Effects of low-pass filtering on intelligibility of periodically interrupted speech

Pranesh Bhargava[a)] and Deniz Başkent[a)]
Department of Otorhinolaryngology/Head and Neck Surgery, University Medical Center Groningen, P.O. Box 30.001, 9700 RB Groningen, The Netherlands
p.bhargava@umcg.nl, d.baskent@umcg.nl

Abstract: The combined effect of low-pass filtering (cut-off frequencies between 500 and 3000 Hz) and periodic interruptions (1.5 and 10 Hz) on speech intelligibility was investigated. When combined, intelligibility was lower than each manipulation alone, even in some conditions where there was no effect from a single manipulation (such as the fast interruption rate of 10 Hz). By using young normal-hearing listeners, potential suprathreshold deficits and aging effects that may occur due to hearing impairment were eliminated. Thus, the results imply that reduced audibility of high-frequency speech components may partially explain the reduced intelligibility of interrupted speech in hearing impaired persons.

PACS numbers: 43.66.Mk, 43.71.Es, 43.71.Rt [QJF]
Date Received: September 23, 2011 **Date Accepted:** November 7, 2011

1. Introduction

The human auditory system has a remarkable capacity for understanding speech in adverse conditions. For example, speech interrupted with periodic silent intervals remains highly intelligible, especially at fast interruption rates, despite omitting up to 50% of the signal (Miller and Licklider, 1950; Powers and Speaks, 1973). The high level intelligibility of interrupted speech is partially attributed to the top-down repair mechanisms of the auditory system that use context, expectations, and linguistic rules (Schnotz *et al.*, 2009; Chatterjee *et al.*, 2010).

Recently, Jin and Nelson (2010) tested intelligibility of interrupted speech in normal-hearing (NH) and hearing-impaired (HI) listeners for interruption rates between 1 and 16 Hz. While slower interruption rates ($\leq$4 Hz) adversely affected both groups, performance by HI listeners was lower at all interruption rates. Başkent (2010) further found that, at slow interruption rates (1–2 Hz), the performance of the HI listeners was negatively correlated with the degree of hearing loss (and the age of the listener).

It is not yet clear what factors cause the poor intelligibility of interrupted speech in hearing impairment. While threshold elevation and reduced audibility (Jin and Nelson, 2006) and suprathreshold deficits, such as temporal and spectral degradation (Horwitz *et al.*, 2002; Gnansia *et al.*, 2010), due to hearing loss could affect the bottom-up speech cues, reduced cognitive resources due to aging (Gordon-Salant and Fitzgibbons, 1993) could affect the top-down repair mechanisms.

In this study, as a first step toward investigating the factors causing reduced intelligibility of interrupted speech with hearing impairment, we aimed to explore the effect of audibility alone, without any other potential factors mentioned before. A systematic loss of high-frequency (HF) speech components was induced with young NH listeners by low-pass (LP) filtering speech stimuli with varying cut-off frequencies and filter orders (similar to Horwitz *et al.*, 2002) prior to applying the interruptions. Thus, potential effects of suprathreshold deficits and aging were eliminated, and only the effect of audibility on perception of interrupted speech is studied.

[a)]Also at: School of Behavioural and Cognitive Neuroscience, University of Groningen, The Netherlands.

2. Materials and methods

2.1 Test materials

Meaningful Dutch sentences, digitally recorded at the sampling rate of 44.1 kHz, were taken from the Vrij University corpus (Versfeld *et al.*, 2000). Each sentence was four to nine words long, containing words up to three syllables long. The corpus has two subsets (one spoken by a male speaker and the other by a female speaker) of 39 lists each. Each list has 13 sentences. For the present study, the first 37 lists from male (the first for training, the next 36 for experiment) and the first 36 lists from female talker were used.

2.2 Participants

Eight Dutch native speakers of both genders (ages 19–22 years; average age about 20 years), who were undergraduate students of the Psychology Department at the University of Groningen and reported no hearing problems, participated in the study. Course credit was given for participation. Written information about the study was provided and written informed consent was collected prior to the experiment. The study was approved by the Ethical Committee of the Psychology Department.

