Across-Frequency Comparison of Temporal Speech Information by Listeners With Normal and Impaired Hearing

Listenors with normal hearing (NH) and with sensorineural hearing impairment (HI) were tested on a speech-recognition task requiring across-frequency integration of temporal speech information. Listeners with NH correctly identified a majority of key words in everyday sentences when presented with a synchronous pair of speech-modulated tones at 750 and 3000 Hz. They could tolerate small amounts (12.5 ms) of across-frequency asynchrony, but performance fell as the delay between bands was increased to 100 ms. Listeners with HI performed more poorly than those with NH when presented with synchronous across-frequency information. Further, performance of listeners with HI fell as a function of asynchrony more steeply than that of their NH counterparts. These results suggest that listeners with HI have particular difficulty comparing and effectively processing temporal speech information at different frequencies. The increased influence of asynchrony indicates that these listeners are especially hindered by slight disruptions in across-frequency information, which implies a less robust comparison mechanism. The results could not be attributed to differences in signal or sensation level, or in listener age, but instead appear to be related to the degree of hearing loss. This across-frequency deficit is unlikely to be attributed to known processing difficulties and may exist in addition to other known disruptions.

KEY WORDS: hearing impairment, speech recognition, temporal, asynchrony, across-frequency processing

Listeners with hearing impairment (HI) experience a variety of difficulties, many of which are associated with portions of the acoustic signal lying below threshold. However, even when appropriate audibility is provided, listeners with moderate to severe sensorineural hearing losses may have problems understanding speech. These suprathreshold processing deficits can lead to significant performance difficulties and can be quite intractable. Perhaps one of the reasons these deficits are not well accounted for is that the specific nature of the disruption in processing is not well understood.

One possibility is that temporal processing is disrupted in listeners with HI. Indeed, these listeners may show reduced performance on psychoacoustic measures of temporal processing, such as modulation detection (e.g., Bacon & Viemeister, 1985) or gap detection (e.g., Florentine

* Currently affiliated with the University of South Carolina, Columbia.
Another potential limitation associated with HI involves broadened auditory tuning. Listeners with hearing losses as small as 45 dB HL might have frequency tuning measures two times larger than normal or more (cf. Moore, 1995). Attempts have been made to measure the impact of reduced frequency selectivity on speech perception. Some studies have found a correlation between tuning and performance on consonant recognition (Thibodeau & van Tasell, 1987) or vowel recognition (Turner & Henn, 1989), whereas other studies have found none (for consonants: Dubno & Dirks, 1989; for vowels: Dubno & Dorman, 1987). Another technique is to present listeners having NH with speech stimuli altered to mimic the reduced tuning associated with HI. These studies have found that performance on broadband speech in quiet is quite good despite extensive spectral smearing (Baer & Moore, 1993). Indeed, spectral information can be reduced to contrasts between only three or four broad contiguous amplitude-modulated noise bands with intelligibility remaining high (Shannon, Zeng, Kamath, Wygonski, & Ekelid, 1995). It is only when speech is presented in background noise that performance has been shown to suffer with increased spectral smearing (e.g., Baer & Moore, 1993, 1994; ter Keurs, Festen, & Plomp, 1992, 1993).

Turner et al. (1999) recently conducted an experiment that led them to suggest that an additional suprathreshold processing deficit may exist. The authors speculated that, although listeners with HI can process single or overall temporal speech patterns quite well, they may have difficulty comparing temporal speech patterns at different frequencies. This ability, of course, is required to assemble the spectro-temporal speech stimulus. The experiment that led to this suggestion involved the recognition of consonants represented by 1, 2, 4, and 8 broad contiguous amplitude-modulated noise bands. Although performance for listeners with NH and with HI was similar for one band (demonstrating intact overall temporal processing for the listeners with HI), the groups diverged when the speech spectrum was represented by two or more temporal patterns. Because listeners with HI can apparently process single temporal patterns normally, poorer performance than normal when two or more bands were presented is consistent with a deficit in the ability to process contrasts between temporal patterns.

The ability of listeners with NH to integrate temporal information across frequencies and to tolerate disruptions in across-frequency information was recently examined in a series of experiments (Healy & Bacon, 2000; Healy & Warren, 2002). These experiments differed from those of Turner et al. in that narrowband speech-modulated patterns were used. Narrow bands of speech were used to modulate narrowband noises or tones. The patterns had various frequency separations, allowing an examination of integration across various regions of the speech spectrum. Using these stimuli, it was found that listeners with NH can effectively integrate temporal information both from adjacent frequencies and from widely separated frequencies (Healy & Warren, 2002). Another study found that listeners with NH employ precise across-frequency timing cues provided by the synchronous temporal speech patterns: Sentence intelligibility fell when 25 ms or more of asynchrony was introduced between members of the band pair (Healy & Bacon, 2000). This influence of asynchrony was similar for band pairs residing together in the low-frequency region (1/3-octave bands centered at 530 and 1100 Hz), mid-frequency region (1100 and 2100 Hz), and also the high-frequency region (2100 and 4200 Hz).

