The effect of hearing impairment on the identification of speech that is modulated synchronously or asynchronously across frequency

Joseph W. Hall III, **Emily Buss**, and **John H. Grose**
Department of Otolaryngology/Head and Neck Surgery, University of North Carolina School of Medicine, Chapel Hill, NC 27599

Abstract

This study investigated the effect of mild-to-moderate sensorineural hearing loss on the ability to identify speech in noise for vowel-consonant-vowel tokens that were either unprocessed, amplitude modulated synchronously across frequency, or amplitude modulated asynchronously across frequency. One goal of the study was to determine whether hearing-impaired listeners have a particular deficit in the ability to integrate asynchronous spectral information in the perception of speech. Speech tokens were presented at a high, fixed sound level and the level of a speech-shaped noise was changed adaptively to estimate the masked speech identification threshold. The performance of the hearing-impaired listeners was generally worse than that of the normal-hearing listeners, but the impaired listeners showed particularly poor performance in the synchronous modulation condition. This finding suggests that integration of asynchronous spectral information does not pose a particular difficulty for hearing-impaired listeners with mild/moderate hearing losses. Results are discussed in terms of common mechanisms that might account for poor speech identification performance of hearing-impaired listeners when either the masking noise or the speech is synchronously modulated.

INTRODUCTION

The purpose of this study was to gain insight into the factors affecting the ability of listeners with sensorineural hearing impairment to understand speech in the presence of masking noise. It is widely recognized that hearing-impaired listeners often show particular difficulty in understanding speech at poor signal-to-noise ratios (e.g., Plomp and Mimpen, 1979; Festen and Plomp, 1990; Peters et al., 1998). The present study focused on the ability of hearing-impaired listeners to integrate asynchronous, spectrally-distributed speech information. We have noted that such an ability may be of benefit in processing signals in fluctuating noise backgrounds, where the signal-to-noise ratio varies dynamically with respect to both frequency and time (Buss et al., 2003). For example, in a given temporal epoch the signal-to-noise ratio may be favorable at Frequency A but not at Frequency B, but in a successive epoch the signal-to-noise ratio may be favorable at Frequency B but not Frequency A. In such cases, a listener may construct an auditory target from asynchronous information arising from different spectral regions.

Previous studies of speech perception in normal-hearing listeners have demonstrated evidence for the ability to combine spectrally-distributed, asynchronous speech information. An innovative "checkerboard speech" study by Howard-Jones and Rosen (1993) was perhaps the

---

Corresponding author, Joseph W. Hall III.

first study to test this ability rigorously. That study assessed the perception of consonants in a masker composed of multiple, contiguous noise bands that were amplitude modulated. In one condition, modulation of neighboring bands was out of phase, such that when odd-numbered bands were gated on, even numbered bands were gated off, and vice versa; in this condition, masking noise formed a 'checkerboard' pattern when displayed as a spectrogram. Speech identification thresholds in the checkerboard conditions were sometimes better than those obtained in conditions where only the odd-numbered masker bands or only the even numbered masker bands were present, and this was interpreted as reflecting the combination of asynchronous cues for speech identification. Evidence for the combination of asynchronous, spectrally-distributed speech information has also been obtained in studies where speech is amplitude modulated in such a way that the availability of simultaneous across-frequency consonant information is limited (Buss et al., 2003;2004a). Carlyon et al. (2002) also showed that vowels can be identified when two formants are presented such that they do not overlap in time. Overall, these results suggest a robust ability to integrate spectrally-distributed, asynchronous speech cues and are consistent with the notion that successful hearing in noise may involve the processing of successive "glimpses" of spectrally and temporally fragmented target sounds (e.g., Miller and Licklider, 1950; Howard-Jones and Rosen, 1993; Assmann and Summerfield, 2004; Buss et al., 2004a; Cooke, 2006). The particular approach used here follows that of Buss et al. (2004a) where consonant identification in a VCV context is determined under two different conditions of modulation. In this approach, speech tokens are filtered into a number of contiguous, log-spaced frequency bands, and the bands are then amplitude modulated such that the pattern of modulation is either in phase across bands (synchronous modulation) or 180° out of phase for adjacent bands (asynchronous modulation).

Although the effect of hearing loss on the ability to integrate asynchronous speech information is largely unknown, a study by Healy and Bacon (2002) provides at least a suggestion that this capacity might be reduced in listeners with sensorineural hearing impairment. These investigators employed a speech perception method similar to that used in several recent studies where speech is divided into a number of frequency bands and the envelope of each band is used to modulate a corresponding frequency region of either a tonal or noise carrier (e.g., Shannon et al., 1995; Turner et al., 1995; Dorman et al., 1997; Turner et al., 1999). Such methods have demonstrated that good speech perception can occur on the basis of envelope fluctuations imposed upon the carrier(s). The study by Healy and Bacon (2002) employed two widely separated tonal carriers. The individual, modulated carriers did not result in recognizable speech, but the pair of modulated carriers did. Of greater potential relevance here was the finding that small temporal delays imposed between the modulated carriers adversely affected performance to different degrees in listeners with normal hearing and hearing impairment. Listeners with normal hearing were able to maintain relatively good performance for temporal delays of 12.5–25 ms between stimuli in the two spectral regions, a finding that is in agreement with previous reports (e.g., Greenberg and Arai, 1998). In contrast, most of the hearing-impaired listeners of Healy and Bacon showed steep declines in performance for such delays. One possible interpretation of this result is that hearing-impaired listeners have a reduced ability to recognize speech on the basis of information that is asynchronous across frequency. A potential practical consequence of this finding is related to signal processing associated with some digital hearing aid strategies. For example, Stone and Moore (2003) noted that digital hearing aid processing that attempts to mimic the frequency selectivity of the normal ear can result in speech delays that vary across frequency. It is possible that such effects could interact with hearing loss to produce undesirable consequences on speech intelligibility.

