Recognition of temporally interrupted and spectrally degraded sentences with additional unprocessed low-frequency speech

Deniz Başkent[a,b] and **Monita Chatterjee**[c]
Deniz Başkent: d.baskent@med.umcg.nl; Monita Chatterjee: mchatterjee@hesp.umd.edu

[a] Department of Otorhinolaryngology/Head and Neck Surgery, University Medical Center Groningen, The Netherlands [b] School of Behavioral and Cognitive Neuroscience, University of Groningen, The Netherlands [c] Cochlear Implants and Psychophysics Laboratory, Department of Hearing and Speech Sciences, The University of Maryland, College Park, MD, USA

Abstract

Recognition of periodically interrupted sentences (with an interruption rate of 1.5 Hz, 50% duty cycle) was investigated under conditions of spectral degradation, implemented with a noiseband vocoder, with and without additional unprocessed low-pass filtered speech (cutoff frequency 500 Hz). Intelligibility of interrupted speech decreased with increasing spectral degradation. For all spectral-degradation conditions, however, adding the unprocessed low-pass filtered speech enhanced the intelligibility. The improvement at 4 and 8 channels was higher than the improvement at 16 and 32 channels: 19% and 8%, on average, respectively. The Articulation Index predicted an improvement of 0.09, in a scale from 0 to 1. Thus, the improvement at poorest spectral-degradation conditions was larger than what would be expected from additional speech information. Therefore, the results implied that the fine temporal cues from the unprocessed low-frequency speech, such as the additional voice pitch cues, helped perceptual integration of temporally interrupted and spectrally degraded speech, especially when the spectral degradations were severe. Considering the vocoder processing as a cochlear-implant simulation, where implant users' performance is closest to 4 and 8-channel vocoder performance, the results support additional benefit of low-frequency acoustic input in combined electric-acoustic stimulation for perception of temporally degraded speech.

Keywords

Perceptual integration/restoration; top-down compensation; temporally interrupted speech; spectral degradation; cochlear implant; combined electric-acoustic stimulation

1. Introduction

Recognition of periodically interrupted speech by normal-hearing listeners can be remarkably good, even when a substantial amount of speech information is removed (Miller and Licklider, 1950; Schnotz et al, 2009). Such high level performance with speech that is temporally degraded and reduced in redundancy is attributed to the top-down compensation and effective use of expectations, context, and linguistic rules (Bashford et al., 1992). The

Corresponding author: Deniz Başkent, Dept. Otorhinolaryngology, University Medical Center Groningen, PO Box 30.001 9700 RB Groningen, The Netherlands, d.baskent@med.umcg.nl, Phone: +31-50-3612554.

minimal disruption of the temporal gaps to speech intelligibility, therefore, implies a successful perceptual grouping of speech segments across the temporal gaps.

Studies with hearing-impaired listeners and users of cochlear implants have shown a deficiency in recognition of temporally interrupted speech (Gordon-Salant and Fitzgibbons, 1993; Nelson and Jin, 2004; Başkent, 2010; Başkent et al., 2010; Chatterjee et al., 2010; Jin and Nelson, 2010). This deficiency implies that degradations in bottom-up speech signal, either due to hearing impairment or front-end processing of hearing devices, may make top-down compensation more difficult. Additionally, top-down processing capabilities could be decreased due to advanced age. Consequently, these listeners may not be able to temporally integrate the speech segments and restore speech into a speech stream that can be decoded properly at higher levels of the auditory system, or the higher level processing in the system itself may be compromised.

One of the important cues for perceptual grouping is the fundamental frequency (F0, or voicing pitch; Cusack and Carlyon, 2004; Chang et al., 2006). In normal hearing, the most salient pitch cues are perceived through a place coding mechanism that uses the resolved low-numbered harmonics, and less salient cues are perceived through a temporal coding mechanism that uses the temporal fine structure. In cochlear-implant signal processing, envelopes extracted in a small number of spectral bands modulate electrical pulses delivered at a fixed stimulation rate. The processed signal, therefore, is reduced in spectral resolution and most temporal fine structure carrier information is removed. As a result, pitch cues from resolved low-numbered harmonics and temporal fine structure are minimal (Turner et al., 2004; Qin and Oxenham, 2006). In cochlear-implant users with residual hearing, however, even if the residual hearing is only confined to low frequencies, the low-numbered harmonics in the acoustic signal may be sufficiently resolved, providing salient voice pitch cues. In fact, substantial improvement has been observed in the simulations of combined electric-acoustic hearing and the actual implant listeners with residual hearing with tasks that require strong pitch cues, such as perception of music, concurrent vowels, and speech in complex background noise (Turner et al., 2004; Kong et al., 2005; Chang et al., 2006; Qin and Oxenham, 2006; Arehart et al., 2010; Zhang et al., 2010),