2.3 Experimental conditions

The sentences were LP filtered using a Butterworth filter (four cut-off frequencies at 500, 1000, 2000, and 3000 Hz, and three filter orders at 1, 3 and 10, with corresponding filter slopes of 6, 18, and 60 dB/octave, respectively). These conditions were selected to simulate a wide range of hearing loss configurations, as well as to retain or reduce specific speech cues, such as voice pitch, vowel formants, and consonants.

LP filtered sentences were either left uninterrupted, or were interrupted by modulating with a periodic square wave of 1.5 or 10 Hz. To prevent LP filtering effects on the square wave the interruption was applied after the filtering. The interruption rates were selected based on previous studies, one a slow phonemic interruption rate producing a significant reduction in intelligibility (Başkent, 2006; Başkent *et al.*, 2010), and one faster rate producing minimal reduction in intelligibility (Jin and Nelson, 2010). The duty cycle was 50%, and a raised cosine ramp of 5 ms was applied to the onsets and offsets of the square wave to prevent spectral splatter.

The experiment consisted of 8 blocks (2 speakers $\times$ 4 cut-off frequencies), with 9 trials each (3 filter orders $\times$ 3 interruption conditions; see Table 1). Thus, data collection comprised of a total of 72 trials with 936 sentences and lasted around 2 h. For the experiment, the lists were presented always in the same order to all listeners, but the order of blocks and the order of trials in the blocks were randomized. Thus, each participant heard the same order of the sentences processed with different conditions.

2.4 Experimental setup and procedure

The stimuli were processed and presented using MATLAB on a Macintosh computer. The processed digital signal was sent through the S/PDIF output of AudioFire 4, the external soundcard of Echo Digital Audio Corporation (California, USA). After conversion to an analogue signal via DA10 digital-to-analog converter of Lavry Engineering Inc. (Washington, USA), it was played back diotically with HD600 headphones of Sennheiser Electronic Corporation (Connecticut, USA) at an rms level of 60 dB SPL. A short beep preceded the stimulus to alert the listener.

Each participant, seated inside an anechoic chamber, listened to the stimulus, and repeated verbally what s/he heard. Listeners were encouraged to guess as much as they feasibly could. When done, the participant requested the next sentence by giving a cue to the experimenter, who then used the graphical user interface displayed on a touch screen monitor to play the next stimulus. The spoken responses of the participants were recorded on DR-100 digital voice recorder by Tascam (California, USA), for offline scoring. A native Dutch speaking student assistant listened to the recordings, and calculated the percent correct scores by the ratio of correctly identified words to

Table 1. Conditions in one block shown with corresponding nine trials. The blocks were repeated for two speakers (male and female), and for four cut-off frequencies (500, 1000, 2000, and 3000 Hz) of the LP filter.

Trial	Rate of interruption	Filter order
1	No interruption	1
2	1.5 Hz	1
3	10 Hz	1
4	No interruption	3
5	1.5 Hz	3
6	10 Hz	3
7	No interruption	10
8	1.5 Hz	10
9	10 Hz	10

the total number of words presented to the listener. Wrong identification of the words was not penalized.

For familiarization with the procedure a short training (with different parameters than actual testing) was provided before the actual experiment. One list of sentences, which was the same for each participant, was used for training. No feedback was provided during the training or data collection.

3. Results

Since the performance of the participants was at ceiling for several conditions, the data are reported in the form of rationalized arcsine transformation unit (RAU) scores (Studebaker, 1985) instead of percent correct scores. Figure 1 shows the mean RAU scores combined for the stimuli spoken by the male and female speakers, as a function of the filter order. The panels show the results for different cut-off frequencies. In each panel, different lines show the RAU transformed intelligibility scores with different interruption conditions.

Figure 1 shows that, without the interruptions (open circles), there was no effect of LP filtering on the speech intelligibility, except for the most aggressive filtering condition, i.e., the lowest cut-off frequency of 500 Hz (top left panel) and the highest filter order of 10. The effect of interruptions alone could best be observed with the least filtering condition, i.e., the highest cut-off frequency of 3000 Hz (right lower panel) and the lowest filter order of 1: While the slow interruption rate of 1.5 Hz reduced speech intelligibility (open square), there was no effect of the fast interruption rate of 10 Hz (open triangle). When combined, however, an interactive and detrimental effect of LP filtering and interruption was observed on speech intelligibility, especially at the low cut-off LP filtering conditions of 500 and 1000 Hz (upper panels). As the severity of filtering, i.e., the filter order, increased, intelligibility of speech decreased. Most strikingly, the faster interruption rate of 10 Hz, which had no effect on intelligibility at the low filter order of 1, reduced intelligibility dramatically at the higher filter order of 10.

