Aging and Spectro-Temporal Integration of Speech

John H. Grose[1], Heather L. Porter[2], and Emily Buss[1]

Abstract

The purpose of this study was to determine the effects of age on the spectro-temporal integration of speech. The hypothesis was that the integration of speech fragments distributed over frequency, time, and ear of presentation is reduced in older listeners—even for those with good audiometric hearing. Younger, middle-aged, and older listeners (10 per group) with good audiometric hearing participated. They were each tested under seven conditions that encompassed combinations of spectral, temporal, and binaural integration. Sentences were filtered into two bands centered at 500 Hz and 2500 Hz, with criterion bandwidth tailored for each participant. In some conditions, the speech bands were individually square wave interrupted at a rate of 10 Hz. Configurations of uninterrupted, synchronously interrupted, and asynchronously interrupted frequency bands were constructed that constituted speech fragments distributed across frequency, time, and ear of presentation. The overarching finding was that, for most configurations, performance was not differentially affected by listener age. Although speech intelligibility varied across condition, there was no evidence of performance deficits in older listeners in any condition. This study indicates that age, per se, does not necessarily undermine the ability to integrate fragments of speech dispersed across frequency and time.

Introduction

Older listeners—even those with clinically normal audiograms—appear less able to benefit from a fluctuating acoustic background, relative to a steady background, in recognizing speech than younger listeners (e.g., Grose, Mamo, & Hall, 2009). One common interpretation is that older listeners are less able to benefit from the glimpses of speech that occur during the masker minima. There are likely multiple factors that contribute to this ranging from *bottom-up* factors like reduced audibility or increased susceptibility to temporal masking (e.g., Dubno, Horwitz, & Ahlstrom, 2003; Gifford & Bacon, 2005; Gifford, Bacon, & Williams, 2007; Peters, Moore, & Baer, 1998) to *top-down* factors like reduced memory capacity or cognitive slowing (e.g., Pichora-Fuller & Singh, 2006; Wingfield & Tun, 2007). In general, bottom-up factors can be considered to reduce the quality of the encoded speech glimpses, whereas the top-down factors undermine the synthesis of meaningful constructs from the encoded snippets. Differentiating these factors is further complicated by the interplay between

top-down deficits and the cognitive benefits of contextual knowledge accumulated over the lifespan (Pichora-Fuller, 2008; Saija, Akyurek, Andringa, & Baskent, 2014). Whereas many of the studies that have examined glimpsed speech have used masked or interrupted speech in which the speech glimpses remain temporally aligned across the spectrum, in many realistic listening conditions, the fluctuating background comprises multiple sources and therefore might not be comodulated across frequency. Consequently, the available fragments of target speech are not spectrally intact or necessarily

[1]Department of Otolaryngology—Head and Neck Surgery, University of North Carolina at Chapel Hill, NC, USA
[2]Hearing and Speech Department, Children's Hospital Los Angeles, CA, USA

Corresponding Author:
John H. Grose, Department of Otolaryngology—Head and Neck Surgery, University of North Carolina at Chapel Hill, 170 Manning Drive, Chapel Hill, NC 27599-7070, USA.
Email: john_grose@med.unc.edu

similar across ears. This introduces yet another issue—namely, the ability to integrate available speech information across frequency and time, as well as across ears. A comprehensive treatment of such spectro-temporal integration of speech as it relates to aging is lacking. Therefore, the purpose of this study is to assess the integration of sparse speech as a function of listener age, where the speech snippets are variously isolated in both the time and frequency domains, as well as in ear of presentation.

To set the context for this study, it is convenient to review the effects of age on specific types of speech-related integration: (a) spectral integration, (b) temporal integration, and (c) binaural integration. The focus here is on studies that incorporated older listeners with good audiometric hearing. In terms of the integration of discrete speech bands across frequency, Peters et al. (1998) found that older listeners with relatively normal audiograms were less able to understand speech masked by noise containing spectral notches than were younger listeners. However, it is difficult to interpret this finding in terms of spectral integration since the older listeners also performed more poorly in masking noise without spectral notches. Spehar, Tye-Murray, and Sommers (2008) measured intelligibility of speech filtered into two relatively narrow fixed bandwidths in younger and older listeners with normal hearing and found that older listeners understood the two-band speech less well than did younger listeners. Although this appears to be evidence for an age-related reduction in spectral integration, this is not necessarily the case since baseline performance for single-band speech was not measured. When speech is filtered into discrete narrow bands, there is typically a *superadditivity* of performance for multiple simultaneous bands relative to any single band (Grant, Braida, & Renn, 1991; Kasturi, Loizou, Dorman, & Spahr, 2002; Lippmann, 1996; Ronan, Dix, Shah, & Braida, 2004; Warren, Riener, Bashford, & Brubaker, 1995). However, to gauge this spectral integration, it is necessary to reference performance to single-band baselines, which was not done in the Spehar et al. (2008) study. Without these baseline measures, it is difficult to assess whether the observed age-related deficits in two-band speech reflect deficient spectral integration or poorer single-band baseline performance.