In the present study, we hypothesized that the salient voice pitch cues from unprocessed low-frequency speech would help listeners to form a stronger sense of continuity across temporally interrupted speech segments, and this would lead to better grouping of the segments into one source, enhancing the interpolation (or restoring) of the missing speech information, under conditions of spectral degradation/simulation of cochlear-implant processing.

2. Materials and methods

2.1 Participants

Young normal-hearing listeners were recruited for the study, to eliminate potential effects of cognitive changes that may occur due to aging. Twelve normal-hearing listeners, ages between 17 and 33 years (average of 23 years), participated in the study. Their pure-tone thresholds were less than 20 dB HL at the audiometric frequencies of 250–6000 Hz in both ears. All participants were native speakers of Dutch. The listeners were recruited through the Psychology Department of the University of Groningen and received credit for participation. The study was approved by the Ethical Committee of the Psychology Department of the University of Groningen, and written informed consents were collected from the listeners prior to the participation.

2.2 Stimuli

Speech stimuli were Dutch sentences, recorded at Vrij University (VU), Amsterdam (Versfeld et al., 2000). The VU database has 78 balanced lists of 13 grammatically and syntactically correct, meaningful and contextual sentences. The sentences represent conversational speech and are semantically neutral. There are 4 to 9 words in each sentence, and 74 to 88 words in each list. Half of the sentences were spoken by a male talker and the other half by a female talker. The male talker's voice pitch range was 80 to 160 Hz and the female talker's was 120 to 290 Hz. The speakers spoke with a similar rate to each other, with around an average duration of 325 ms per word, as estimated from the initial ten sentence recordings for each speaker. Despite the similarity in speech materials and speaking rates between the two talkers, Versfeld et al. (2000) have reported that the overall intelligibility varied. Therefore, the performances with different speakers will not be compared to each other in the present study. Instead, the scores will be pooled with the purpose of enhanced generalization of the results.

2.3 Signal processing

The speech stimuli were first interrupted by modulating with a periodic square-wave envelope, and then spectrally reduced with noiseband vocoder processing, without or with unprocessed low-frequency speech (details given in the subsequent sections). This order in signal processing was preferred, as the secondary aim of the present study was to use the vocoder processing without and with unprocessed low-frequency speech as simulations of cochlear-implant processing and combined electric-acoustic hearing, respectively. Hence, we simulated speech temporally interrupted before it arrived at the microphones of the auditory devices.

2.3.1 Periodic temporal interruptions—Periodic temporal interruptions were applied to the sentences by modulating with a square-wave envelope. The interruption rate was 1.5 Hz, which produced on and off speech phases of 333 ms duration each (50% duty cycle). Based on a pilot and previous studies (Miller and Licklider, 1950; Powers and Speaks, 1973; Nelson and Jin, 2004; Gilbert et al, 2007), this slow interruption rate was selected to yield maximal changes in speech perception with temporal interruptions and to leave maximum space for potential improvement in performance with the addition of the low-frequency speech cues, while avoiding floor and ceiling effects. A 5-ms raised cosine at the onset and offset of the speech segments prevented distortions due to spectral splatter.