These experiments are perfectly suited to examining further the recent suggestion by Turner et al. (1999) involving listeners with HI and to overcome potential limitations of that work. In the study by Turner et al., portions of the signal were inaudible for some of the listeners. This is attributable in part to the use of a broadband stimulus. The use of narrowband temporal patterns within a more restricted range of frequencies (along with high signal levels and listeners having appropriate audiometric configurations) should allow better control over audibility.

Another potential limitation of the previous work involves broadened auditory tuning. In the study by Turner et al. (1999), the listeners with HI performed more poorly than those with NH when only two bands were used, despite the fact that, given their tuning abilities, the listeners with HI should be able to process at least two distinct frequency channels. However, broadened frequency tuning could still play a role in the observed effects by increasing the effective overlap of the contiguous noise bands. It has been previously found that increased acoustic overlap of narrowband temporal patterns can result in lower intelligibility scores (Healy & Warren, 2002). Turner et al. investigated this possibility by introducing a filtered notch (approximately 24 dB in...
depth) to the two-band stimulus. Although the elimination of this information did not increase scores for the subset of listeners with HI who were tested, the use of narrowband temporal patterns allows for far greater channel isolation. In the current experiments, a pair of speech-modulated tones having complete acoustic isolation down to the noise floor (approximately 90 dB) and a nominal spacing of two octaves was used to further reduce auditory interactions and potential influences of tuning.

Finally, in the work by Turner et al. (1999), performance differences attributable to differences in sensation level (SL) were not examined. An additional experiment in the current study provided this information.

The current experiments were designed to provide evidence for or against the suggestion that listeners with HI possess a deficit in their ability to compare temporal speech patterns at different loci. They were designed to overcome potential limitations of the study by Turner et al. (1999) and to extend that study to a situation in which listeners can use natural context cues. Further, the current experiments allow an examination of performance when across-frequency temporal information is disrupted through cross-channel asynchrony. Sentences represented by a pair of narrowband temporal patterns were presented either in temporal synchrony or with various misalignments ranging from 12.5 to 100 ms. The bands were spaced in frequency to eliminate acoustic overlap and to reduce the influence of auditory interactions. They were presented at levels required for full audibility and also at low levels required to examine potential influences of SL. If listeners with HI possess an across-frequency temporal processing deficit, then lower intelligibility scores should be observed when the bands are aligned. Further, the current manipulations allow an examination of the ability of various listeners to tolerate graded disruptions in this across-frequency cue.

Experiment 1: Effect of Asynchrony

Method

Participants

A group of 10 listeners with NH participated. Each had pure-tone audiometric thresholds of 20 dB HL or better at audiometric frequencies from 250 to 8000 Hz (American National Standards Institute [ANSI], 1989) and 15 dB HL or better at the test frequencies (750 and 3000 Hz). The average audiogram (plotted in SPL vs. frequency) for these listeners is shown in Figure 1 as a dotted line without symbols (repeated across panels).

The ages of these participants ranged from 20 to 52 years (mean = 27). An additional group of 10 listeners with HI was also recruited. All had a bilateral sensorineural hearing loss of cochlear origin. They were recruited based on similar pure-tone thresholds, without regard to speech audiometry scores. Pure-tone audiometric thresholds at 750 Hz ranged from 30 to 65 dB HL (average = 47), and those at 3000 Hz ranged from 45 to 75 dB HL (average = 60). Audiograms (SPL vs. frequency) for each of these listeners are shown in separate panels of Figure 1. The ages of these participants ranged from 53 to 79 years (mean = 70). Ages and audiometric thresholds at the test frequencies are listed for the individual listeners with HI in Table 1. All listeners were paid for participation and had no previous exposure to the sentence materials.

Stimuli

The stimuli were based on the Central Institute for the Deaf (CID) “everyday American speech” sentences (Davis & Silverman, 1978; Silverman & Hirsh, 1955). The 100 sentences are arranged in 10 sets of 10 sentences each, with 50 key words in each set for a total of 500 scoring key words. An additional set of 10 practice sentences was taken from the high-predictability subset of the Speech Perception In Noise (SPIN) test (Kalikow, Stevens, & Elliot, 1977). Digital recordings (22 kHz, 16 bit) of the sentences were created by a professional male speaker with a general American dialect from within an audiometric booth.

The sentences were represented by a pair of narrowband temporal patterns created by amplitude-modulating tones by corresponding (frequency-matched) narrow bands of speech. The creation of each speech-modulated tone involved first filtering the sentences to a single 1/3-octave band (96 dB/octave) using dual-cascaded passes