In terms of temporal integration of discrete speech segments, two basic approaches have been taken to segmenting the speech in time: interrupting the speech and masking the continuous speech with modulated maskers. The focus here is on studies that used interrupted speech in order to circumvent effects of temporal masking associated with the use of modulated maskers. Importantly, studies that have compared the perception of interrupted or amplitude-modulated speech to the perception of speech masked by modulated noise have shown a strong association between these two approaches (Buss, Whittle, Grose, & Hall, 2009; Jin & Nelson, 2010). At a cursory level, the reported findings on age-related effects for interrupted speech appear mixed. For example, Kidd and Humes (2012) found that older listeners with and without hearing loss performed only slightly worse than younger listeners, and therefore that the pattern of performance did not vary much with age. On the other hand, several studies have reported age-related deficits irrespective of whether the older listeners had relatively normal audiograms or not (e.g., Gordon-Salant & Fitzgibbons, 1993; Shafiro, Sheft, Risley, & Gygi, 2015). Comparisons across studies, however, are complicated by differences in speech material and interruption rate. In terms of interruption rate, it is known that both word and sentence intelligibility typically improve with increasing interruption rate, at least over the range of about 1 to 100 interruptions/sec (e.g., Dirks & Bower, 1970; Huggins, 1964; Jin & Nelson, 2010; Miller & Licklider, 1950; Powers & Wilcox, 1977; Shafiro, Sheft, & Risley, 2011). Whereas this trajectory generally holds across the age span, differences between younger and older listeners can emerge within particular rate spans. For example, Shafiro et al. (2015) found that younger and older listeners with relatively normal audiograms performed similarly at a very low interruption rate (0.5 Hz) but diverged as rates were increased up to 8 Hz. In a similar vein, Saija et al. (2014) found that younger and older listeners with relatively normal audiograms exhibited equivalent performance at very low rates (0.625–1.25 Hz) but differed at higher rates (2.5–5.0 Hz). The two age groups in this study then converged again at a 10-Hz interruption rate where performance approached ceiling. In summary, the effects of age on the ability to integrate discrete segments of speech over time appear to depend in part on the rate at which the speech is interrupted.

Turning now to binaural integration, favorable glimpses of speech are often not identical across the two ears at any one time in many competitive listening environments, so dichotic integration of speech is key to maximizing benefit (Brungart & Iyer, 2012). In terms of dichotic effects in spectral integration, the Spehar et al. (2008) study also compared intelligibility of pairs of speech bands presented either to the same ear or to opposite ears. They found that performance in both younger and older listeners with relatively normal audiometric hearing declined for dichotic presentation relative to monaural presentation. In terms of dichotic effects in binaural temporal integration, Stewart, Ethan, and Wingfield (2008) measured perception of speech that alternated across ears as a function of alternation rate (i.e., the speech in each ear was asynchronously interrupted). They found that older listeners with relatively normal hearing exhibited the same pattern of

performance as younger listeners in that performance was nonmonotonic with minimum intelligibility occurring at an alternation rate of about 3 to 4 Hz. Although they interpreted this result as demonstrating that temporal integration across ears—or fusion in their terms—was itself intact in the senescent auditory system, one feature of their data was that the older listeners performed more poorly overall—including in the baseline diotic condition. A stronger test of the effects of asynchronous gating across ears would be to ensure that baseline performance was independent of age.

In summary, age-related effects in the integration of speech across frequency, time, and ear of presentation have been separately examined, although questions remain within each of these independent dimensions. However, realistic multisource fluctuating backgrounds result in glimpses of target speech that can vary concurrently across each of these dimensions. The purpose of this study, therefore, was to determine the effects of age on *spectro-temporal* integration of speech within and across ears. It tests the hypothesis that spectro-temporal integration of speech is reduced in older listeners, even in those with audiometrically normal hearing.

Materials and Methods

Participants

The participants comprised three age groups, with 10 participants per group: younger (20–28 years; mean = 23 years), mid-age (44–55 years; mean = 49 years), and older (67–81 years; mean = 71 years). All had audiometric thresholds within normal limits ($\leqslant$20 dB HL) at the octave frequencies 250 to 4000 Hz with the exception of two participants in the older group; one older listener had a threshold of 25 dB HL at 250 Hz and the other had a threshold of 25 dB HL at 4000 Hz (Figure 1).
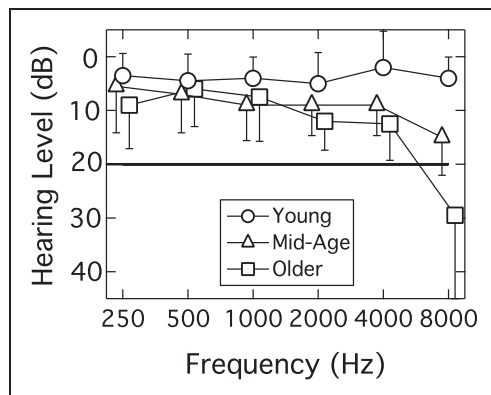

Figure 1. Mean audiograms in the test ear for the three age groups. Symbols are offset for clarity. Error bars are one standard deviation.

All participants provided informed consent and were reimbursed for their participation. The study was approved by the Institutional Review Board of the University of North Carolina at Chapel Hill.

Stimuli

The speech material consisted of the revised Harvard sentences compiled by the Institute of Electrical and Electronic Engineers (IEEE, 1969) and spoken by a male speaker. This corpus consists of 72 lists of 10 sentences per list, where each sentence contains five key words. The corpus was digitally stored as a library of *wav* files using a sampling rate (SR) of 12207 Hz. Depending on the condition, the target speech was filtered into one or two bands as described in further detail in the Procedure section. The lower band was centered at 500 Hz, and the higher band was centered at 2500 Hz. These are the same center frequencies that have been used in similar tests of spectral integration of speech in our lab using the Bamford-Kowal-Bench sentence corpus (Hall, Buss, & Grose, 2008; Mlot, Buss, & Hall, 2010). In some conditions, the filtered speech bands were further square-wave amplitude modulated at a rate of 10 Hz. This rate was selected for several reasons. First, Jin and Nelson (2010) have shown that, for younger normal-hearing listeners, interruption rates of unfiltered IEEE sentences in this range result in relatively high intelligibility but still below ceiling performance (8 Hz: ~82%; 16 Hz: ~90%). Second, Gordon-Salant and Fitzgibbons (1993) have shown that, for full spectrum sentence material presented at a similar interruption rate (12.5 Hz), younger and older listeners with normal hearing exhibit similar performance. In a similar vein, Saija et al. (2014) have shown that younger and older listeners with relatively normal audiometric hearing have similar (and high) intelligibility of unfiltered sentence material interrupted at a rate of 10 Hz in quiet. As with the Jin and Nelson (2010) and Saija et al. (2014) studies, the square-wave modulator applied here was shaped to have slightly tapered transitions; that is, the amplitude transitions between high and low levels consisted of 4-ms ramps comprising half cycles of a raised cosine. The necessary digital signal processing (bandpass filtering, temporal modulation, amplitude scaling, etc.) was applied in real time to each selected sentence prior to presentation using custom MATLAB code (Mathworks, Natick, MA) in conjunction with a digital signal processing platform (RZ6; Tucker-Davis Technologies, Alachua, FL); this platform uses 24-bit sigma-delta digital-to-analog converters that provide an output bandwidth of 0–0.44SR Hz. The speech was presented through Sennheiser HD580 headphones (Wedemark, Germany) at a level that, prior to any processing, was 70 dB SPL as calibrated with a sound level