2.3.2 Spectral degradation—Spectral degradation was implemented with noiseband vocoder (NBV) processing, which has been widely used as a simulation of cochlear-implant processing (Dudley, 1939; Shannon et al, 2005). In the NBV, 4, 8, 16, and 32 channels were implemented. The entire spectral range of the vocoder processing was limited to 150 to 7000 Hz, a range similar to that of real cochlear implants. The lower and higher limits of the spectral range were first converted into corresponding cochlear distances, using the Greenwood mapping function (Greenwood, 1990). Then the individual spectral ranges (in mm) were calculated by dividing the entire cochlear length into equal distances and the ranges in mm were converted into corresponding frequencies in Hz with the mapping function. The resulting frequency cut-offs of individual frequency bands are shown in Table I. These cutoff frequencies applied to both analysis and synthesis filters. The envelopes were extracted from analysis filters by half-wave rectification, followed by a low-pass Butterworth filter (18 dB/oct) with a cutoff frequency of 160 Hz. Filtering white noise with the synthesis filters produced the carrier noise bands. The noise carriers in each channel were modulated with the corresponding extracted envelope. The processed speech was the sum of the modulated noise bands from all vocoder channels.

2.3.3 Unprocessed low-pass filtered speech—The main question of the present study was how adding unprocessed low-frequency speech (LP500) would change the performance with temporally interrupted and spectrally degraded sentences. NBV combined with LP500 can be considered as a simulation of combined electric-acoustic stimulation (Turner et al., 2004). LP500 was produced by low-pass filtering unprocessed speech with a Butterworth filter (18 dB/oct) with a cutoff frequency of 500 Hz. The cutoff value was mainly based on previous studies (Turner et al., 2004; Qin and Oxenham, 2006; Brown and Bacon, 2009a), but it also closely resembled the range of residual hearing observed with real implant users (Reiss et al., 2007; Büchner et al, 2009). The order of the filter was relatively low, imitating listeners with very good residual hearing (for example, S7 in Zhang et al, 2010). The low-pass filtered speech was incorporated into processed stimuli by replacing low-frequency vocoder channels, as shown in Table I. Figure 1 shows the spectra of temporally interrupted and spectrally degraded wideband noise signals, taken from the database, each matching the long-term speech spectrum for male and female talkers (left and right panels, respectively). The top and bottom panels show the spectra with worst and best spectral degradation conditions, respectively. The spectra were calculated from the first 5 s of the noise samples, using Audacity spectrum analyzer with a Hanning window of size 512. For reference, spectra of unprocessed and interrupted noise are also shown.

Using the spectra shown in the figure, the Articulation Index (AI; ANSI S3.5-1969R1986) was calculated to predict the improvement in intelligibility due to the addition of LP500. AI is a measure of audibility, in a scale from 0.0 to 1.0, that is proportionate to the audible average speech energy above the hearing thresholds and/or not masked by background noise. It is based on earlier work by Fletcher and colleagues who performed low-pass and high-pass filtering on speech until each portion produced equal intelligibility (Fletcher and Steinberg, 1929; French and Steinberg, 1947). Studebaker and Sherbecoe (1993) further refined the method, developing a spectral weighting function to determine the contribution of different frequencies to the overall intelligibility. The AI and its variations have since been used to predict speech intelligibility with hearing impairment and hearing aids (Pavlovic et al., 1986; Pavlovic, 1991), cochlear implants (Shannon et al., 2002; Faulkner et al., 2003), and other distortions, such as reverberation (Steeneken and Houtgast, 1980).

In the AI calculations of the present study, LP500 spectrum was the average speech spectrum. Because portions of it were masked by the accompanying NBV processing in the higher frequencies, the NBV-processed part was the average noise spectrum. The audiometric thresholds were that of normal hearing, i.e., 0 dB HL at all frequencies. With these calculations, AI predicted an improvement of 0.09 (or 9% in percent correct) due to the additional audible speech information provided by LP500.

The effects of temporal interruption, spectral degradation, and adding low-frequency speech are further shown in Figure 2 with one sentence spoken by the male talker. The figure shows spectrograms of the original waveform (in the upper panel) and the processed stimuli (in the lower panels). The middle panels show the spectrograms with 4-channel conditions, similar to the processing shown in the upper left panel of Figure 1. The lower panels show the spectrograms with 32-channel conditions, similar to the processing shown in the lower left panel of Figure 1. The left and right panels show the effects of temporal interruption and spectral degradation without and with additional LP500, respectively. In each panel, the solid line shows the F0 pitch contours. The spectrograms and the pitch contours were calculated using PRAAT software (Boersma, 2001) with standard settings, except for the spectral bandwidth of 0 to 7000 Hz, a window length of 50 ms, the choice of Hamming window instead of Gaussian, and a dynamic range of 50 dB. To make the spectrogram details more visible, especially at the lower frequencies where the harmonics can be observed, the high-frequency pre-emphasis filter was turned off. The standard settings for

the spectrogram were a time step of 2 ms and a frequency step of 20 Hz. The standard pitch calculation was based on the autocorrelation method, with a frequency range of 75 to 600 Hz.