<table>
<thead>
<tr>
<th>Listener</th>
<th>Age</th>
<th>Threshold (dB HL)</th>
</tr>
</thead>
<tbody>
<tr>
<td>HI 1</td>
<td>60</td>
<td>45</td>
</tr>
<tr>
<td>HI 2</td>
<td>79</td>
<td>55</td>
</tr>
<tr>
<td>HI 3</td>
<td>67</td>
<td>65</td>
</tr>
<tr>
<td>HI 4</td>
<td>76</td>
<td>50</td>
</tr>
<tr>
<td>HI 5</td>
<td>65</td>
<td>50</td>
</tr>
<tr>
<td>HI 6</td>
<td>53</td>
<td>30</td>
</tr>
<tr>
<td>HI 7</td>
<td>79</td>
<td>45</td>
</tr>
<tr>
<td>HI 8</td>
<td>73</td>
<td>45</td>
</tr>
<tr>
<td>HI 9</td>
<td>78</td>
<td>45</td>
</tr>
<tr>
<td>HI 10</td>
<td>70</td>
<td>35</td>
</tr>
</tbody>
</table>

Table 1. Ages and audiometric thresholds for listeners with hearing impairment.
Figure 1. Audiometric thresholds (converted to dB SPL) for each of the 10 listeners with hearing impairment (HI) (circles) are presented along with the stimulus levels heard by each listener (crosses). The average audiometric thresholds for the 10 listeners with normal hearing (NH) are represented by the dotted line. Signal levels for the listeners with NH were 90 dB SPL.
through 8th-order digital Butterworth filters. The narrow speech band was then full-wave rectified by taking the absolute value of each sample in the digital file. A sinusoidal tone having a frequency corresponding to the center of the speech band was multiplied with the rectified speech on a sample point-by-point basis. To ensure that temporal fluctuation rates were limited only by the narrowband filtering, no low-pass filtering of the rectified speech was employed. Because the amplitude modulation produced a spectral broadening of the tone through the addition of frequency components that act as modulation side bands, the modulated patterns were refiltered to ensure that all energy was restricted to the original 1/3-octave region. Processing of the stimuli was performed digitally using MATLAB software.

Individual speech-modulated tones at 750 and 3000 Hz were created. The long-term root mean square (RMS) average amplitude spectra of the stimuli is shown in Figure 2. These center frequencies were selected because they provide a wide frequency separation, appropriate to best examine integration and to minimize the potential influence of broadened auditory tuning, and also because they provide high intelligibility when presented in temporal synchrony (Healy & Warren, 2002). The sentences in each band were scaled so that the peak of the slow-response RMS average was within 1 dB for each sentence. The speech-modulated tones were then saved to right and left channels of a stereo file for mixing into a contrasting pair of patterns. The use of a stereo file allowed independent level adjustment of each individual band. Files were created with the bands in temporal synchrony and with asynchronies of 12.5, 25, 50, and 100 ms. Proper alignment (or accurate misalignment) was ensured by correcting for the frequency-specific phase shifts of the IIR (Infinite-duration Impulse Response) filters using an additional correction delay of 9 ms for the high-frequency band relative to the low-frequency band. The value required to correct for the filter delay was determined empirically by passing a single-sample click through the filters used to create the bands and measuring the time between the centers of the resulting broad pulses. Two conditions were prepared for each asynchrony—one in which the high-frequency band led and the low-frequency band lagged and one in which the low-frequency band led and the high-frequency band lagged.

The stereo stimuli were played back from a personal computer using a high-quality digital-to-analog (D/A) converter (EMU 10k1) and were routed to a Mackie mixer (1202 VLZ) for mixing and amplification before transduction by one earphone of Sony headphones (MDR 7506). The listeners with NH heard each band set to a peak level of 90 dB SPL (level of the peaks of the slow-response RMS average), to match the high levels required by the listeners with HI. This level was determined to be the maximum comfortable level for these dual-band stimuli by a panel of five experienced laboratory personnel. The listeners with HI heard each band set to a minimum of 90 dB SPL. However, if the audiometric threshold at either signal frequency was 65 dB HL or worse, the listener was allowed to adjust that individual band to most comfortable loudness (MCL). That level was used unless it exceeded 100 dB SPL, in which case a level of 100 dB SPL was used. The only exception to this was the first listener tested with HI (HI 3), who adjusted both bands simultaneously, rather than individually. Five of the 10 listeners with

Figure 2. Long-term average amplitude spectra for the speech-modulated tones at 750 and 3000 Hz.
HI simply heard the stimuli at 90 dB SPL, 4 performed this MCL adjustment at 3000 Hz, and one performed it for both bands simultaneously. The average stimulus levels for the listeners with HI were 90.3 dB SPL at 750 Hz and 93.7 dB SPL at 3000 Hz. The stimulus levels heard by each individual listener with HI are displayed in Figure 1 as crosses. As the figure shows, the level adjustment procedure maximized audibility while ensuring comfort. Because each of the temporal patterns provide virtually no intelligibility when presented individually (Healy & Warren, 2002), the listeners received no practice from the level-setting procedure. All level measurements were made acoustically following transduction using a sound-level meter and flat-plate coupler.

**Procedure**

Listeners were tested individually, seated with the experimenter in a double-walled audiometric booth. They participated in a single session lasting roughly 1 to 1.5 hours. Pure-tone audiograms were performed for all listeners immediately before testing. Listeners with NH heard the stimuli monaurally in the right ear, and listeners with HI heard them in the more appropriate ear (the ear having more similar thresholds across the two stimulus frequencies, with values closest to 45–60 dB HL).