meter and flat-plate coupler (Model 824 and Model AEC101; Larson-Davis, Provo, UT). For monaural conditions, the default ear for testing was the right ear. There were three exceptions to this in the older group where the audiometric profile for the left ear was marginally better than the right and so this ear was used instead as the test ear.

Procedure

The participant sat in a double-walled, sound-attenuating booth and listened to the speech over headphones. The participant was instructed to repeat back aloud as much of each presented sentence as was perceived, even if it did not make semantic or grammatical sense. Outside the booth, the experimenter monitored the participant's response through a microphone feedback system. Concurrent with the acoustic presentation of a target sentence to the participant, a written transcription of the sentence appeared on the monitor in front of the experimenter with each key word highlighted within a position-sensitive rectangle. The experimenter mouse-clicked on each key word that was either omitted or repeated incorrectly, and the computer thereby registered the accuracy of the response at the word level. No participant received the same sentence more than once across all conditions, and the starting sentence within the IEEE corpus varied across participants.

Phase 1 of the experiment consisted of adaptively varying the bandwidth of both the low and high bands in isolation in order to achieve a relatively low-percent correct score for that band presented alone. A somewhat ad hoc stepping rule was derived that empirically converged on the target performance region. By this rule, a response to the sentence-level trial was recorded as incorrect if none, or only one, of the key words in the sentence were correctly identified. If two or more key words were correctly identified, the trial response was scored as correct. Following one correct response, the filter bandwidth was decreased by a factor of 1.21. Following two incorrect responses in a row, the filter bandwidth was increased by the same factor. A transition from narrowing to broadening of the filter bandwidth, or vice versa, constituted a reversal, and the track was terminated after eight reversals. The criterion bandwidth estimate for each track was taken as the geometric mean of the filter bandwidths at the final six reversal points of the track. To ensure that the adaptively varying bandwidths for each of the two filters could not converge on values that would result either in overlap of the low and high filters or on filters that were too narrow, boundary values were placed on the adaptive track that flagged whether the adaptively varying bandwidth ever fell below or exceeded values, as a proportion of filter center frequency, of either

0.01 (floor) or 1.5 (ceiling). Three bandwidth estimates were collected for each of the two frequency regions, and the final bandwidth for that region was taken as the mean of the three estimates.

Phase 2 of the experiment was initiated for each participant once the two criterion bandwidths had been individually measured for that participant. Phase 2 measured percent-correct word recognition across a sequence of 25 sentences for each of seven conditions that incorporated the individually tailored pair of speech bands. The conditions constituted combinations of spectral, temporal, and binaural integration and are shown schematically in Figure 2. The first two conditions were monaural baseline conditions comprising: (a) the low band alone (*Low*) and (b) the high band alone (*High*). The next two conditions examined strictly spectral integration where the two speech bands were presented together continuously either to the same ear or to opposite ears. These conditions were as follows: (c) low-plus-high bands monaural; that is, both bands in the test ear (*Mon L+H*); and (d) low-plus-high bands dichotic; that is, high band in the test ear, low band in the contralateral ear (*Dich L+H*). These two conditions were also accompanied by a control condition (*Low Contr*; not shown in Figure 2), wherein the low band alone was presented to the ear contralateral to the test ear since, in the *Dich L+H* configuration, the low band was presented to the ear in which its criterion bandwidth had *not* been measured. The final three conditions incorporated configurations that allowed for spectro-temporal integration both within and across ears. These conditions were as follows: (e) the pair of monaural speech bands synchronously interrupted at a rate of 10 Hz (*Mon Sync*); (f) the pair of speech bands synchronously interrupted at 10 Hz within an ear but asynchronously across ears such that the pair of bands alternated across ears (*Dich Alt*); and (g) the pair of speech bands asynchronously interrupted at 10 Hz within an ear *and* across ears, thereby alternating in frequency within an ear and across ears such that the two ears never received the same frequency band simultaneously (*Dich Async*). Here, information from each speech band was continuously available but never synchronously within an ear or across ears. Subsequent to completion of the main study, a supplementary condition was tested (see Discussion section) where the two speech bands, presented monaurally, were interrupted asynchronously (*Mon Async*).