A repeated measures three-way analysis of variance (ANOVA) on the RAU scores, with the variables of cut-off frequency, filter order, and interruption rate showed a significant main effect of all three independent variables, with significant interactions between them (Table 2).

4. Discussion

In this study, the effect of LP filtering (to simulate missing HF speech information that may occur in hearing loss) on intelligibility of temporally interrupted speech was

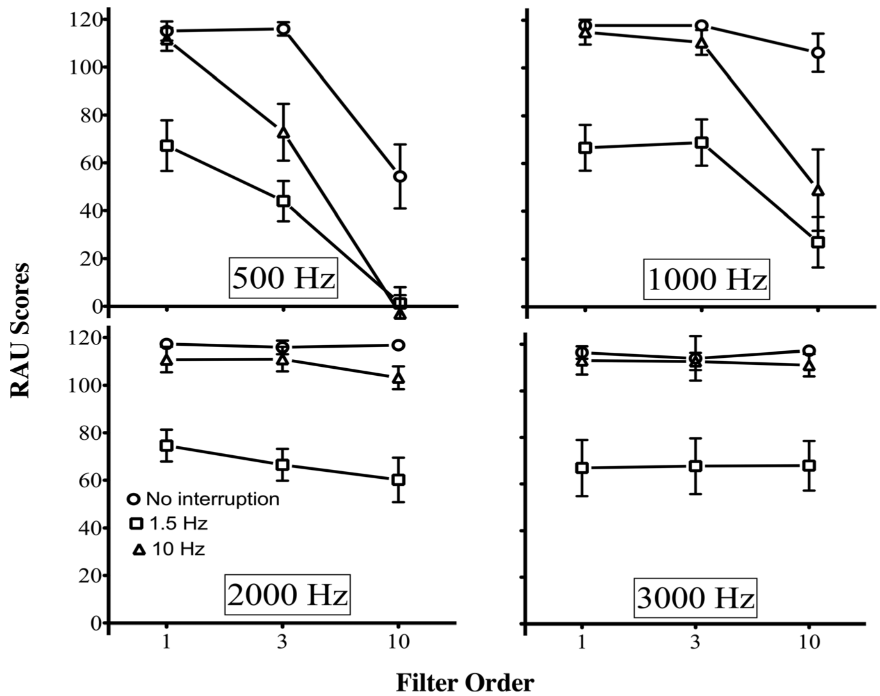

Fig. 1. The mean RAU scores shown as a function of the filter order. The panels show the results for different cut-off frequencies, and within each panel, the results are shown separately for different interruption conditions. The error bars denote the standard deviations.

investigated. The cut-off frequencies of LP filtering ranged from 500 to 3000 Hz, while the interruption rates were slow and fast at 1.5 and 10 Hz, respectively.

LP filtering alone had little effect on the intelligibility of uninterrupted speech, despite removing speech information important for phoneme identification. The most important frequencies for understanding speech are 1–4 kHz, as this range contains important formant information for vowel identification, as well as cues for consonant identification (Fletcher, 1953; Owens *et al.*, 1972; Walden *et al.*, 1981). However, with uninterrupted meaningful sentences, the syntactic and semantic cues available from the linguistic context of the uninterrupted words likely compensate for the impoverished HF speech cues. As a result, LP filtered uninterrupted speech may remain highly intelligible.

The effect of interruption alone was best visible at the least aggressive LP conditions (lowest filter order and highest cut-off frequency). Similar to previous studies

Table 2. Results of the three-way repeated measures ANOVA on the RAU scores. For all main effects and interactions, the p value was highly significant at $p < 0.001$.

Main effects and interactions	F value
1. Filter order	$F(2,14) = 174.34$
2. Interruption rate	$F(2,14) = 37.44$
3. Cut-off frequency	$F(3,21) = 28.94$
4. Filter order × Interruption rate	$F(4,28) = 217.05$
5. Filter order × Cut-off frequency	$F(6,42) = 120.87$
6. Interruption rate × Cut-off frequency	$F(6,42) = 240.17$
7. Order × Cut-off × Interruption	$F(12,84) = 207.60$

(Miller and Licklider, 1950; Jin and Nelson, 2010; Başkent, 2010; Shafiro *et al.*, 2011), the slow interruption rate of 1.5 Hz was detrimental for intelligibility of interrupted speech, whereas the fast interruption rate of 10 Hz was not. The difference in these effects is attributed to the obliteration of entire words from the speech stream at the slower rate, in contrast to the increased looks per word at the faster rate.