The listeners started with practice involving the 10 SPIN sentences, hearing first the broadband (unprocessed) signal and then the aligned pair of temporal patterns. During this practice phase, individual sentences were repeated if necessary to familiarize the listeners with the stimuli. They then heard the practice list once more processed to match their first test condition. Following this familiarization phase, the listeners heard the first of the five test conditions (asynchronies of 0, 12.5, 25, 50, or 100 ms). The practice list was presented again before each of the subsequent test conditions, once aligned, and again processed to match the upcoming condition. The listeners heard two lists of 10 CID test sentences in each of the five asynchronous conditions. They repeated the sentences aloud after hearing each, guessing if they were unsure of the content. The experimenter recorded the number of key words correctly reported and controlled the presentation of sentences. Each test sentence was played only once.

The test condition-to-sentence list correspondence was balanced so that each pair of sentence lists was heard once in each condition for each subset of 5 listeners. The order that conditions were heard was balanced so that every condition was heard first across each subset of 5 listeners. The remaining conditions were presented in random order. This balancing procedure was used to control for the potential influence of learning and practice within the test session, which previous analysis has shown to be most pronounced within the first-run condition. Half of the listeners with NH and half with HI heard the high-frequency band lead and the low-frequency band lag, and the other half heard the opposite arrangement. Due to time constraints, the low-intelligibility 100-ms asynchrony condition was skipped for 3 of the 10 listeners with HI.

**Results**

Individual intelligibility scores were calculated for each listener based on the percentage of key words correctly reported, and these scores were averaged to generate group means. A preliminary analysis compared the 5 listeners with NH who heard the high-frequency band lag to the 5 who heard the high-frequency band lead. Consistent with previous research (Healy & Bacon, 2000), the differences between these groups at the four asynchronies were small (average absolute difference = 4.3%), and no significant difference was observed at any of the asynchronies (12.5 to 100 ms; t ≤ 1.1, p > .05). These data were therefore pooled for this and all subsequent analyses and presentations.

The upper panel of Figure 3 shows the group mean intelligibility scores and 95% confidence intervals for both listener groups as a function of dual-band asynchrony. In accord with previous research (Healy & Bacon, 2000), the listeners with NH could correctly report a majority of the key words when the individually unintelligible temporal patterns were aligned or were misaligned by 12.5 ms, but intelligibility dropped as asynchrony increased to 100 ms.

The average performance of the listeners with HI was worse than that of the NH group when the bands were aligned and at all asynchronies. In the lower panel of Figure 3, intelligibility is plotted as a proportion of the score in the aligned condition. This plot shows that cross-channel asynchrony had a larger effect on the listeners with HI, as their scores fell more sharply as a function of asynchrony than did the scores from the NH group. The scores at 50 ms of asynchrony are reduced by 0.67 (to 0.33) for the listeners with NH and by 0.87 (to 0.13) for the listeners with HI relative to the scores.

---

1 To control for the effects of practice within the session, the presentation order of conditions was balanced and randomized. However, because it was planned to examine individual (rather than group mean) data from the listeners with HI, the first-run conditions were repeated after the end of the session to examine potential increases in performance. Only one of the listeners (HI 2) displayed a change in performance. The score in her first-heard condition (aligned) increased from 7% to 33%. The latter score was included in the group means and analysis because it appeared to reflect a truer measure of her abilities (see Figure 4).
in the aligned conditions. This corresponds to a reduction in score proportion of 0.013/ms for those with NH and 0.017/ms for those with HI.

These observations are supported by statistical analysis. A two-way (2 listener groups \times 5 asynchronies) repeated-measures ANOVA revealed significant main effects of listener group \( [F(1, 18) = 22.0, p < .01] \) and asynchrony \( [F(4, 69) = 93.1, p < .01] \). Post hoc analysis indicated that the difference between listener groups was significant at each of the five asynchrony conditions \( (p < .05) \). The interaction in the two-way ANOVA was not significant \( [F(4, 69) = 1.4, p > .05] \), but when the different levels of asynchrony were examined separately within each significant listener group factor [one-way repeated-measures ANOVA for NH: \( F(4, 36) = 57.6, p < .01 \); one-way repeated-measures ANOVA for HI: \( F(4, 33) = 39.6, p < .01 \)], post hoc analysis indicated that intelligibility did not change for the listeners with NH as asynchrony was increased from 0 to 12.5 ms \( (p > .05) \). However there was a significant drop for the listeners with HI with this very small asynchrony \( (p < .05) \). Also, the performance of listeners with HI reached the floor by 50 ms of asynchrony \( (50 vs. 100 ms, p > .05) \), whereas the performance of listeners with NH did not \( (50 vs. 100 ms, p < .05) \).