Results

Phase 1

The results of the adaptive bandwidth phase of the experiment are summarized in Figure 3 where criterion bandwidth is expressed as a proportion of center
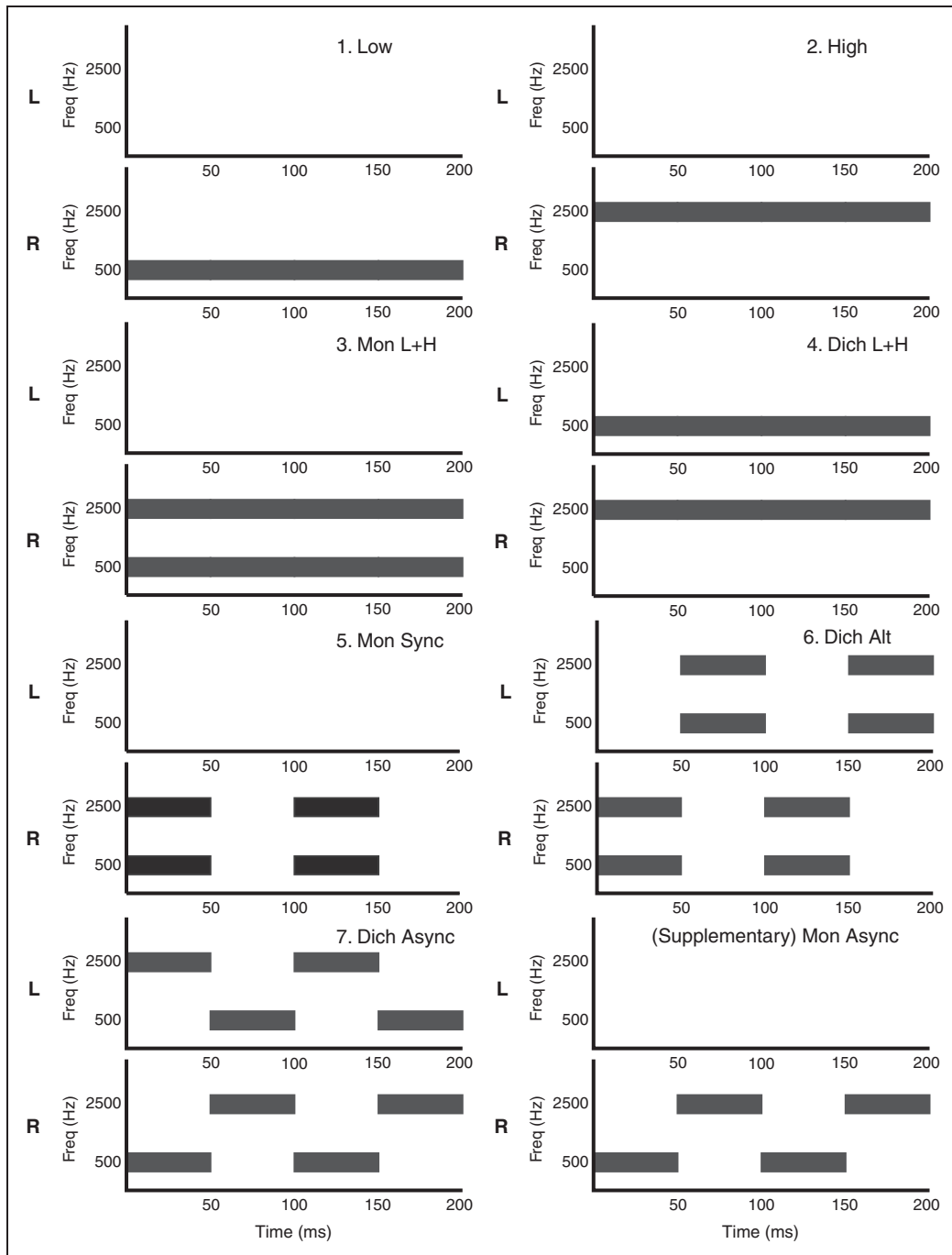
Figure 2. Stimulus configuration schematics. Each panel displays a time-frequency schematic for left and right ears for the condition noted in the upper right corner.

frequency. The mean criterion bandwidths are plotted for the low-frequency band centered at 500 Hz and the high-frequency band centered at 2500 Hz, with age group as the parameter. These data were log-transformed and submitted to a repeated measures analysis of variance (RMANOVA) that indicated a significant effect of frequency band region, ($F(1, 27) = 187.370$; $p < .01$), but no effect of age group, ($F(2, 27) = 1.60$; $p = .22$), and no interaction between these two factors, ($F(2, 27) = 1.244$; $p = .30$). Although the criterion bandwidth was proportionally higher in the 500-Hz region than in the 2500-Hz region, the absolute bandwidth was substantially larger in the 2500-Hz region ($\sim$1545 Hz) than in the 500-Hz region ($\sim$309 Hz).

Phase 2

The results of the speech intelligibility phase of the experiment are summarized in Figure 4; all percent-correct intelligibility scores have been transformed into rationalized arcsine units (RAUs) (Studebaker, 1985), and all statistics were performed on these transformed scores. In Figure 4, proceeding from left to right across the panel, results are shown for the two baseline conditions (*Low, High*), followed by the monaural and dichotic spectral integration conditions (*Mon L + H, Dich L + H*), and then the remaining spectro-temporal integration conditions (*Mon Sync, Dich Alt, Dich Async*). The data, parameterized by age group, are shown as
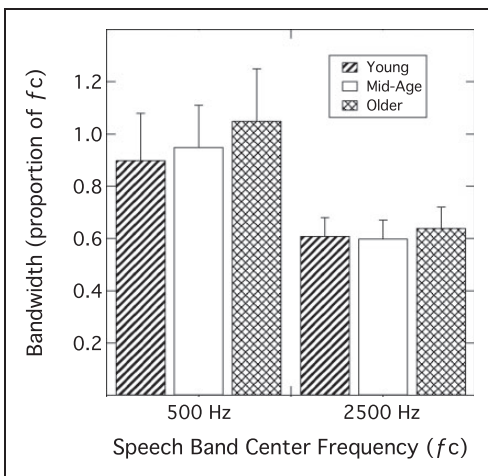

Figure 3. Mean criterion bandwidths (proportion of center frequency) at the low (500 Hz) and high (2500 Hz) frequency regions for the three age groups. Error bars are one standard deviation.