As noted above, the results of Healy and Bacon (2002) suggest that hearing-impaired listeners may have a reduced ability to integrate asynchronous speech information when compared with normal-hearing listeners under conditions where different speech frequency regions are delayed with respect to each other. Under such circumstances, the temporal relationship of

information across frequency is disrupted. In contrast, when speech is presented in a masker with spectro-temporal modulations, the available speech cues may be sparse but the temporal relationship between those cues is unchanged. The cues presented in such a masker may be simulated using the *band-AM* paradigm, where the speech stimulus is filtered into contiguous narrow bands which are then amplitude modulated independently (Buss et al., 2004a). When amplitude modulation is out of phase across neighboring bands, there are asynchronous cues, and the temporal relationship between those cues that remain is natural and unaffected by stimulus processing. At the outset of this study it was hypothesized that the poorer ability to integrate asynchronous speech information demonstrated by Healy and Bacon (2002) using the *delayed-band* paradigm may reflect a more general reduction in the capacity of hearing-impaired listeners to combine asynchronous, across-frequency information, like that associated with the *band-AM* paradigm.

Methods

Listeners

Listeners with normal hearing and listeners with sensorineural hearing loss participated. There were eight normal-hearing listeners, two male and six female, ranging in age from 24 to 55 years (with a mean age of 36.5 years and standard deviation of 11.7 years). These listeners had no history of hearing problems, and had pure-tone thresholds of 15 dB HL or better at octave frequencies from 250 Hz to 8000 Hz (ANSI, 2004). The hearing-impaired listeners had mild to moderate sensorineural hearing losses that were relatively flat in configuration (see Table I). There were seven listeners in this group, two male and five female, ranging in age from 24 to 55 years (with a mean age of 46.2 years and standard deviation of 5.4 years). All listeners were paid for their participation.

Stimuli

The stimulus processing was very similar to that used by Buss et al. (2004a) in a study of normal-hearing listeners. The stimuli were 12 vowel-consonant-vowel (VCV) tokens spoken by an American English speaking female. There were five separate samples of each of the 12 tokens, and each token was in the form /a/C /a/ (e.g., /aka/). The consonants were b, d, f, g, k, m, n, p, s, t, v, and z. The VCV samples were 528–664 ms in duration with a mean duration of 608 ms. The speech tokens were scaled to have equal total RMS level, and the presentation level of the tokens prior to modulation was approximately 75 dB SPL. The tokens were digitally filtered into 4 or 16 contiguous bands, with edge frequencies logarithmically spaced from 0.1 to 10 kHz. Our previous research (Buss et al., 2004a) indicated that listeners with normal hearing showed evidence of an ability to integrate asynchronous spectral information in the perception of speech for 2, 4, 8, or 16 contiguous bands. The present choice of 4 and 16 bands was motivated by a desire to determine whether listeners with hearing impairment also showed an ability to integrate asynchronous spectral information in the perception of speech over a similarly wide range of spectral bands. The filtered speech tokens were saved to disk in separate files containing either the odd-numbered bands or the even-numbered bands. Modulation was accomplished by multiplying the speech tokens by a 20-Hz, raised square wave having a 50% duty cycle. The abrupt duty cycle transitions were replaced by 5-ms $\cos^2$ ramps.

In one condition, unprocessed speech tokens were presented. This condition was intended to allow an assessment of group differences related to the identification of unprocessed speech in noise. All other conditions were associated with modulated speech. In *just-odd* conditions, only the odd-numbered bands were included and modulation began with the "on" half of the duty cycle. In the *just-even* conditions, only the even-numbered bands were included and modulation began with the "off" half of the duty cycle. The *asynchronous* conditions were formed by the combination of the *just-odd* and *just-even* stimuli as defined above. For the

synchronous condition, both the odd and even numbered bands were included; modulation was in phase across all bands, and the modulation began with the "on" half of the duty cycle. Previous findings (Buss et al., 2004a) showed that the starting phase of the modulation cycle ("on" versus "off") did not affect the outcome for these stimuli when modulation was synchronous across bands. Note that the number of bands (4 or 16) did not matter for the synchronous conditions, as the contiguous bands formed identical stimuli when bands were modulated synchronously.

The masker was a noise with a spectral shape matching the long-term spectrum of the speech stimuli. The level of this masker was adjusted adaptively, and played continuously over the duration of a threshold track. Stimuli were presented monaurally via a Sennheiser HD265 headphone. For normal-hearing listeners, the stimuli were presented to the left ear. For hearing-impaired listeners, the ear tested was the poor ear in the one case of a unilateral hearing loss, the better ear in the one case where hearing thresholds were beyond audiometric limits in one ear, and was assigned randomly in cases of bilateral hearing loss. A continuous, 50-dB SPL speech-shaped noise was presented to the contralateral ear of all listeners in order to mask possible cross-over to the contralateral ear.

Procedure

In stage 1 of the procedure, the threshold level for detecting the speech-shaped masking noise was obtained. This threshold was ascertained because the main procedure determined the level of the speech-shaped noise that just masked the identification of the speech token. It was therefore important to know whether the masker level that just masked the speech was above the detection threshold of the masker, in order to identify possible floor effects. An observation interval was marked visually, and the listener pressed a button to report whether or not the sound had been detected. Based on the listener's response, the experimenter raised or lowered the noise level in steps of 2 dB and bracketed the threshold level based upon the yes/no responses of the listener. In stage 2 of the procedure, listeners were presented with the samples of the unprocessed and modulated VCVs presented in quiet. The purpose of this stage was to give the listeners a general familiarity with the speech material on which they would be tested.