In the original spectrogram in the upper panel, spectral details, such as the harmonics of the voice pitch, the pitch contours, and the formants, are visible. In the spectrograms in the left panels, which show the effect of temporal interruption and spectral degradation due to NBV processing, the harmonics and the pitch contours are mostly not visible. The spectrogram in the lower left panel shows that, at higher spectral resolution of 32 channels, some formants are recovered, but only a very small ratio of pitch contours is accurately estimated. After LP500 is added, as shown in the right panels, the lower voice harmonics are discernible and the pitch contours are accurately estimated, at both spectral-degradation conditions of 4 and 32 channels. This figure again emphasizes that adding LP500 brings back the voice pitch cues that were lost as a result of NBV processing.

2.4 Experimental set-up

The listeners were seated in an anechoic chamber facing a touchscreen monitor. The routing of the stimuli was through the S/PDIF output of the external Echo AudioFire 4 soundcard that was connected to the Lavry DA10 D/A converter. The stimuli were then presented diotically with Sennheiser HD 600 headphones to the listener.

2.5 Procedure

A Matlab program was used for signal processing and playback of the stimuli. With a Matlab Graphical User Interface (GUI), the listener could control the pace of the experiment, by pressing "Next" button on a touchscreen monitor, when they were ready for the next sentence. The presentation level of the stimuli was 60 dBA (as measured at the headphone output that was placed on Bernie, the KEMAR head). A short tone was played before each sentence to alert the listener to the stimulus. The listeners heard one sentence at a time and immediately repeated what they heard. They were instructed to say any words they have heard even if they had to guess or even if they could understand only one word. The responses from the listeners were recorded during the test with a digital voice recorder and were later transcribed by a student assistant, who did not know the experimental conditions. All words of the sentences were included in the scoring. The percent correct scores were calculated from the ratio of the number of the correctly identified words to the total number of the words within each list.

Listeners of the present study were naïve. Before actual data collection, a short practice was given with one sentence list with uninterrupted 4-channel vocoder processing. An intelligibility score of 30% was accepted as threshold, based on pilot data with more experienced listeners. The practice run was repeated with each listener until the threshold was reached. During data collection, half of the listeners were first tested with NBV conditions without LP500, and the other half first with NBV conditions with LP500. In each block of data collection, the order of the conditions was random. For each condition, one list of thirteen sentences was used and no list was repeated for any of the listeners. No feedback was provided during training or data collection.

3. Results

Figure 3 shows the mean percent correct scores, combined for male and female talkers, as a function of spectral degradation. The top panel shows the intelligibility with temporally interrupted and spectrally degraded speech, with or without the LP500 (open and filled symbols, respectively). The gray score shows the effect of interruptions alone, with no

spectral degradation. The lower panel shows the difference in performance with and without LP500, hence these results show the improvement in performance as a result of adding LP500 to NBV. The dashed line shows the improvement predicted by the AI calculations.

Similar to previous studies, the performance with interruptions only, shown with the gray symbols, was high, 68% on average. When spectral degradation was added with NBV, as shown by the black symbols, there was a significant reduction in performance. Even at the highest number of the vocoder channels, 32, the performance at 56% was significantly lower than the baseline score of 68% (paired t-test, p<0.01).

In general, reducing the number of the vocoder channels reduced the performance, both with NBV and NBV with LP500, implying a negative effect of spectral degradation on intelligibility of interrupted speech. However, for all numbers of vocoder channels, the performance with LP500 was higher than without. A repeated-measures two-factor ANOVA showed significant main effects of the number of vocoder channels (F=246.54, p<0.001) and the presence/absence of LP500 (F=89.48, p<0.001). A posthoc Tukey test indicated that the improvement due to LP500 was significant at all vocoder-channel conditions. There was also a significant interaction between these factors (F=7.08, p<0.001), implying that the amount of the improvement due to LP500 differed at different numbers of channels. In fact, the lower panel of Figure 3 shows the varying effect. An improvement of 18–19% was observed at smaller numbers of channels (4, 8) that was larger than both the improvement of 7–9% observed at higher numbers of channels (16, 32) and the predicted improvement of 9% by the AI.