box-and-whisker plots where each rectangle denotes the 25% to 75% range, the horizontal line represents the median, and the capped bars denote the 10% to 90% range. Dealing first with the baseline *Low* and *High* conditions, performance was generally slightly better for the high-frequency band alone than the low-frequency band alone. This was confirmed with a RMANOVA that showed a significant effect of band center frequency, ($F(1, 27) = 7.952$; $p < .01$), but no effect of age group, ($F(2, 27) = 0.081$; $p = .92$), or interaction between these two factors, ($F(2, 27) = 1.586$; $p = .22$). The mean RAU score for the low band alone was about 31, whereas that for the high band alone was about 37. When the two bands were presented together continuously, superadditivity occurred in that performance jumped to high levels irrespective of whether the bands were presented to the same ear or different ears. A RMANOVA on these *Mon L + H* and *Dich L + H* results indicated no main effect of monaural versus dichotic presentation mode, ($F(1, 27) = 2.341$; $p = .14$), and no main effect of age group, ($F(2, 27) = 0.129$; $p = .88$). However, the interaction between these two factors was significant, ($F(2, 27) = 6.086$; $p < .01$). Simple main effects testing showed that, whereas the presentation mode had no effect on the younger and older age groups, the mid-age group performed more poorly in the dichotic condition than in the monaural condition ($p < .01$). Recall that the criterion bandwidths were measured in one ear only per listener but the *Dich L + H* condition entailed presentation of the low band to the ear contralateral to the test ear. To ensure that recognition of this low band of speech was not dependent upon ear of presentation, the control condition *Low Contr* was run wherein the low-band speech alone was presented to the nontest ear. A RMANOVA showed that recognition of this low-band
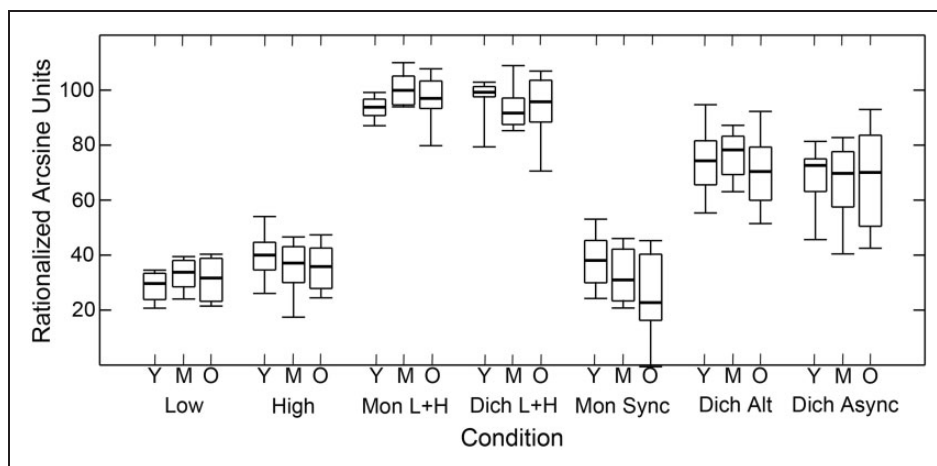

Figure 4. Word recognition performance (RAU scores) for each condition and age group (Y: younger; M: mid-age; O: older). Each rectangle = 25% to 75%, horizontal line = median; capped bars = 10% to 90%.

speech was not affected by the ear of presentation, $(F(1, 27) = 1.019; p = .32)$, or age group, $(F(2, 27) = 0.177; p = .84)$, and that the interaction of these factors was also not significant, $(F(2, 27) = 2.516; p = .10)$. Thus, the poorer performance of the mid-age group in the dichotic condition (*Dich L + H*) relative to the monaural condition (*Mon L + H*) was not due to poorer recognition of the low-band speech in the nontest ear.

Turning now to the spectro-temporal integration conditions, the introduction of 10-Hz temporal interruptions into the two-band filtered speech was generally disruptive to performance. Relative to performance for the two-band speech presented continuously to a single ear, the introduction of synchronous interruptions to the monaural speech (*Mon Sync*) caused performance to drop to single-band baseline levels. This was confirmed by comparing the performance for each listener in the *Mon Sync* condition to that listener's average performance in the baseline *Low* and *High* conditions. The RMANOVA indicated no difference between these two performance measures, $F(1, 27) = 1.139; p = .30$, no effect of age group, $F(2, 27) = 2.522; p = .10$, and no interaction between these factors, $F(2, 27) = 2.865; p = .07$. When the synchronously interrupted speech bands were presented alternately to both ears (*Dich Alt*), there was some recovery of performance. However, performance recovery did not return to the peak levels associated with the continuous two-band speech (*Mon L + H* and *Dich L + H*). To verify this, the performance for each listener in the *Dich Alt* condition was compared with that listener's performance in the *Mon Sync* condition and to his or her average performance in the *Mon L + H* and *Dich L + H* conditions. The RMANOVA indicated that there was a significant effect of condition, $F(2, 54) = 402.432; p < .01$, but no effect of age group, $F(2, 27) = 2.000; p = .16$, and no interaction between these factors, $F(4, 54) = 1.464; p = .23$. Simple contrasts for the condition effect showed that condition *Dich Alt* differed significantly from both the *Mon Sync* condition and the average of the *Mon L + H* and *Dich L + H* conditions, being intermediate between the two. The final spectro-temporal condition, *Dich Async*, contained the same aggregate speech-band information as the *Dich Alt* condition, although presented in a different pattern, but the range of performance was the same. To confirm this, a RMANOVA was performed on the two conditions stratified by age group. The analysis indicated no effect of condition, $F(1, 27) = 3.916; p = .06$, no effect of age group, $F(2, 27) = 0.209; p = .81$, and no interaction between these factors, $F(2, 27) = 0.470; p = .63$.