In stage 3 of the procedure, computer-controlled threshold runs were obtained to determine masked VCV identification thresholds. During trials on these runs, listeners were presented with a randomly selected VCV token. The listeners were then visually presented with the 12 possible consonants and asked to enter a response via the keyboard. No feedback regarding the correct response was provided. The level of the masker was adjusted using a 1-up, 1-down adaptive rule, estimating the masker level necessary to obtain 50% correct identification. There were 26 total reversals per threshold run. The first two reversals were made with 3-dB steps, and the final 24 reversals were made in steps of 2 dB. The threshold estimate was taken as the mean masker level at the last 24 track reversals. Three to five threshold runs were obtained and averaged for each condition. If, during the course of a threshold run, a total of six reversal values occurred at or below the masker threshold (as determined in stage 1), it was assumed that the threshold was influenced by a floor effect, and the masked VCV identification threshold was considered to be unmeasurable.

Results

Fig. 1 provides a summary of the results of the conditions where the speech was unprocessed, modulated synchronously, and modulated asynchronously. All of the normal-hearing and hearing-impaired listeners obtained measurable thresholds in these conditions. Table II shows individual data for these conditions and also for the conditions where only odd or even bands were present. The latter data will be considered last because there were many cases where the thresholds of the hearing-impaired listeners were unmeasurable in these conditions. In

interpreting the masked threshold values from the experimental conditions, it should be kept in mind that higher values represent better performance, because threshold represents the masker level associated with 50% correct identification of a fixed-level speech signal.

Group differences for unprocessed speech

In agreement with the findings of several previous studies (Plomp and Mimpen, 1979; Dubno et al., 1984; Glasberg and Moore, 1989; Turner et al., 1992; Peters et al., 1998), the present results for the unprocessed speech indicated that the hearing-impaired listeners required a higher signal-to-noise ratio than the normal-hearing listeners at the masked speech identification threshold. As shown in the points to the left of Fig. 1, the mean masker level at threshold was 77.0 dB SPL for the normal-hearing listeners and 74.3 dB SPL for the hearing-impaired listeners. An independent samples t-test indicated that this difference was statistically significant (t_{13}=2.8;p=0.01).

Modulated speech conditions

The results of central interest in the present study concerned differences in performance between the synchronously modulated and asynchronously modulated speech conditions. Our previous data for normal-hearing listeners and a 20-Hz rate of modulation indicated a relatively good ability to integrate asynchronous, across-frequency speech information, with masked thresholds being approximately the same for conditions of synchronous and asynchronous modulation (Buss et al., 2004a). The present results are in agreement with this: the normal-hearing listeners had an average threshold of 70.4 dB SPL for both the 4-band and 16-band asynchronous conditions, which compares closely to the average threshold of 70.2 dB SPL for the synchronous modulation condition.

Inspection of Fig. 1 suggests generally poorer performance by the hearing-impaired listeners in the modulated speech conditions, with particularly poor performance for the synchronous modulation condition. In order to evaluate this impression, a repeated measures analysis of variance was performed on the modulated speech data with a within-subjects factor of condition (synchronous modulation, 4-band asynchronous modulation, and 16-band asynchronous modulation) and a between-subjects grouping factor of hearing impairment. Results of this analysis indicated a significant effect of condition ($F_{2,26}$=5.4; p=0.01), a significant effect of hearing impairment ($F_{1,13}$=7,100; p<0.001), and a significant interaction between condition and hearing impairment ($F_{2,26}$=4.3; p=0.02). The interaction reflects the finding that whereas the normal-hearing listeners showed comparable performance across the asynchronous and synchronous modulated speech conditions, the hearing-impaired listeners performed relatively poorly in the synchronous modulation condition. A further indication of the poor performance of the hearing-impaired listeners in the synchronously modulated speech condition is the fact that the hearing-impaired were, on average, 8.2 dB worse than normal in this condition, compared to only 2.7 dB worse than normal in the unprocessed speech condition.

Performance in just-odd and just-even conditions

As can be seen in Table II, many of the thresholds of the hearing-impaired listeners were at floor (unmeasurable) in the conditions where just even numbered or just odd numbered bands were present. Statistical analyses were therefore not performed for these thresholds. The available data are nevertheless informative. For the normal-hearing listeners, where no thresholds were unmeasurable, it was always the case that the thresholds for the asynchronous conditions were better than for either the *just-odd* or *just-even* conditions. This finding is consistent with the interpretation that the threshold in the asynchronous condition did not simply reflect the better of the *just-odd* or *just-even* conditions, but instead reflected an integration of the asynchronously presented information; a previous study, where data were collected on individual speech tokens, also supported this interpretation (Buss *et al.*, 2004). A

similar pattern of results was found for the hearing-impaired listeners, where the thresholds for the asynchronous conditions were better than the thresholds associated with the *just-odd* or *just-even* conditions (with a number of the *just-odd* or *just-even* thresholds being unmeasurable).

General discussion

The discussion begins with the central question of the study, whether hearing-impaired listeners appear to have a deficit in the ability to utilize spectrally distributed, asynchronous cues for speech identification. Possible accounts are then considered for the effects of hearing loss on the processing of synchronously and asynchronously modulated speech signals.