4. Discussion

The baseline percent correct scores with interrupted speech showed that normal-hearing listeners can achieve high levels of intelligibility even when a significant amount of speech information is missing, similar to previous findings (Miller and Licklider, 1950; Nelson and Jin, 2004; Gilbert et al, 2007; Chatterjee et al., 2010). However, once speech was also spectrally degraded with noiseband vocoder processing, performance dropped significantly. At 4 channels of processing, for example, the performance was very low, at around 5% on average. Nelson and Jin (2004) reported similar low speech perception performance with slow interruption rates of 1 to 2 Hz. Even at the mildest spectral degradation condition of 32 channels, the performance was significantly lower than the baseline by 12 percentage points. With a similar finding, Chatterjee et al. concluded that even a minimal spectral degradation in the bottom-up signal made a top-down restoration more difficult. Such spectral degradation occurs inherently in cochlear implants, for reasons such as limited number of electrodes, irregular patterns in surviving neurons and the proximity of the electrodes to them, and possible channel interactions due to current spread and cross-talk between multiple electrodes (Friesen et al., 2001; Fu and Nogaki, 2005), and this is possibly one of the factors contributing to reduced performance by implant users with interrupted speech (Nelson and Jin, 2004; Chatterjee et al., 2010).

An important cue in perceptual grouping is pitch, which is not strongly delivered in cochlear implants due to reduced spectral resolution and temporal fine structure. In the present study, we hypothesized that adding unprocessed low-pass filtered speech to the interrupted and spectrally degraded speech would improve the performance. The hypothesis was based on the idea that the low-frequency speech would provide salient voice pitch cues that are consistent across speech segments, and these would help listeners to form a stronger sense of continuity and better grouping of these segments into a single speech signal, which in return would help restore the speech into a stream that can be decoded more easily and effectively with top-down resources from the higher level auditory system. Previous studies have shown

that such narrowband low-frequency speech by itself provides minimal intelligibility, for example in background noise (Kong et al., 2005). Therefore, the main advantage it provides is considered to be stronger low-frequency voicing cues (Qin and Oxenham, 2006; Brown and Bacon, 2009a) that possibly help segregating speech segments from background noise and grouping the appropriate temporal cues to identify the speech signal (Chang et al., 2006). In the present study, using the AI model, we predicted that the additional intelligibility provided from the audible portions of LP500 would be 9%. The observed improvement was the same as the predicted improvement at large number of NBV channels (16 and 32). At small numbers of NBV channels (4 and 8), however, an improvement larger than predicted was observed, at 18–19%. Hence, at worst spectral degradation conditions, the contribution from LP500 was more than would be expected from added speech information alone. We have attributed this benefit to the additional effects from other speech cues especially important for perceptual grouping, such as the voice pitch cues. Thus, the present data supported the hypothesis that additional voice pitch cues from the unprocessed low-frequency speech provide a strong bottom-up cue and help perceptual integration and top-down compensation/restoration of interrupted and spectrally degraded speech, especially at poor spectral resolution conditions.