Discussion

The primary result of this study was the general absence of an age effect in spectral and spectro-temporal integration for the conditions tested here. The one deviation from this general result was in the spectral integration conditions (*Mon L + H* and *Dich L + H*) where the mid-age group performed more poorly in the dichotic condition than in the monaural condition, unlike the younger and older groups who did not show a difference between these two conditions. The dissimilar spectral integration performance of the mid-age group eludes a straightforward explanation. Although their pattern of results is in line with the findings of Spehar et al. (2008), who found poorer performance in filtered sentence perception in a dichotic condition relative to a monaural condition in their normal-hearing participants, this concurrence is undermined by the fact that the participants in the Spehar et al. (2008) study were younger (mean age = 21 years) and older (mean age = 73 years) listeners—there were no middle-aged participants. Therefore, their pattern of results actually contrasts with the present study where no effect of condition was observed for normal-hearing younger and older listeners. An interaction between age and monaural or dichotic presentation of pairs of speech bands has been reported for a filtered word test (Palva & Jokinen, 1975), but the pattern of that interaction does not match that of the present study. In the Palva and Jokinen (1975) study, listeners over the age of about 60 years tended to perform better in the dichotic condition than in a monaural condition, unlike younger adults who generally gave similar performance for both monaural and dichotic conditions. However, the data pattern for their older listeners was driven by an ear asymmetry wherein the monaural two-band speech was more intelligible in one ear than the other, while performance in the dichotic condition generally tracked the better monaural condition. Thus, performance in the dichotic condition was better than that of the *poorer* monaural condition in the older listeners. Note that the Palva and Jokinen (1975) results should be interpreted cautiously in terms of age-related effects because participant inclusion criteria tolerated some degree of hearing loss; this may have confounded the factors of age and hearing loss. The interaction between age and condition in the present study, in which only the mid-age group showed a monaural/dichotic difference in spectral integration, does not lend itself to a straightforward interpretation and remains a somewhat anomalous finding. Future studies with larger test populations may resolve this issue. It is important to bear in mind, however, that performance in these conditions of spectral integration was generally very high across the age groups, raising the possibility that encroachment upon ceiling performance might have affected comparisons across age groups.

The introduction of a 10-Hz square-wave modulator to synchronously interrupt the pair of filtered speech bands (condition *Mon Sync*) was detrimental to

performance, but to the same extent across listeners, thereby resulting in no age-related differences. This lack of an age effect for interrupted speech contrasts with some reports in the literature (e.g., Gordon-Salant & Fitzgibbons, 1993; Shafiro et al., 2015) but two important stimulus characteristics must be considered when making such comparisons: (a) the rate of interruption and (b) the speech material used. In terms of the rate of interruption, rates in the region of 10 Hz can have little effect on the intelligibility of otherwise intact sentence material in both younger and older normal-hearing listeners, whereas lower interruption rates can result in marked deficits (Saija et al., 2014; Shafiro, Sheft, & Risley, 2016). Even for rates in the 10-Hz region, however, performance depends on the speech material. The high intelligibility reported by Shafiro et al. (2016) for 8- and 16-Hz interruptions was for the Hearing in Noise Test [HINT] sentences while that reported by Saija et al. (2014) for 10-Hz interruptions was for everyday Dutch sentences. In contrast, Gordon-Salant and Fitzgibbons (1993) used the low-predictability Revised Speech Perception in Noise [LP R-SPIN] sentences and found a significant reduction in performance for both younger and older normal-hearing listeners when the sentences were interrupted at a rate of 12.5 Hz. A reasonable conclusion, therefore, is that performance for both younger and older normal-hearing listeners in perceiving speech interrupted at $\sim$10 Hz depends in part on the redundancy of the speech material. The present results support this possibility in that the sentence material was already reduced in redundancy by virtue of the band-pass filtering, and therefore the 10-Hz interruption rate markedly reduced its intelligibility. Irrespective of the level of performance with the 10-Hz interruptions, the present study found no age-related differences in performance. Whereas Saija et al. (2014) also found no difference between their normal-hearing younger and older listeners at 10 Hz, this could have reflected a ceiling effect since both groups performed near 100% correct for this rate. At lower rates of 2.5 Hz and 5 Hz, the older group performed significantly more poorly than the younger group, with performance converging again for rates $\leqslant \sim$1 Hz. The data of Gordon-Salant and Fitzgibbons (1993) also suggest an absence of an age effect for younger and older listeners with good audiometric hearing for an interruption rate of 12.5 Hz (cf. their Figure 4), although they reported a general age-related deficit across the range of interruptions 12.5 to 100 Hz. In contrast, Shafiro and coworkers (2015, 2016) did find an age effect for interrupted speech at rates of 8 Hz and 16 Hz in younger and older normal-hearing listeners, and this effect extended to lower rates until the groups converged again at a rate of 0.5 Hz. In summary, then, the cumulative evidence indicates that age-related deficits in the perception of interrupted speech do exist for some

combinations of interruption rate and speech material; however, such age effects do not extend to all combinations—such as those used in the present study.

The monaural presentation of 10-Hz synchronously interrupted speech (condition *Mon Sync*) reduced performance to that of the single-band baseline levels (conditions *Low* and *High*). Buss, Hall, and Grose (2004) have shown that, for amplitude-modulated vowel-consonant-vowel stimuli, some listeners perceive synchronously modulated speech better than speech that is filtered into contiguous bands, with alternate bands modulated out-of-phase. In other words, for these listeners, synchronously interrupted speech is more intelligible than asynchronously interrupted speech. This raises the possibility that performance for some listeners in the present paradigm might have been even poorer than that associated with the *Mon Sync* condition if the two monaural speech bands had been interrupted asynchronously. This would amount to performance in the asynchronous condition dropping below that of the single-band levels. Consideration of this possibility arose after collection of the present dataset had been completed, so a supplementary dataset was collected at a later time that compared *Mon Sync* performance to a new condition where the 10-Hz interrupted speech bands were presented in an alternating pattern to the test ear (condition *Mon Async*). The participant inclusion criteria for this supplementary dataset were the same as for the main experiment except that only younger and older listeners were recruited (10 younger and 8 older participants; of these, only 6 older listeners had participated in the main experiment). All participants underwent Phase 1 testing to identify criterion speech bandwidths, and then were tested in Phase 2 on conditions *Mon Sync* and *Mon Async*. For the *Mon Sync* condition, the median RAU score (and interquartile range) was 32.6 (13.6) for the younger listeners and 28.4 (11.4) for the older listeners. For the *Mon Async* condition, it was 24.9 (15.8) and 25.0 (19.3) for the younger and older listeners, respectively. The RMANOVA indicated no effect of condition, ($F(1, 16) = 2.696$; $p = .12$), no effect of age group, ($F(1, 16) = 0.075$; $p = .79$), and no interaction between these factors, ($F(1, 16) = 0.641$; $p = .44$). Thus, there is no indication that the asynchronously interrupted speech was more difficult than the synchronously interrupted speech for the conditions tested here.