Performance for synchronous versus asynchronous modulation

The primary motivation for the present study was to determine whether sensorineural hearing impairment might be associated with a reduced ability to integrate asynchronous, spectrally-distributed speech information. Because this ability may aid the processing of speech at poor signal-to-noise ratios, such a reduced capacity might help to account for the commonly reported difficulties of hearing-impaired listeners understanding speech in noisy backgrounds. Rather than indicating that hearing-impaired listeners have particular difficulty processing asynchronous speech information, the present results indicated that many hearing-impaired listeners actually performed better for asynchronous modulation of spectrally-distributed speech information than for synchronously modulated speech. As noted in the introduction, a possible interpretation of the results of Healy and Bacon (2002) is that hearing-impaired listeners have a deficit in the ability to integrate asynchronous speech information: in that study listeners with sensorineural hearing loss were more adversely affected than normal-hearing listeners by small temporal asynchronies between spectrally separated bands of speech information. The fact that the present study found that hearing-impaired listeners performed better for sparse cues that are distributed over time than for cues that are clustered in time (asynchronous vs synchronous AM) does not constitute a conflict with the findings or conclusions of Healy and Bacon, however: these two studies used very different means of presenting speech information and of manipulating asynchrony. The present results suggest that mild/moderate sensorineural hearing loss is not associated with a general deficit in the ability to integrate spectrally-distributed, asynchronous speech information. Before considering the results of the asynchronous modulation conditions further, we will discuss the possible significance of the poor performance of the hearing-impaired listeners in the synchronous modulation condition.

Performance of hearing-impaired listeners for modulated speech and its possible relation to performance in modulated noise

As noted above, the average difference in performance between the normal-hearing listeners and the hearing-impaired listeners was relatively large in the synchronous modulation condition. This finding is noteworthy because of its possible relation to previous results from studies where the masking noise was modulated instead of the speech signal (e.g., Wilson and Carhart, 1969; Festen and Plomp, 1990; Takahashi and Bacon, 1992; Eisenberg et al., 1995; Peters et al., 1998; George et al., 2006). A common outcome in such studies is that performance deficits for speech identification by hearing-impaired listeners are greater in modulated noise than in unmodulated noise. Although reduced audibility may sometimes contribute to this effect, the effect persists when the audibility factor is controlled (e.g., Eisenberg et al., 1995; George et al., 2006). Perhaps the most obvious factor that might underlie poor speech identification performance in modulated noise by hearing-impaired listeners is reduced temporal resolution (e.g., greater than normal forward masking), with a resulting reduction in the ability to take advantage of the good signal-to-noise ratios associated with masker envelope

minima (e.g., Zwicker and Schorn, 1982; Festen and Plomp, 1990). The present finding that hearing-impaired listeners also show relatively poor performance when the speech is synchronously modulated instead of the masker suggests that factors in addition to temporal resolution may play an important role in the relatively poor speech identification performance obtained by hearing-impaired listeners in modulated noise. This follows because, in the present modulated speech masking conditions, it would seem reasonable to attribute most of the masking to the speech-shaped masker energy that occurs *simultaneously* with the speech signal rather than to forward masking.

As noted above, poor speech identification performance of hearing-impaired listeners in modulated noise persists when the audibility factor is controlled (e.g., Eisenberg et al., 1995; George et al., 2006). In the present paradigm, a single speech level was employed and it is therefore difficult to draw firm conclusions about possible effects related to audibility. However, there are bases for speculating that audibility was not a major factor contributing to the difference in results between the synchronous and asynchronous modulation conditions obtained here for the hearing-impaired listeners. First, because the speech level was the same in the synchronous and asynchronous modulation conditions, it seems reasonable to infer that the difference in results between these conditions depended upon factors other than audibility. Second, if audibility contributed strongly to the pattern of results in the hearing-impaired listeners, significant correlations might be expected between audiometric thresholds and the speech identification threshold in synchronously modulated speech (which was relatively poor in the hearing-impaired group), and/or between audiometric thresholds and the difference in speech identification thresholds for synchronously versus asynchronously modulated speech (which was relatively large in the hearing-impaired group). To examine this question, the correlation was determined between these speech measures and the audiometric thresholds averaged over 500, 1000, 2000, and 4000 Hz. The correlations with audiometric threshold did not approach significance for either the speech identification threshold for synchronously modulated speech ($r=0.24$; $p=0.61$), or for the difference in speech identification thresholds for synchronously versus asynchronously modulated speech ($r=0.34$; $p=0.46$).

The fact that hearing-impaired listeners show a marked deficit in speech identification performance if either the masking noise *or* the speech is modulated invites speculation about a processing mechanism that may be common to both types of modulation. Two related (and not mutually exclusive) possibilities are considered below[1].

1. Interaction between hearing impairment and speech redundancy

One possibility that could account for poor speech perception in hearing-impaired listeners when either the masking noise or the speech is synchronously modulated involves an interaction between hearing impairment and a reduction of speech redundancy. One feature that is shared whether the noise or the speech is modulated is that some of the listening epochs are systematically corrupted (by periodically increasing the masker level or decreasing the speech level). It is possible that the reduction of speech cue redundancy resulting from noise or speech modulation interacts with the degradation of speech cues associated with sensorineural hearing loss (Plomp, 1978) to produce a more substantial deterioration than

---

[1]In these accounts, it is assumed that the same factors may underlie the relatively poor performance for hearing-impaired listeners when either the speech signal or a masking noise is modulated. Although these accounts have the virtue of parsimony, a caveat is that there are potentially important perceptual differences associated with the noise modulation and speech modulation paradigms. For modulated noise, informal listening suggests that speech has a relatively natural quality at both high and low signal-to-noise ratios (e.g., 10–15 dB above the 50% identification threshold versus a few dB above the 50% identification threshold). For modulated speech, the speech signal has a somewhat unnatural quality at high signal-to-noise ratios, but a more natural quality at low signal-to-noise ratios, perhaps related to the induction effect, where background noise may promote "filling in" of missing parts of an auditory image (e.g., Bashford and Warren, 1987). Thus, although the modulated noise and modulated speech paradigms have conceptual similarities, they are associated with perceptual differences, and it is not clear how such perceptual differences might interact with the factor of hearing impairment.