When the performance of the listeners with HI was examined in further detail, it became clear that these listeners fell into two distinct groups based on performance in the aligned condition. The majority of the listeners (6 of 10) performed markedly worse than the NH group. These individual scores are plotted in the upper panel of Figure 4 along with the group mean for the listeners with NH. Only 1 of the 27 individual scores (recall that 3 of the listeners were not tested at 100 ms of asynchrony) reached the average performance level of the listeners with NH. The average performance for this subgroup of “poor performing” listeners with HI is plotted in the middle panel of Figure 4 along with the group mean for the listeners with NH. The lower panel shows intelligibility as a proportion of performance in the aligned condition.

In contrast, a minority of the listeners with HI performed as well as (or even better than) the NH group in the aligned condition. Individual scores for these 4 listeners are presented in the upper panel of Figure 5, and the group mean for these “well performing” listeners is plotted in the middle panel. Although performance in the aligned condition was quite good for these listeners, 3 of the 4 exhibited a sharp drop in intelligibility as a very small amount of asynchrony was introduced \( (12.5 ms) \). Although some recovery was observed for 2 of these listeners as asynchrony was increased to 25 ms, all scores remain below the NH mean at all asynchronies. The poor-performing listeners with HI do not show this dramatic drop in scores at 12.5 ms (although the data at this delay are quite variable and scores for some listeners are reduced). However, the tendency for performance to drop more sharply with across-channel asynchrony was also present for these listeners: The lower panel of Figure 4 shows that after 12.5 ms, performance fell more steeply for the listeners with HI. Together, these results indicate that only 1 of the 10 listeners with HI (HI 7) exhibited performance similar to that of the listeners with NH up to 25 ms, after which even his performance fell below that of the listeners with NH.

\[ \text{This extremely steep drop was not due to presentation order of conditions, as both the aligned condition and the 12.5-ms asynchrony condition were presented in 2.67th position on average for the 3 listeners who displayed the drop (HI 8, 9, 10). Participant number does not reflect the order of listener participation.} \]

\[ \text{It was later learned that this individual is a former professor of music and professional conductor. He has since participated in other experiments in the laboratory and has displayed remarkable listening skills.} \]
Experiment 2: Effect of Sensation Level

The stimulus configuration and presentation levels ensured that the stimuli were clearly audible for all listeners (see Figure 1). However, significant differences in SL exist between the listeners with NH and those with HI. To examine whether the results obtained in Experiment 1 are attributable to these differences in SL, additional trials with listeners having NH were run at low SPLs matching the SLs of the HI listeners.

Method

Participants

An additional group of 10 listeners having the audiometric qualifications of the NH group in Experiment 1 participated. Their ages ranged from 22 to 42 years (mean = 27).

Stimuli and Procedure

The same dual-band stimuli and apparatus were employed. The average SLs for the well-performing listeners with HI (those shown in Figure 5) were found to be slightly greater (by 6/4 dB at 750/3000 Hz) than those for the poor-performing listeners with HI (shown in Figure 4). It was therefore decided to have half of the current group with NH hear the stimuli at a level corresponding to the well-performing SL and half at a level corresponding to the poor-performing SL. However, it was noted that an even lower “worst case” SL could be obtained if listener HI 1 was omitted from the calculation of the poor-performing SL. Recall from Figure 4 that this listener is the only one to have any score reach the average level of the listeners with NH. Although this listener should still be classified as a poor performer, a more conservative SL comparison could be obtained by omitting this person. The omission of HI 1 lowered the poor-performing SL an additional 0.7/2.5 dB (750/3000 Hz). Five randomly selected listeners with NH therefore heard the stimuli at levels corresponding to the SL of the well-performing listeners with HI in Experiment 1 (39.0/27.3 dB) and another 5 heard them at the SL of the (worst-case) poor-performing listeners with HI (32.1/20.5 dB). The level of each band was adjusted for individual listeners based on their pure-tone audiometric thresholds measured immediately before testing.

As before, the presentation order of conditions was balanced and randomized, and the sentence list-to-condition correspondence was balanced for each subgroup of 5 listeners. Half the listeners heard the high-frequency band lead and the low-frequency band lag, and the other half heard the opposite arrangement. All other procedures were the same as those described previously.

Results

The first analysis examined differences between the two subgroups of listeners hearing the higher versus the lower SLs. These differences were small (average absolute difference = 3.9%), and no significant difference was observed at any of the five asynchrony conditions (0 to 100 ms; t ≤ 1.2, p > .05). Indeed, the scores produced by those hearing the lower SL were actually better than the scores produced by those hearing the
higher SL by a (signed) average of 2% across conditions. These data were therefore pooled for further analysis and presentation.

The upper panel of Figure 6 displays the results from Experiment 1 for the listeners with NH and HI at high SPLs (from Figure 3), along with the pooled data for the listeners with NH at low SLs. Scores for these listeners are somewhat reduced compared to their high-SPL NH counterparts, but they are not reduced to the level of the listeners with HI. A two-way (3 listener groups × 5 asynchronies) repeated-measures ANOVA revealed significant main effects of listener group \(F(2, 27) = 14.8, p < .01\) and asynchrony \(F(4, 104) = 163.3, p < .01\) and a nonsignificant interaction \(F(8, 104) = 0.8, p > .05\). Post hoc analysis on the scores collapsed across asynchronies indicated that the low-SL group did not differ from its NH counterparts \((p > .05)\), but did differ from the HI group \((p < .05)\).

