2008**). "Processing F0 with cochlear implants: Modulation frequency discrimination and speech intonation recognition," Hear. Res. **235**, 143–156.
Friesen, L. M., Shannon, R. V., Başkent, D., and Wang, X. (**2001**). "Speech recognition in noise as a function of the number of spectral channels: Comparison of acoustic hearing and cochlear implants," J. Acoust. Soc. Am. **110**, 1150–1163.
Fu, Q.-J. (**2006**). TigerSpeech technology: Innovative speech software, version 1.05.02, available from http://www.tigerspeech.com/tst_tigercis.html/ (Last viewed Sept. 15, 2009).
Fu, Q.-J., Chinchilla, S., and Galvin, J. J. (**2004**). "The role of spectral and temporal cues in voice gender discrimination by normal-hearing listeners and cochlear implant users," J. Assoc. Res. Otolaryngol. **5**, 253–260.
Gonzalez, J., and Oliver, J. C. (**2005**). "Gender and speaker identification as a function of the number of channels in spectrally reduced speech," J. Acoust. Soc. Am. **118**, 461–470.
Laures, J. S., and Weismer, G. (**1999**). "The effects of a flattened fundamental frequency on intelligibility at the sentence level," J. Speech Lang. Hear. Res. **42**, 1148–1156.
Nelson, P. B., and Jin, S.-H. (**2004**). "Factors affecting speech understanding in gated interference: Cochlear implant users and normal-hearing listeners," J. Acoust. Soc. Am. **115**, 2286–2294.
Nilsson, M., Soli, S. I., and Sullivan, J. A. (**1994**). "Development of the hearing in noise test for the measurement of speech reception thresholds in quiet and in noise," J. Acoust. Soc. Am. **95**, 1985–1099.
Miller, G. A., and Licklider, J. C. (**1950**). "The intelligibility of interrupted speech," J. Acoust. Soc. Am. **22**, 167–173.
Peng, S. C., Lu, N., and Chatterjee, M. (**2009**). "Effects of cooperating and conflicting cues on speech intonation recognition by cochlear implant users and normal hearing listeners," Audiol. Neuro-Otol. **14**, 327–337.
Powers, G. L., and Wilcox, J. C. (**1977**). "Intelligibility of temporally interrupted speech with and without intervening noise," J. Acoust. Soc. Am. **61**, 195–199.
Shannon, R. V., Zeng, F. G., Kamath, V., Wygonski, J., and Ekelid, M. (**1995**). "Speech recognition with primarily temporal cues," Science **270**, 303–304.
Warren, R. M., and Obusek, C. J. (**1971**). "Speech perception and phonemic restorations," Percept. Psychophys. **9**, 358–362.