occurs for unprocessed speech in unmodulated masking noise. This account is similar to that developed by Baer and Moore (1994) who spectrally "smeared" speech material presented to normal-hearing listeners in order to simulate the effects of reduced frequency selectivity that is common in listeners with sensorineural hearing impairment. Using this technique, both Baer and Moore (1994) and ter Keurs et al. (1993) have reported that spectral smearing has a greater deleterious effect for a fluctuating masker than for a steady masker. Baer and Moore reasoned that although the fluctuating masker was associated with sporadic good listening epochs (masker envelope minima), the masker energy present during envelope maxima tended to reduce the redundancy of the speech signal and interacted with the spectral smearing manipulation to cause a relatively large performance deficit. The results of the present study, where speech was modulated, and past studies, where noise was modulated, are consistent with an interpretation that corruption of speech information associated with modulation of either the speech or the noise can interact with the further corruption of speech information that results from hearing impairment to cause such a performance deficit. Although psychoacoustic abilities were not evaluated in the present hearing-impaired listeners, one likely way in which the coding of the auditory stimulus could have been corrupted is reduced frequency selectivity (e.g., Tyler et al., 1984; Stelmachowicz et al., 1985; Leek and Summers, 1996), analogous to the spectral smearing manipulation that was used in the Baer and Moore study.

The above reasoning is compatible with the present finding that the results of the hearing-impaired listeners were generally less abnormal for the asynchronously modulated speech. For example, it is possible that the spectral discontinuities associated with asynchronous speech modulation may have had the beneficial consequence of reducing the potential for interactions among widely separated spectral components of speech. Such interactions would be minimal in the normal ear, which is highly frequency selective, but could occur in listeners with sensorineural hearing loss. This account has some similarity to previous research on the question of whether signal pre-processing can aid speech identification in hearing-impaired listeners. Several investigators have noted that it is theoretically possible to ameliorate effects related to poor frequency selectivity via forms of speech signal processing intended to sharpen the spectrum of the speech signal (e.g., Summerfield et al., 1985; Simpson et al., 1990; Baer et al., 1993; Miller et al., 1999). Although such approaches have not resulted in large improvements in the speech identification abilities of hearing-impaired listeners, they have met with modest success (e.g., Baer et al., 1993). The asynchronous modulation of speech investigated here, with its reduced potential for interaction among widely-spaced spectral components, might be regarded as a special case of "spectral sharpening" when compared with the synchronous condition.

2. Temporal envelope versus temporal fine structure cues

A second, related, possibility that could account for poor performance of hearing-impaired listeners when either the masker or the speech stimulus is synchronously modulated concerns the relative utility of temporal envelope versus temporal fine structure cues for speech perception. Although both temporal envelope and temporal fine structure cues contribute importantly to normal speech perception (Rosen, 1992), it has been hypothesized that the ability to process speech envelope cues in modulated noise may be limited, due to the modulations of the masker interfering with the ability of the listener to process the envelope modulations of the speech stimulus (e.g., Nelson et al., 2003; Lorenzi et al., 2006). By this account, envelope cues have a reduced role for speech perception in modulated noise, elevating the relative importance of temporal fine structure speech cues that are available in the envelope minima of the masker. Although the results of psychoacoustical (e.g., Bacon and Viemeister, 1985) and speech studies (e.g., Turner et al., 1995) suggest that hearing-impaired listeners are often capable of using temporal envelope cues well (at least for stimuli presented in quiet), there is growing behavioral evidence that such listeners often have a reduced ability to benefit from

temporal fine structure cues (e.g., Lacher-Fougere and Demany, 1998; Moore and Moore, 2003; Buss et al., 2004b). Lorenzi et al. (2006) hypothesized that poor speech perception of hearing-impaired listeners in modulated noise may result from 1) the increased importance of temporal fine structure cues for speech in such noise, and 2) the reduced ability of listeners with sensorineural hearing loss to utilize temporal fine structure cues. This reasoning is also consistent with the present finding that listeners with sensorineural hearing loss showed relatively poor performance for synchronously modulated speech: this external modulation may reduce the utility of speech envelope cues, thereby increasing the importance of temporal fine structure cues for which the hearing-impaired listeners have a diminished processing ability.

The above account raises the question of why asynchronous modulation of speech did not appear to be as problematic as synchronous modulation for the hearing-impaired listeners. One possibility is that it is easier to follow speech envelope cues over time in the asynchronous modulation conditions than in the synchronous modulation condition. It is well known that the envelopes associated with different frequency regions of speech are often correlated (e.g., Remez et al., 1994). This raises the possibility that speech envelope cues can be followed across the odd/even-band phases of asynchronous modulation, allowing better utilization of speech envelope cues than is possible in the case of synchronous modulation. This would be particularly likely in hearing-impaired listeners, as reduced frequency selectivity would increase the likelihood that adjacent odd/even speech bands would stimulate common frequency channels in which speech envelope cues could be represented.

Conclusions

When compared to normal-hearing results, the thresholds of the hearing-impaired listeners were more elevated in the synchronous modulation condition than for unprocessed speech in noise. This finding with modulated speech mirrors that demonstrated in the literature for speech in modulated noise. Given that the present results were not likely to have been influenced significantly by forward masking, these results suggest that factors other than temporal resolution play an important role in the speech perception deficits demonstrated here and those shown previously with modulated noise. One hypothesis for this poorer performance is that the corruption of speech information associated with modulation interacts with the additional corruption of speech information associated with hearing loss. A second, related hypothesis is that synchronous modulation of the speech signal reduces the ability of the listener to benefit from cues related to the speech envelope. With the reduction in the availability of speech envelope cues, the performance of the hearing impaired suffers because the listeners are forced to depend upon a poorly encoded cue, temporal fine structure.