In a similar study, based on similar reasoning explained above, Chatterjee et al. (2010) hypothesized that changing F0 contours would affect perception of interrupted and spectrally degraded speech. However, the hypothesis was not supported by the data; there was no difference in performance when natural or flattened F0 contours were used. A comparison between the present study and the study by Chatterjee et al. implies that while the presence or absence of the low-frequency F0 cues is important for perceptual grouping of interrupted and spectrally degraded speech, the shape of the existing F0 contours does not seem to be. Note that there were a number of differences between the two studies. Chatterjee et al. used a vocoder simulation where the envelope cutoff frequency was high, 400 Hz, which preserved more F0 cues in the NBV. In the present study, the cutoff frequency of the envelope filter was 160 Hz, a lower value than the Chatterjee study. In addition, the interruption rate used by Chatterjee et al. was faster at 5 Hz, compared to the interruption rate of the present study at 1.5 Hz. Different interruption rates leave out different speech structures, possibly producing different effects on intelligibility and on top-down restoration of interrupted speech (Nelson and Jin, 2004; Gilbert et al, 2007; Başkent et al, 2010). These differences may have contributed to generally lower scores with interrupted and spectrally degraded speech of the present study, leaving more room for improvement. Regardless, all results combined, the presence or absence of the F0 cues seems to be a more important factor than varying the inherent F0 contours for interrupted speech perception with spectral degradations of implant processing. This point was further supported by studies by Brown and Bacon (2009a, 2009b). They showed that adding a simple tone at F0 that was amplitude-modulated with speech envelope was sufficient to produce an improvement in speech perception with implants and implant simulations.

The results of the present study have practical implications for cochlear implants. The noiseband vocoder processing can be considered as a simulation of cochlear implant processing, while the additional unprocessed low-pass filtered speech can be considered as a simulation of the acoustic input in combined electric-acoustic stimulation. The results of the present study, therefore, imply an additional advantage of combined electric-acoustic stimulation, namely stronger grouping with electric-acoustic stimulation than the implant alone.

There are a number of factors that may potentially limit the applicability of these results to actual users of implants. The first such factor is the level of the hearing thresholds for residual hearing, which may not be as good as simulated in the present study. Hence, the

listeners may have less access to the low-frequency speech cues. On a similar note, the AI prediction of the present study may differ once applied to the implant population. A second factor is a potential degradation in the top-down mechanisms themselves, for example, due to aging. In the present study, the listeners are young, potentially with best cognitive capacity, and therefore all results were attributed to the changes in signal processing. In real implant listeners, the cognitive abilities may vary and therefore produce a different outcome. On the other hand, the results showed that the effect was strongest at smaller numbers of vocoder channels, 4 and 8. This is a positive finding for implant users, considering that Friesen et al. (2001) have shown that spectral resolution by implant users was functionally closest to 4 to 8-channel vocoder simulations for perception of speech in quiet and in noise, and Nelson and Jin (2004, figure 6) have shown that implant users' performance was close to 4-channel vocoder simulation for perception of temporally interrupted speech. As a result, the present findings imply that the additional acoustic input has great potential to help implant users with residual hearing in complex listening environments.

Acknowledgments

References

American National Standards Institute. American national standard methods for the calculation of the articulation index. ANSI S3.5-1969 R 1986.

Arehart KH, Souza PE, Muralimanohar RK, Miller CW. Effects of age on concurrent vowel perception in acoustic and simulated electro-acoustic hearing. J Sp Lang Hear Res. 201010.1044/1092-4388(2010/09-0145)

Bashford JA Jr, Riener KR, Warren RM. Increasing the intelligibility of speech through multiple phonemic restorations. Percept Psychophys 1992;51:211–217. [PubMed: 1561046]

Başkent D. Phonemic restoration in sensorineural hearing loss does not depend on baseline speech perception scores. J Acoust Soc Am 2010;128(3)

Başkent D, Eiler CL, Edwards B. Phonemic restoration by hearing-impaired listeners with mild to moderate sensorineural hearing loss. Hear Res 2010;260:54–62. [PubMed: 19922784]

Boersma P. Praat, a system for doing phonetics by computer. Glot International 2001;5:341–345.

Brown CA, Bacon SP. Low-frequency speech cues and simulated electric-acoustic hearing. J Acoust Soc Am 2009a;125:1658–1665. [PubMed: 19275323]

Brown CA, Bacon SP. Achieving electric-acoustic benefit with a modulated tone. Ear Hear 2009b; 30:489–493. [PubMed: 19546806]

Büchner A, Schüssler M, Battmer RD, Stöver T, Lesinski-Schiedat A, Lenarz T. Impact of low-frequency hearing. Audiol Neurotol 2009;14(Supp 1):8–13.