The main interest of the present study was the combined effect of LP filtering and temporal interruption. Despite producing minimal effects when applied individually, intelligibility was reduced drastically at some combined conditions. In a similar study, Lacroix *et al.* (1979) found that in the presence of temporal distortions, such as interruption, reverberation, or temporal compression, the degree of intelligibility heavily relies on frequencies above 2 kHz. In our study, our listeners showed more tolerance to LP filtering. The difference could be due to the more severe filter slope, at 96 dB/octave, of the previous study. Regardless, both studies indicate that in the absence of information redundancy provided by the HF speech components, interruptions lead to a greater loss of intelligibility, possibly because of the degradation of low-level acoustic speech cues (due to LP filtering) that are needed for the top-down repair mechanism in the case of interrupted speech. Understanding a sentence requires understanding its component words. Interruption causes obliteration of complete or partial words, which may cause a disruption in linguistic context. Due to LP filtering, the remaining (partial) words lose the robust low-level speech cues that may be necessary to access higher order linguistic information (Shafiro *et al.*, 2011). Hence, combined, this would result in the loss of intelligibility.

Since we wanted to explore the effect of audibility alone on the perception of interrupted speech, we employed young NH listeners, and simulated the loss of HF speech cues due to hearing loss with LP filtering. Thus, we could eliminate other factors relevant to hearing impairment, such as the suprathreshold deficits or age-related cognitive decline. A more deleterious effect of the combination of two distortions, e.g., spectral smearing and background noise, as compared with sole distortion, e.g., spectral smearing, is known (Baer and Moore, 1993). Similarly, the combined effect of suprathreshold auditory deficits and periodic interruptions is known to be detrimental to speech intelligibility, e.g., amplitude modulation and frequency modulation filtering and interruptions (Gnansia *et al.*, 2010; Nelson and Jin, 2004; Gilbert *et al.*, 2007). Since speech has a lot of inherent redundancy, the loss of audibility should not affect the intelligibility. This is why the combined effect of audibility loss and interruption found in this study is interesting. It seems that richness in the speech is important for understanding interrupted speech. The findings of this study imply that apart from suprathreshold or cognitive deficits, the decrease in audibility itself may be a major reason for poor intelligibility of interrupted speech in HI listeners (Jin and Nelson, 2006; Başkent, 2010).

The presence of suprathreshold deficits and their potential effects on perception of degraded speech is still under debate. For example, Fabry and van Tasell (1986) concluded that the effect of sensorineural hearing loss on speech perception is simply due to the reduction or elimination of speech cues from the loss in hearing sensitivity. On the other hand, Başkent (2006) observed that, despite correcting for presentation levels, HI listeners did not seem to take advantage of increased spectral resolution for perception of spectrally degraded speech. While the results of the present study do not support or reject potential effects of such other factors on perception of interrupted speech, they at least imply that a reduction in speech redundancy alone can have substantial detriment to perception of degraded speech.

Thus, in line with previous research (Schnotz *et al.*, 2009; Chatterjee *et al.*, 2010), we conclude that understanding linguistic context, which is a top-down process, is necessary to restore words, but bottom-up auditory cues are necessary to understand the context words. The bottom-up disruption in audibility hampers the top-down restoring of the interrupted speech. Hence, the loss of intelligibility of interrupted speech for listeners with HF hearing loss might mainly be caused due to loss of audibility of HF components, in addition to the potential suprathreshold or cognitive deficits.

Acknowledgment

This study was funded by the Rosalind Franklin Fellowship from the University Medical Center Groningen, University of Groningen, and the VIDI grant from the Netherlands Organisation for Scientific Research (NWO).

References and links

Baer, T., and Moore, B. C. J. (**1993**). “Effects of spectral smearing on the intelligibility of sentences in noise,” J. Acoust. Soc. Am. **94**, 1229–1241.