When interrupted speech is alternated across ears, intelligibility typically improves relative to the interrupted speech restricted to a single ear—although the degree of improvement depends on the rate of interruption (e.g., Wingfield, 1977). However, Stewart et al. (2008) found that in both younger and older listeners the recovery of intelligibility did not return to that of the uninterrupted level, at least across the range of

interruption rates 1 to 16 Hz. This was also the case in the present dataset, where 10-Hz alternation of the comodulated speech bands across ears (*Dich Alt* condition) improved performance in all age groups to a level intermediate between the single-ear interrupted speech (*Mon Sync* condition) and the continuous speech (*Mon L + H* and *Dich L + H* conditions). This improvement relies on the integration of speech information that completely switches between ears. Relative to this level of integration, it was of interest to determine the effect of maintaining a continuous feed of speech information to each ear, but where the content of that information alternated across frequency in each ear and, concomitantly, across ears as in the *Dich Async* condition. Successful integration of asynchronous speech glimpses presented dichotically has been demonstrated using a masking paradigm (Ozmeral, Buss, & Hall, 2012). In that study, speech was filtered into multiple contiguous bands, with alternate bands presented to opposite ears, and this continuous speech was then masked by a modulated masker where the modulator phase could be inverted across ears to render the speech glimpses asynchronous across ears. In the present study, the *Dich Async* configuration had the two speech bands interrupted asynchronously within an ear and asynchronously across ears such that the two ears did not receive the same frequency information simultaneously. The results showed the intelligibility of the *Dich Async* condition to be equivalent to the *Dich Alt* condition where the comodulated speech alternated across ears. This demonstrates that binaural integration of speech segments does not depend on the full complement of speech information being available within a single ear at any one time, at least for the conditions tested here. However, integration of the speech fragments distributed across time and frequency as well as across ear does not return performance to the same peak level associated with continuous, uninterrupted speech.

The general lack of an age effect in this study of interrupted speech is particularly noteworthy given the age-related deficits often reported in the perception of speech that is masked by a modulated masker. Numerous studies have shown that when target speech is segmented into temporal glimpses by the presence of a modulated masker, older listeners—even those with relatively normal audiometric hearing—generally perform worse than younger listeners (e.g., Dubno et al., 2003; George et al., 2007; Gifford et al., 2007; Grose et al., 2009). This dichotomy in the intelligibility of speech fragmented into glimpses by means of interruption versus temporal masking—even for similar rates—suggests that the age-related deficits associated with speech in modulated maskers are not necessarily due to a failure to integrate available speech glimpses but, rather, to a reduction in the fidelity of the target speech extracted during the masker minima. This has positive ramifications for

efforts directed at improving speech-in-noise intelligibility by means of algorithms that identify and isolate time-frequency windows with momentarily favorable signal-to-noise ratios (e.g., Brungart, Chang, Simpson, & Wang, 2006; Healy, Yoho, Wang, & Wang, 2013). The successful application of such binary mask approaches depends ultimately on the ability of the listener to integrate speech fragments from across frequency, time, and space (i.e., ear). The results of this study suggest that age, per se, does not necessarily undermine this potential. One caveat to bear in mind, however, is that this study did not assess listening effort associated with performance, and there is strong evidence that listening effort changes with age (e.g., Heinrich & Schneider, 2011; Pichora-Fuller, 2003)

Summary and Conclusion

The purpose of this study was to determine whether spectro-temporal integration of speech declines with age. The experiment tested the hypothesis that spectro-temporal integration of speech is reduced in older listeners even for those with good audiometric hearing. The experimental conditions comprised speech that was band-pass filtered into two bands where each band could be independently square wave interrupted at a rate of 10 Hz. Configurations of uninterrupted, synchronously interrupted, and asynchronously interrupted speech were constructed that exemplified speech fragments distributed across frequency, time, and ear of presentation. As such, these configurations were designed to be simplified analogs of the types of fragmented target speech that might be encountered in real-world, multisource fluctuating backgrounds. The over-arching finding of the study was that performance was not affected by listener age; thus, the null hypothesis was not rejected. This general finding can be viewed as encouraging in the context of the development of speech-in-noise processing algorithms that are designed to extract time-frequency windows in which the signal-to-noise ratio is momentarily high. The success of these algorithms depends ultimately on the ability to integrate the glimpses into meaningful speech constructs. The results of this study suggest that age, per se, does not necessarily undermine this potential.

Declaration of Conflicting Interests

The authors declared the following potential conflicts of interest with respect to the research, authorship, and/or publication

of this article: The authors declare that there are no conflicts of interest.

References

Brungart, D. S., Chang, P. S., Simpson, B. D., & Wang, D. (2006). Isolating the energetic component of speech-on-speech masking with ideal time-frequency segregation. *Journal of the Acoustical Society of America*, *120*, 4007–4018.

Brungart, D. S., & Iyer, N. (2012). Better-ear glimpsing efficiency with symmetrically-placed interfering talkers. *Journal of the Acoustical Society of America*, *132*, 2545–2556.

Buss, E., Hall, J. W. 3rd, & Grose, J. H. (2004). Spectral integration of synchronous and asynchronous cues to consonant identification. *Journal of the Acoustical Society of America*, *115*, 2278–2285.

Buss, E., Whittle, L. N., Grose, J. H., & Hall, J. W. 3rd (2009). Masking release for words in amplitude-modulated noise as a function of modulation rate and task. *Journal of the Acoustical Society of America*, *126*, 269–280.

Dirks, D., & Bower, D. (1970). Effects of forward and backward masking on speech intelligibility. *Journal of the Acoustical Society of America*, *47*, 1003–1008.