The lower panel of Figure 6 shows that scores from the low-SL group fell as a function of asynchrony somewhat more steeply than those of their high-SPL NH counterparts, but not as steeply as the listeners with HI.

### Discussion

**Listener Groups**

**Listeners With NH**

The ability of listeners with NH to correctly identify the majority of words within everyday sentences based on a pair of synchronous narrowband temporal speech patterns is consistent with results obtained previously (Healy & Warren, 2002). Because the intelligibility of these narrow speech-modulated bands was...
shown to be at or near zero when presented individually, accurate recognition must be based on the contrast provided by the pair of simultaneous patterns. The current conditions therefore provide a simple test of across-channel integration of temporal speech information.

Listeners with NH appear able to tolerate very small disruptions in across-frequency timing, as no significant change in scores was observed when 12.5 ms of asynchrony was introduced between bands. However, intelligibility fell as asynchrony increased, reaching values below 10% by 100 ms. These data are consistent with those of Healy and Bacon (2000), who presented synchronous and asynchronous temporal patterns at frequency positions different from those used here to listeners with NH. An asynchrony value of 100 ms therefore appears sufficient to virtually eliminate the advantage that the contrasting pair of temporal patterns provides over a single pattern. This result is in general agreement with those of Fu and Galvin (2001), who used speech represented by four broad contiguous bands of speech-modulated noise and introduced various asynchronies between the bands up to a maximum of 240 ms. However, when the speech spectrum was represented by a larger number (16) of speech-modulated carriers, the influence of asynchrony was reduced, presumably due to the increase in spectral detail and decrease in asynchrony between neighboring patterns at any given maximum asynchrony. The influence of asynchrony can be substantially reduced when component frequency bands of speech, rather than speech-modulated carriers, are desynchronized (Greenberg & Arai, 1998; Greenberg, Arai, & Silipo, 1998). This effect is likely due to the presence of aligned temporal contrasts (spectral fine structure) within each individual speech band. The stimulus configuration used in this study has the advantages of (a) reducing the examination of asynchrony to the simplest (2-band) case and (b) isolating comparisons across bands due to the lack of within-band contrasts (an attribute of speech-modulated carriers). Under these conditions, a value of 100 ms appears sufficient to disrupt the across-frequency speech comparison mechanism for listeners with NH.

**Listeners With HI**

On average, the listeners with HI performed more poorly than those with NH when presented with synchronous temporal patterns. This is despite the fact that the stimuli were clearly audible for all listeners (see Figure 1) and that substantial acoustic isolation was provided between the narrow bands (see Figure 2), reducing the potential influence of broadened auditory tuning. This result therefore supports the conclusion that at least some listeners with HI possess a deficit in their ability to effectively integrate temporal speech patterns across different frequencies.

The majority of listeners with HI (6 of 10) produced intelligibility scores far below those of their NH counterparts in the aligned condition. However, a minority of listeners (4 of 10) produced comparable scores. Although this result suggests that the deficit may not be present in this minority of listeners with HI, the results obtained with asynchronous patterns (discussed below) suggest some disruption in across-frequency temporal processing even for those listeners who performed well with synchronous patterns.

On average, intelligibility scores for the listeners with HI tended to fall as a function of asynchrony more steeply than those for listeners with NH. This result indicates that the listeners with HI are less able to handle disruptions in across-frequency timing information and is consistent with the possibility of a less robust across-frequency processing system. For these listeners, a value of 50 ms (rather than the 100 ms required by the listeners with NH) appears sufficient to almost completely disrupt the across-frequency comparison mechanism, regardless of whether the scores in the aligned condition are as good as (Figure 5) or worse than (Figure 4) those of the listeners with NH. In addition to differences in the amount of asynchrony required to obliterate performance, differences exist between listener groups when small amounts of asynchrony are introduced: A steep drop in performance with 12.5 to 25 ms of asynchrony was seen both in the listeners with HI classified as poor-performing (see the lower panel of Figure 4) and as well-performing (see the lower panel of Figure 5). Noteworthy is the especially steep drop apparent for 3 of the 4 listeners who performed as well as the listeners with NH in the aligned condition. Together, these data provide support for a deficit in the across-frequency comparison mechanism in listeners with HI. Although the results for the well-performing listeners were analyzed separately, it is likely that their processing deficit is not fundamentally different from that of the poor-performing listeners. The well-performing listeners appear to possess a deficient cross-frequency processing mechanism that, although it allows them to process synchronous across-frequency temporal information as well as their NH counterparts, is not sufficiently robust to tolerate even slight disruptions in this information. For the members of the poor-performing group, cross-frequency processing is poorer than normal even in the synchronous condition.

However, before concluding that the listeners with HI have a deficit in cross-frequency processing, it is...
important to consider other potential reasons for the observed differences between listener groups. Potential reasons include (a) differences in stimulus presentation level, (b) differences in sensation level, and (c) differences in listener age. Each is examined in the following sections.