Başkent, D. (**2006**). “Speech recognition in normal hearing and sensorineural hearing loss as a function of the number of spectral channels,” J. Acoust. Soc. Am. **120**, 2908–2925.

Başkent, D. (**2010**). “Phonemic restoration in sensorineural hearing loss does not depend on baseline speech perception scores,” J. Acoust. Soc. Am. **128**, EL169–EL174.

Başkent, D., Eiler, C. L., and Edwards, B. (**2010**). “Phonemic restoration by hearing-impaired listeners with mild to moderate sensorineural hearing loss,” Hear. Res. **260**, 54–62.

Chatterjee, M., Peredo, F., Nelson, D., and Başkent, D. (**2010**). “Recognition of interrupted sentences under conditions of spectral degradation,” J. Acoust. Soc. Am. **127**, 37–41.

Fabry, D. A., and van Tasell, D. J. (**1986**). “Masked and filtered simulation of hearing loss: Effects on consonant recognition,” J. Speech Hear. Res. **29**, 170–178.

Fletcher, H. (**1953**). *Speech and Hearing in Communication* (Van Nostrand, New York).

Gilbert, G., Bergeras, I., Voillery, D., and Lorenzi, C. (**2007**). “Effects of periodic interruptions on the intelligibility of speech based on temporal fine-structure or envelope cues,” J. Acoust. Soc. Am. **122**, 1336–1339.

Gnansia, D., Pressnitzer, D., Pean, V., Meyer, B., and Lorenzi, C. (**2010**). “Intelligibility of interrupted and interleaved speech in normal-hearing listeners and cochlear implantees,” Hear. Res. **265**, 46–53.

Gordon-Salant, S., and Fitzgibbons, P. J. (**1993**). “Temporal factors and speech recognition performance in young and elderly listeners,” J. Speech Hear. Res. **36**, 1276–1285.

Horwitz, A. R., Dubno, J. R., and Ahlstrom, J. B. (**2002**). “Recognition of low-pass-filtered consonants in noise with normal and impaired high-frequency hearing,” J. Acoust. Soc. Am. **111**, 409–416.

Jin, S-H., and Nelson, P. B. (**2006**). “Speech perception in gated noise: The effects of temporal resolution,” J. Acoust. Soc. Am. **119**, 881–889.

Jin, S-H., and Nelson, P. B. (**2010**). “Interrupted speech perception: The effects of hearing sensitivity and frequency resolution,” J. Acoust. Soc. Am. **128**, 3097–3108.

Lacroix, P. G., Harris, J. D., and Randolph, K. J. (**1979**). “Multiplicative effects on sentence comprehension for combined acoustic distortions,” J. Speech Hear. Res. **22**, 259–269.

Miller, G. A., and Licklider, J. C. (**1950**). “The intelligibility of interrupted speech,” J. Acoust. Soc. Am. **22**, 167–173.

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.

Owens, E., Benedict, M., and Schubert, E. D. (**1972**). “Consonant phonemic errors associated with pure-tone configurations and certain kinds of hearing impairment,” J. Speech Hear. Res. **15**, 308–315.

Powers, G. L., and Speaks, C. (**1973**). “Intelligibility of temporally interrupted speech,” J. Acoust. Soc. Am. **54**, 661–667.

Schnotz, A., Digeser, F., and Hoppe, U. (**2009**). “Speech with gaps: Effects of periodic interruptions on speech intelligibility,” Folia Phoniatr. Logop. **61**, 263–268.

Shafiro, V., Sheft, S., and Risley, R. (**2011**). “Perception of interrupted speech: Cross-rate variation in the intelligibility of gated and concatenated sentences,” J. Acoust. Soc. Am. **130**, 106–114.

Studebaker, G. A. (**1985**). “Research Note: A ‘rationalized’ arcsine transform,” J. Speech Hear. Res. **28**, 455-462.

Versfeld, N. J., Daalder, L., Festen, J. M., and Houtgast, T. (**2000**). “Method for the selection of sentence materials for efficient measurement of the speech reception threshold,” J. Acoust. Soc. Am. **107**, 1671–1684.

Walden, B. E., Schwartz, D. M., Montgomery, A. A., and Prosek, R. A. (**1981**). “A comparison of the effects of hearing impairment and acoustic filtering on consonant recognition,” J. Speech Hear. Res. **24**, 32–43.