Chatterjee M, Peredo F, Nelson D, Başkent D. Recognition of interrupted sentences under conditions of spectral degradation. J Acoust Soc Am 2010;127:EL37–EL41. [PubMed: 20136176]

Cusack, R.; Carlyon, RP. Auditory scene analysis inside and outside the laboratory. In: Neuhoff, Jon, editor. Ecological Psychoacoustics. 2004.

Dudley H. Remaking speech. J Acoust Soc Am 1939;11:169–177.

Faulkner A, Rosen S, Stanton D. Simulations of tonotopically-mapped speech processors for cochlear implant electrodes varying in insertion depth. J Acoust Soc Am 2003;113:1073–1080. [PubMed: 12597200]

Fletcher H, Steinberg JC. Articulation testing methods. Bell Systems Tech J 1929;8:806–854.

French NR, Steinberg JC. Factors governing the intelligibility of speech sounds. J Acoust Soc Am 1947;19:90–119.

Friesen LM, Shannon RV, Başkent D, Wang X. Speech recognition in noise as a function of the number of spectral channels: Comparison of acoustic hearing and cochlear implants. J Acoust Soc Am 2001;110:1150–1163. [PubMed: 11519582]

Fu Q-J, Nogaki G. Noise susceptibility of cochlear implant users: The role of spectral resolution and smearing. J Assoc Res Otolaryn 2005;6:19–27.

Gilbert G, Bergeras I, Voillery D, Lorenzi C. Effects of periodic interruptions on the intelligibility of speech based on temporal fine-structure or envelope cues (L). J Acoust Soc Am 2007;122:1336–1339. [PubMed: 17927396]

Gordon-Salant S, Fitzgibbons PJ. Temporal factors and speech recognition performance in young and elderly listeners. J Sp Hear Res 1993;36:1276–1285.

Greenwood DD. A cochlear frequency-position function for several species—29 years later. J Acoust Soc Am 1990;87:2592–2605. [PubMed: 2373794]

Jin S-H, Nelson PB. Interrupted speech perception: The effects of hearing sensitivity and frequency resolution. J Acoust Soc Am 2010;128:881–889. [PubMed: 20707457]

Kong Y-Y, Stickney GS, Zeng F-G. Speech and melody recognition in binaurally combined acoustic and electric hearing. J Acoust Soc Am 2005;117:1351–1361. [PubMed: 15807023]

Miller GA, Licklider JC. The intelligibility of interrupted speech. J Acoust Soc Am 1950;22:167–173.

Nelson PB, Jin S-H. Factors affecting speech understanding in gated interference: Cochlear implant users and normal-hearing listeners. J Acoust Soc Am 2004;115:2286–2294. [PubMed: 15139640]

Pavlovic CV, Studebaker GA, Sherbecoe RL. An articulation index based procedure for predicting the speech recognition performance of hearing-impaired individuals. J Acoust Soc Am 1986;80:50–57. [PubMed: 3745665]

Pavlovic CV. Speech recognition and five articulation indexes. Hearing Instruments 1991;42:20–23.

Powers GL, Speaks C. Intelligibility of temporally interrupted speech. J Acoust Soc Am 1973;54:661–667. [PubMed: 4754389]

Qin MK, Oxenham AJ. Effects of introducing unprocessed low-frequency information on the reception of envelope-vocoder processed speech. J Acoust Soc Am 2006;119:2417–2426. [PubMed: 16642854]

Reiss LAJ, Turner CW, Erenberg SR, Gantz BJ. Changes in pitch with a cochlear implant over time. J Assoc Res Otolaryn 2007;8:241–257.

Schnotz A, Digeser F, Hoppe U. Speech with gaps: Effects of periodic interruptions on speech intelligibility. Folia Phoniatrica et Logopaedica 2009;61:263–268. [PubMed: 19696487]

Shannon RV, Zeng F-G, Kamath V, Wygonski J, Ekelid M. Speech recognition with primarily temporal cues. Science 1995;270:303–304. [PubMed: 7569981]

Shannon RV, Galvin JJ 3rd, Başkent D. Holes in hearing. J Assoc Res Otolaryngol 2002;3:185–199. [PubMed: 12162368]

Steeneken HJM, Houtgast T. A physical method for measuring speech transmission quality. Journal of the Acoustical Society of America 1980;67:318–326. [PubMed: 7354199]

Studebaker, GA.; Sherbecoe, RL. Frequency – importance functions for speech recognition. In: Studebaker, GA.; Hochberg, I., editors. Acoustical Factors Affecting Hearing and Performance. Allyn and Bacon; Boston: 1993.