**Effect of Stimulus Level**

Speech understanding can be reduced when presentation levels are excessively high (Fletcher & Galt, 1950; French & Steinberg, 1947). Because audibility was a primary concern in the current study, an MCL level-setting procedure was used causing the stimulus levels heard by the listeners with NH (90/90 dB SPL at 750/3000 Hz) to differ slightly from the average levels heard by those with HI (90.3/93.7 dB SPL). However, according to recent data (Studebaker & Sherbecoe, 1995), increasing the speech level by 4 dB should reduce the articulation index (ANSI, 1986) value only very slightly for listeners with NH (0.003 units/dB). Additional data suggest that the influence of level may be slightly increased in listeners with HI (Studebaker, Sherbecoe, McDaniel, & Gwaltney, 1999). However, an increase in signal strength of 2 (on average for the two bands) to 4 dB (maximum) corresponded to a drop in performance of only a couple of percentage points. Thus, differences in stimulus presentation levels cannot account for the substantial differences between listener groups observed in the current study.

**Effect of Sensation Level**

The scores produced by the listeners with NH at low SPLs (matching the average SLs of the listeners with HI) are slightly reduced relative to their high-SPL NH counterparts. Their scores are, however, significantly better than those produced by the listeners with HI, indicating that SL alone cannot account for the differences between listener groups observed in Experiment 1. Further, there were no significant differences observed between the listeners with NH who heard the stimuli at levels matching the worst-case SL of the poor-performing (26 dB on average for the two bands) versus the well-performing (33 dB) listeners with HI. These results are consistent with data showing that performance on psychophysical measures of temporal resolution (e.g., modulation detection) are poorer near threshold, but are asymptotic once stimulus levels reach about 25 dB SL (Bacon & Gleitman, 1992; Bacon & Viemeister, 1985).

**Effect of Listener Age**

It has been reported that listener age, independent of hearing loss, can negatively affect some measures of temporal auditory processing (e.g., Gordon-Salant & Fitzgibbons, 1999; Strouse, Ashmead, Ohde, & Grantham, 1998), although modulation detection may be unaffected by age, at least within the age ranges tested here (cf. Takahashi & Bacon, 1992). Additional information was therefore gathered to examine potential contributions of age differences between the listener groups in the current study.

1. An additional younger listener with bilateral cochlear HI was recruited. He was 38 years old and had pure-tone audiometric thresholds (L ear) of 55 and 70 dB HL at 750 and 3000 Hz. Using the MCL procedure described earlier, he adjusted the stimulus levels to 93 and 100 dB SPL for the low- and high-frequency bands. He heard four dual-band conditions (0, 12.5, 25, 50 ms asynchrony) in random order using procedures identical to those described earlier. His scores were within 2%–5% of the HI group mean in the 0- and 12.5-ms asynchrony conditions but then fell sharply below the mean at 25 ms. His score remained low at 50 ms, once again matching (within 2%) the HI mean. His performance was therefore well below that of the listeners with NH who more closely matched his age but quite similar to the older listeners with HI who matched his hearing status. In addition, he displayed an especially steep drop in performance with asynchrony, characteristic of the (older) listeners with HI.

2. An additional (unreported) group of listeners with NH was run on these same dual-band conditions at 80 dB, rather than 90 dB SPL as reported here. The oldest of these listeners (64 years old) produced scores within 3% (average of absolute differences across the five asynchrony conditions) of the mean for the NH group in Experiment 1. There was no significant difference between the group run at 80 dB and that run at 90 dB in any condition ($t \leq 1.5$, $p > .05$).

3. Another older (59 years old) listener with NH heard a similar set of dual-band misalignment conditions as part of another experiment (Healy & Bacon, 2000). His scores varied unsystematically about the mean (average of absolute differences = 6% across the same five asynchrony conditions), even though the average age of the remaining 9 listeners in his group was 24 years. Further, he did not show the steep drop in performance as a function of asynchrony typically seen in the (older) listeners with HI.

4. The listeners with HI classified as poor-performing and shown in Figure 4 were actually younger (average age = 67 years) than those classified as well-performing and shown in Figure 5 (average age = 75 years).

5. The only listener with HI in the current study to approximate the performance of the listeners with NH (HI 7) matched the age of the oldest participant (79 years).
Mechanisms and Implications

Considerable evidence exists showing that listeners with a hearing loss can identify speech items based on single temporal speech patterns as well as their NH counterparts (Turner et al., 1995, 1999). Thus, any difficulty they experience is likely not related to the ability to extract linguistic information from the temporal waveform itself. The current results involving simultaneous temporal speech patterns at different frequencies suggest that listeners with HI have specific difficulty performing across-frequency integration of speech information. These difficulties do not appear to be due to stimulus level, sensation level, or listener age. A major remaining difference between listener groups involves hearing status: Figure 1 displays the vast differences in audiometric thresholds that exist between the listeners with NH and those with HI. If the presence and degree of hearing loss is the basis of the observed differences in performance and the root of the processing deficit, then it might be expected that a difference in degree of loss would exist between the listeners with HI classified as poor-performing and those classified as well-performing. Indeed, the average audiometric thresholds for the poor-performing listeners were higher than those of the well-performing by 6.7 dB at 750 Hz (57.7 vs. 51.0 dB HL) and by 7.5 dB at 3000 Hz (72.0 vs. 64.5 dB HL). Although these differences are relatively small, they may be sufficient to account for the patterns of results displayed in Figures 4 and 5.