The results of the present study did not suggest a general deficit by hearing-impaired listeners to integrate asynchronous spectral information in the perception of speech. In fact, many of the hearing-impaired listeners in this study showed better performance in the asynchronous conditions, where adjacent bands of speech were modulated out-of-phase, than in the synchronous conditions where the modulation of all bands of speech was in-phase. Two interpretations of this result were considered: 1) the better performance in the asynchronous modulation condition occurred because deleterious effects related to poor frequency selectivity were reduced under conditions of asynchronous modulation; 2) speech envelope cues were more accessible in conditions where the modulation of the speech was asynchronous rather than synchronous.

Acknowledgements

References

ANSI. ANSI S3.6–2004, Specification for audiometers. American National Standards Institute, American National Standards Institute; New York: 2004.

Assmann, PF.; Summerfield, AQ. The perception of speech under adverse conditions. In: Greenberg, S.; Ainsworth, WA.; Popper, AN.; Fay, RR., editors. Speech Processing in the Auditory System. Springer Verlag; New York: 2004.

Bacon SP, Viemeister NF. Temporal modulation transfer functions in normal-hearing and hearing-impaired subjects. Audiology 1985;24:117–134. [PubMed: 3994589]

Baer T, Moore BCJ. Effects of spectral smearing on the intelligibility of sentences in the presence of interfering speech. J Acoust Soc Am 1994;95:2277–2280. [PubMed: 8201124]

Baer T, Moore BCJ, Gatehouse S. Spectral Contrast Enhancement of Speech in Noise for Listeners with Sensorineural Hearing Impairment - Effects on Intelligibility, Quality, and Response-Times. J Rehabil Res Dev 1993;30:49–72. [PubMed: 8263829]

Bashford JA Jr, Warren RM. Effects of spectral alternation on the intelligibility of words and sentences. Percept Psychophys 1987;42:431–438. [PubMed: 3696938]

Buss E, Hall JW 3rd, Grose JH. Spectral integration of synchronous and asynchronous cues to consonant identification. J Acoust Soc Am 2004a;115:2278–2285. [PubMed: 15139639]

Buss E, Hall JW 3rd, Grose JH. Temporal fine-structure cues to speech and pure tone modulation in observers with sensorineural hearing loss. Ear Hear 2004b;25:242–250. [PubMed: 15179115]

Buss E, Hall JW, Grose JH. Effect of amplitude modulation coherence for masked speech signals filtered into narrow bands. J Acoust Soc Am 2003;113:462–467. [PubMed: 12558283]

Carlyon RP, Deeks JM, Norris D, Butterfield S. The Continuity Illusion and Vowel Identification. Acta Acustica United with Acustica 2002;88:408–415.

Cooke M. A glimpsing model of speech perception in noise. J Acoust Soc Am 2006;119:1562–1573. [PubMed: 16583901]

Dorman MF, Loizou PC, Rainey D. Speech intelligibility as a function of the number of channels of stimulation for signal processors using sine-wave and noise-band outputs. J Acoust Soc Am 1997;102:2403–2411. [PubMed: 9348698]

Dubno JR, Dirks DD, Morgan DE. Effects of age and mild hearing loss on speech recognition in noise. J Acoust Soc Am 1984;76:87–96. [PubMed: 6747116]

Eisenberg LS, Dirks DD, Bell TS. Speech recognition in amplitude-modulated noise of listeners with normal and listeners with impaired hearing. J Speech Hear Res 1995;38:222–233. [PubMed: 7731213]

Festen JM, Plomp R. Effects of fluctuating noise and interfering speech on the speech-reception threshold for impaired and normal hearing. J Acoust Soc Am 1990;88:1725–1736. [PubMed: 2262629]

George EL, Festen JM, Houtgast T. Factors affecting masking release for speech in modulated noise for normal-hearing and hearing-impaired listeners. J Acoust Soc Am 2006;120:2295–2311. [PubMed: 17069325]

Glasberg B, Moore B. Psychoacoustics abilities of subjects with unilateral and bilateral cochlear hearing impairments and their relationship to the ability to understand speech. Scandanavian Audiology Supplement 32 1989:1–25.

Greenberg, S.; Arai, T. Speech intelligibility is highly tolerant of cross-channel spectral asynchrony. Joint Proceedings of the Acoustical Society of America and the International Congress on Acoustics; Seattle. 1998.

Healy EW, Bacon SP. Across-frequency comparison of temporal speech information by listeners with normal and impaired hearing. J Speech Lang Hear Res 2002;45:1262–1275. [PubMed: 12546492]

Howard-Jones PA, Rosen S. Uncomodulated glimpsing in "checkerboard" noise. J Acoust Soc Am 1993;93:2915–2922. [PubMed: 8315155]

Lacher-Fougere S, Demany L. Modulation detection by normal and hearing-impaired listeners. Audiology 1998;37:109–121. [PubMed: 9547924]

Leek MR, Summers V. Reduced frequency selectivity and the preservation of spectral contrast in noise. J Acoust Soc Am 1996;100:1796–1806. [PubMed: 8817905]

Lorenzi C, Gilbert G, Carn H, Garnier S, Moore BC. Speech perception problems of the hearing impaired reflect inability to use temporal fine structure. Proc Natl Acad Sci U S A 2006;103:18866–18869. [PubMed: 17116863]

Miller GA, Licklider JCR. The intelligibility of interrupted speech. Journal of the Acoustical Society of America 1950;22:167–173.