Turner CW, Gantz BJ, Vidal C, Behrens A, Henry BA. Speech recognition in noise for cochlear implant listeners: Benefits of residual acoustic hearing. J Acoust Soc Am 2004;115:1729–1735. [PubMed: 15101651]

Versfeld NJ, Daalder L, Festen JM, Hourgast T. Method for the selection of sentence materials for efficient measurement of the speech reception threshold. J Acoust Soc Am 2000;107:1671–1684. [PubMed: 10738820]

Zhang T, Dorman MF, Spahr AJ. Information from the voice fundamental frequency (F0) region accounts for the majority of the benefit when acoustic stimulation is added to electric stimulation. Ear Hear 2010;31:63–69. [PubMed: 20050394]
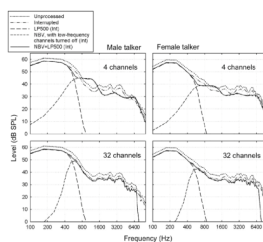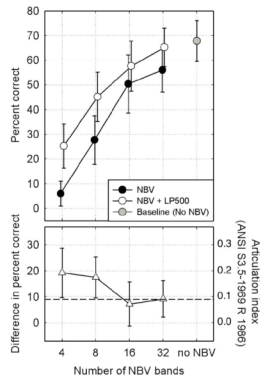
Figure 1.
Spectra of noise matching the long-term speech spectrum of male and female talkers, shown after periodic temporal interruptions and highest (4 channels) and lowest (32 channels) spectral degradations were applied. The spectra of unprocessed and only-interrupted noise are added as a reference.
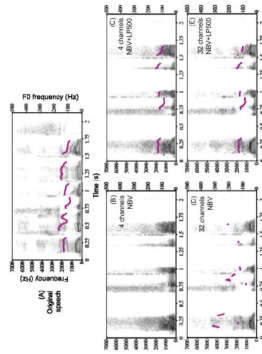
Figure 2.
Spectrograms of one sentence spoken by the male talker and processed with different experimental conditions. (A) shows the original sentence. 500 Hz (the cut-off frequency for LP500) is marked with the straight dashed line as a reference. The solid lines superimposed with the spectrogram are the pitch contours (time-varying F0 frequency) extracted by PRAAT software. The middle and lower panels show the sentence with temporal interruption and spectral degradation with 4 and 32 channels, respectively. The left and right panels show the sentence with NBV and NBV+LP500 processing, respectively.

Figure 3.
The top panel shows the mean percent correct scores of interrupted and spectrally degraded sentences, with and without the LP500, as a function of the number of the vocoder channels. The gray symbol shows the baseline score with full-bandwidth interrupted speech with no spectral degradation. The lower panel shows the improvement in performance due to the addition of LP500 for each vocoder-channel condition. The dashed line shows the improvement predicted by the Articulation Index. The error bars denote one standard deviation.

TABLE I

Filter cut-off frequencies for noiseband vocoder (NBV) processing, with and without the unprocessed low-frequency speech (LP500).

4-channel vocoder			8 to 32-channel vocoder				
Channel number	NBV (Low to high, Hz)	NBV with LP500 (Low to high, Hz)	8 ch Channel number	16 ch Channel number	32 ch Channel number	NBV (Low to high, Hz)	NBV with LP500 (Low to high, Hz)
1	150–510	Unprocessed (Low-pass filter cutoff = 500 Hz)	1	1–2	1–4	150–295	Unprocessed (Low-pass filter cutoff = 500 Hz)
			2	3–4	5–8	295–510	510–830
2	510–1308	510–1308	3	5–6	9–12	510–830	830–1308
3	1308–3077	1308–3077	4	7–8	13–16	830–1308	1308–2019
4	3077–7000	3077–7000	5	9–10	17–20	1308–2019	2019–3077
			6	11–12	21–24	2019–3077	3077–4653
			7	13–14	25–28	3077–4653	4653–7000
			8	15–16	29–32	4653–7000	