Much like the speech processing abilities examined here, the psychoacoustic across-channel integration abilities of listeners with HI have not been extensively studied. From the work that has been done, it appears that modulation detection interference (MDI), in which the threshold for detecting amplitude modulation is increased when a second modulated sound is introduced, is essentially normal in listeners with HI (Bacon & Opie, 2002; Grose & Hall, 1994). Comodulation masking release (CMR), in which the threshold for a signal in a modulated masker is reduced when a second modulated masker is introduced, can be reduced in listeners with HI, although it is unclear to what extent that simply reflects broadened auditory tuning (Hall, Davis, Haggard, & Pillsbury, 1988; Hall & Grose, 1989). The psychoacoustic task that perhaps most resembles the current speech experiments involves the identification of correlated versus uncorrelated noise envelope patterns at different frequencies (Hall & Grose, 1993). When two bands were presented, it was found that listeners with HI performed slightly worse than those with NH, but the difference between groups did not reach statistical significance in most conditions, perhaps due to large individual differences. The differences between groups were largely absent when more than two simultaneous bands were presented. It is difficult to know the extent to which these psychoacoustic tasks assess the same processing mechanisms examined in the current study. However, if a consistent lack of correspondence emerges, then different processing mechanisms are implicated.

An examination of phonetically balanced (PB) word recognition scores from audiological examinations obtained at various times before experimentation revealed that the listeners with HI classified as well-performing based on the results of the current study had better scores on average (92%) than those classified as poor-performing (74%). This clinical result based on synchronous across-frequency speech information is therefore in accord with the current experimental results. The experimental task was designed to isolate as well as possible the across-frequency integration of speech information and to control for other characteristics of hearing impairment that can contribute to poor performance. It can therefore be argued that those listeners who have lower PB scores and are generally poorer performers in real-world settings may do poorly in part because they possess the proposed across-frequency processing deficit to a greater degree than do their counterparts.

The finding that listeners with HI have particular difficulty with disruptions in across-frequency synchrony may have implications for the perception of speech in reverberant environments. Reverberation times can be longer for low frequencies than for high due to the greater absorption of high-frequency components (cf. Knudsen & Harris, 1978). Low-frequency components can therefore be present at greater delays than high-frequency components. This could potentially create cross-channel asynchrony to the extent that listeners receive the indirect signal. Thus, the inability of listeners with HI to tolerate asynchrony may contribute to their difficulty in reverberant environments (cf. Moore, 1995).

Summary and Conclusions

On a speech-recognition task requiring across-frequency integration of temporal speech information, listeners with normal hearing could accurately identify a majority of component words within everyday speech sentences. Intelligibility remained high when a very slight across-channel asynchrony was introduced, indicating that these listeners possess a tolerance for very slight disruptions in across-frequency timing information. When the asynchrony between temporal patterns was increased to 100 ms, intelligibility was reduced close to the low level of an individual pattern, indicating a virtually complete disruption in across-frequency processing. Listeners with sensorineural hearing loss performed more poorly when presented with synchronous...
temporal speech patterns, indicating a deficit in their ability to perform the required across-frequency integration. In addition, scores from these listeners fell when very small across-channel asynchronies were introduced, indicating an inability to tolerate even slight disruptions in across-frequency timing information. An asynchrony of 50 ms appeared sufficient to almost completely disrupt across-frequency processing in these listeners.

Together, these results suggest that listeners with hearing loss may have a deficit in their ability to compare and effectively process speech patterns at different loci. The results could not be attributed to differences in signal or sensation level or in listener age but instead appear to be associated with the existence and degree of hearing loss: Those having more severe losses exhibited more substantial deficits. The current experiments accounted for audibility by employing a band-limited stimulus and high signal levels. Broadened auditory tuning was accounted for by using a pair of narrowband patterns having substantial acoustic isolation between bands. Because the results could not be attributed to these known disruptions, this across-frequency deficit may exist in addition to known processing difficulties. These results may serve to account for some of the difficulty listeners with HI experience when processing speech and may deepen our understanding of their processing abilities and deficits.

Acknowledgments

This research was supported by a grant from The National Organization for Hearing Research Foundation (EWH) and by NIH/NIDCD grant no. DC01376 (SPB). Manuscript preparation was supported in part by NIH/NIDCD grant no. DC05796 (EWH). The authors thank Sarah Avrech for valuable assistance in performing the experiments.

References


Received August 9, 2001
Accepted March 7, 2002
DOI: 10.1044/1092-4388(2002/101)

Contact author: Eric W. Healy, PhD, Department of Communication Sciences and Disorders, University of South Carolina, Columbia, SC 29208. E-mail: ewh@sc.edu