Miller RL, Calhoun BM, Young ED. Contrast enhancement improves the representation of/epsilon/-like vowels in the hearing-impaired auditory nerve. J Acoust Soc Am 1999;106:2693–2708. [PubMed: 10573886]

Moore BC, Moore GA. Discrimination of the fundamental frequency of complex tones with fixed and shifting spectral envelopes by normally hearing and hearing-impaired subjects. Hear Res 2003;182:153–163. [PubMed: 12948610]

Nelson PB, Jin SH, Carney AE, Nelson DA. Understanding speech in modulated interference: cochlear implant users and normal-hearing listeners. J Acoust Soc Am 2003;113:961–968. [PubMed: 12597189]

Peters RW, Moore BC, Baer T. Speech reception thresholds in noise with and without spectral and temporal dips for hearing-impaired and normally hearing people. J Acoust Soc Am 1998;103:577–587. [PubMed: 9440343]

Plomp R. Auditory handicap of hearing impairment and the limited benefit of hearing aids. J Acoust Soc Am 1978;63:533–549. [PubMed: 670550]

Plomp R, Mimpen AM. Speech-reception threshold for sentences as a function of age and noise level. J Acoust Soc Am 1979;66:1333–1342. [PubMed: 500971]

Remez RE, Rubin PE, Berns SM, Pardo JS, Lang JM. On the perceptual organization of speech. Psychol Rev 1994;101:129–156. [PubMed: 8121955]

Rosen S. Temporal information in speech: acoustic, auditory and linguistic aspects. Philos Trans R Soc Lond B Biol Sci 1992;336:367–373. [PubMed: 1354376]

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

Simpson AM, Moore BCJ, Glasberg BR. Spectral enhancement to improve the intelligibility of speech in noise for hearing-impaired listeners. Acta Otolaryngologica Suppl 1990;469:101–107.

Stelmachowicz PG, Jesteadt W, Gorga MP, Mott J. Speech perception ability and psychophysical tuning curves in hearing-impaired listeners. J Acoust Soc Am 1985;77:620–627. [PubMed: 3973233]

Stone MA, Moore BCJ. Tolerable hearing aid delays" III Effects on speech production and perception of across-frequency variation in delay. Ear Hear 2003;24:175–183. [PubMed: 12677113]

Summerfield AQ, Foster J, Tyler R, Bailey PJ. Influences of formant narrowing and auditory frequency selectivity on identification of place of articulation in stop consonants. Speech Commun 1985;4:213–229.

Takahashi GA, Bacon SP. Modulation detection, modulation masking, and speech understanding in noise in the elderly. J Speech Hear Res 1992;35:1410–1421. [PubMed: 1494284]

ter Keurs M, Festen JM, Plomp R. Effect of spectral envelope smearing on speech reception II. J Acoust Soc Am 1993;93:1547–1552. [PubMed: 8473608]

Turner CW, Chi SL, Flock S. Limiting spectral resolution in speech for listeners with sensorineural hearing loss. Journal of Speech, Language, & Hearing Research 1999;42:773–784.

Turner CW, Fabry DA, Barrett S, Horwitz AR. Detection and recognition of stop consonants by normal-hearing and hearing-impaired listeners. J Speech Hear Res 1992;35:942–949. [PubMed: 1405550]

Turner CW, Souza PE, Forget LN. Use of temporal envelope cues in speech recognition by normal and hearing-impaired listeners. J Acoust Soc Am 1995;97:2568–2576. [PubMed: 7714274]

Tyler RS, Hall JW, Glasberg BR, Moore BCJ, Patterson RD. Auditory filter asymmetry in the hearing impaired. J Acoust Soc Am 1984;76:1363–1376. [PubMed: 6512098]

Wilson RH, Carhart R. Influence of pulsed masking on the threshold for spondees. J Acoust Soc Am 1969;46:998–1010. [PubMed: 5824043]

Zwicker E, Schorn K. Temporal resolution in hard of hearing patients. Audiology 1982;21:474–492. [PubMed: 7181741]
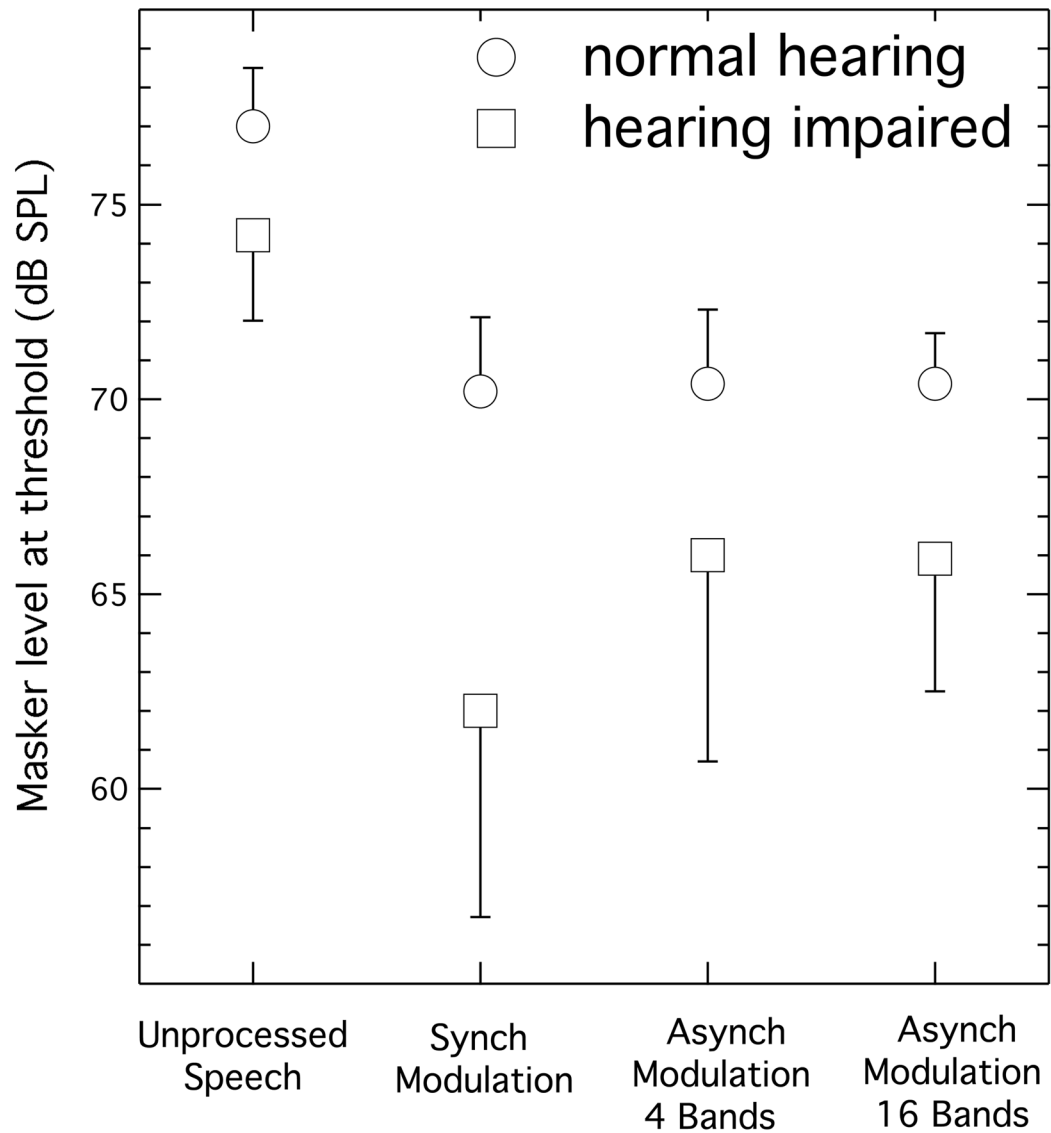
Fig. 1.
Masker level at threshold for the unprocessed speech condition, the synchronous modulation condition, the asynchronous modulation/4-band condition, and the asynchronous modulation/16-band condition. The error bars show +1 standard deviation for the normal-hearing listeners and −1 standard deviation for the hearing-impaired listeners.

Table I

Audiometric data for hearing-impaired listeners (HI1–HI7). The ear tested is associated with the entries in bold. The %Sp column refers to speech recognition for monosyllabic words in quiet with presentation level approximately 30 dB above the speech recognition threshold for spondee words. "NR" indicates that the threshold was above the limits of audiometric testing and "DNT" indicates that the test was not performed.

	Left							Right						
	.25	.5	1.0	2.0	4.0	8.0	%Sp	.25	.5	1.0	2.0	4.0	8.0	%Sp
HI1	**40**	**45**	**45**	**35**	**45**	**85**	**84**	NR	NR	NR	NR	NR	NR	DNT
HI2	50	55	60	60	60	65	92	**40**	**35**	**45**	**45**	**45**	**45**	**88**
HI3	**25**	**30**	**50**	**50**	**50**	**35**	**92**	25	30	40	50	50	40	92
HI4	20	20	20	20	20	15	100	**50**	**55**	**55**	**40**	**40**	**40**	**92**
HI5	35	35	25	30	40	60	100	**25**	**30**	**30**	**30**	**45**	**65**	**100**
HI6	**25**	**40**	**45**	**30**	**55**	**85**	**88**	15	15	10	5	5	10	100
HI7	**25**	**50**	**60**	**45**	**35**	**50**	**88**	10	40	45	45	45	35	96

Table II

Masker level at threshold (dB SPL) for the various experimental conditions is shown for the normal-hearing (N1–N8) and hearing-impaired (HI1–HI7) listeners. The numbers in parentheses are the inter-listener standard deviations. Dashes indicate thresholds that were unmeasurable. Synchronous and Asynchronous modulation conditions are labeled "Synch" and "Asynch," respectively.

| | Unprocessed | -----------4 Bands----------- | | | | --------16 Bands-------- | | |
		Synch	Asynch	Just-Odd	Just-Even	Asynch	Just-Odd	Just-Even
N1	78.9	72.6	73.1	68.3	65.2	72.5	65.3	67.5
N2	79.6	73.4	71.6	59.0	37.9	71.6	61.2	68.5
N3	77.0	70.1	72.1	47.2	55.2	70.5	58.0	63.7
N4	75.8	68.5	67.8	30.7	56.5	69.7	56.0	60.0
N5	76.0	69.4	68.1	35.0	46.9	68.8	55.0	58.3
N6	76.6	69.2	69.9	39.5	45.7	70.1	56.8	63.8
N7	76.4	70.2	69.3	38.9	55.5	68.8	62.1	60.7
N8	75.5	67.8	71.5	47.9	59.6	70.8	53.8	65.9
Mean	77.0 (1.5)	70.2 (1.9)	70.4 (1.9)	45.8 (12.6)	52.8 (8.7)	70.4 (1.3)	58.5 (4.0)	63.6 (3.7)
HI1	72.2	62.8	58.2	-	-	63.0	-	47.1
HI2	76.6	68.4	69.7	-	-	70.1	63.1	-
HI3	75.5	57.0	60.9	-	56.5	64.3	50.0	-
HI4	76.7	70.3	73.3	57.1	-	71.0	62.2	66.1
HI5	74.9	59.0	68.8	-	-	66.8	-	-
HI6	72.3	58.6	67.7	48.6	-	63.1	48.8	46.0
HI7	71.6	58.1	63.5	-	56.0	63.3	-	-
Mean	74.3 (2.2)	62.0 (5.3)	66.0 (5.3)			65.9 (3.4)		