Dubno, J. R., Horwitz, A. R., & Ahlstrom, J. B. (2003). Recovery from prior stimulation: Masking of speech by interrupted noise for younger and older adults with normal hearing. *Journal of the Acoustical Society of America*, *113*, 2084–2094.

George, E. L., Zekveld, A. A., Kramer, S. E., Goverts, S. T., Festen, J. M., Houtgast, T. (2007). Auditory and nonauditory factors affecting speech reception in noise by older listeners. *Journal of the Acoustical Society of America*, *121*, 2362–2375.

Gifford, R. H., & Bacon, S. P. (2005). Psychophysical estimates of nonlinear cochlear processing in younger and older listeners. *Journal of the Acoustical Society of America*, *118*, 3823–3833.

Gifford, R. H., Bacon, S. P., & Williams, E. J. (2007). An examination of speech recognition in a modulated background and of forward masking in younger and older listeners. *Journal of Speech Language Hearing Research*, *50*, 857–864.

Gordon-Salant, S., & Fitzgibbons, P. J. (1993). Temporal factors and speech recognition performance in young and elderly listeners. *Journal of Speech Language Hearing Research*, *36*, 1276–1285.

Grant, K. W., Braida, L. D., & Renn, R. J. (1991). Single band amplitude envelope cues as an aid to speech reading. *Quarterly Journal of Experimental Psychology. A, Human Experimental Psychology*, *43*, , 621–645.

Grose, J. H., Mamo, S. K., & Hall, J. W. 3rd (2009). Age effects in temporal envelope processing: Speech unmasking and auditory steady state responses. *Ear and Hearing*, *30*, 568–575.

Hall, J. W. 3rd, Buss, E., & Grose, J. H. (2008). Spectral integration of speech bands in normal-hearing and hearing-impaired listeners. *Journal of the Acoustical Society of America*, *124*, 1105–1115.

Healy, E. W., Yoho, S. E., Wang, Y., & Wang, D. (2013). An algorithm to improve speech recognition in noise for hearing-impaired listeners. *Journal of the Acoustical Society of America*, *134*, 3029–3038.

Heinrich, A., & Schneider, B. A. (2011). Elucidating the effects of ageing on remembering perceptually distorted word pairs. *Quarterly Journal of Experimental Psychology*, *64*, 186–205.

Huggins, A. W. F. (1964). Distortion of temporal pattern of speech: Interruption and alternation. *Journal of the Acoustical Society of America*, *36*, 1055–1064.

Institute of Electrical and Electronic Engineers. (1969). IEEE recommended practice for speech quality measurements [subjective measurements subcommittee]. *Institute of Electrical and Electronic Engineers Transactions on Audio and Electroacoustics*, *AU-17*, 225–246.

Jin, S. H., & Nelson, P. B. (2010). Interrupted speech perception: The effects of hearing sensitivity and frequency resolution. *Journal of the Acoustical Society of America*, *128*, 881–889.

Kasturi, K., Loizou, P. C., Dorman, M., & Spahr, T. (2002). The intelligibility of speech with "holes" in the spectrum. *Journal of the Acoustical Society of America*, *112*, 1102–1111.

Kidd, G. R., & Humes, L. E. (2012). Effects of age and hearing loss on the recognition of interrupted words in isolation and in sentences. *Journal of the Acoustical Society of America*, *131*, 1434–1448.

Lippmann, R. (1996). Accurate consonant perception without mid-frequency speech energy. *Institute of Electrical and Electronic Engineers Transactions on Speech and Audio Processing*, *4*, 66–69.

Miller, G. A., & Licklider, J. C. R. (1950). The intelligibility of interrupted speech. *Journal of the Acoustical Society of America*, *22*, 167–173.

Mlot, S., Buss, E., & Hall, J. W. 3rd (2010). Spectral integration and bandwidth effects on speech recognition in school-aged children and adults. *Ear and Hearing*, *31*, 56–62.

Ozmeral, E. J., Buss, E., & Hall, J. W. (2012). Asynchronous glimpsing of speech: Spread of masking and task set-size. *Journal of the Acoustical Society of America*, *132*, 1152–1164.

Palva, A., & Jokinen, K. (1975). The role of the binaural test in filtered speech audiometry. *Acta Oto-Laryngologica*, *79*, 310–314.

Peters, R. W., Moore, B. C., & Baer, T. (1998). Speech reception thresholds in noise with and without spectral and temporal dips for hearing-impaired and normally hearing people. *Journal of the Acoustical Society of America*, *103*, 577–587.

Pichora-Fuller, M. K. (2003). Processing speed and timing in aging adults: Psychoacoustics, speech perception, and comprehension. *International Journal of Audiology*, *42*, S59–S67.

Pichora-Fuller, M. K. (2008). Use of supportive context by younger and older adult listeners: Balancing bottom-up and top-down information processing. *International Journal of Audiology*, *47*, S72–S82.

Pichora-Fuller, M. K., & Singh, G. (2006). Effects of age on auditory and cognitive processing: Implications for hearing aid fitting and audiologic rehabilitation. *Trends in Amplification*, *10*, 29–59.

Powers, G. L., & Wilcox, J. C. (1977). Intelligibility of temporally interrupted speech with and without intervening noise. *Journal of the Acoustical Society of America*, *61*, 195–199.

Ronan, D., Dix, A. K., Shah, P., & Braida, L. D. (2004). Integration across frequency bands for consonant identification. *Journal of the Acoustical Society of America*, *116*, 1749–1762.

Saija, J. D., Akyurek, E. G., Andringa, T. C., & Baskent, D. (2014). Perceptual restoration of degraded speech is preserved with advancing age. *Journal of the Association for Research in Otolaryngology*, *15*, 139–148.

Shafiro, V., Sheft, S., & Risley, R. (2011). Perception of interrupted speech: Effects of dual-rate gating on the intelligibility of words and sentences. *Journal of the Acoustical Society of America*, *130*, 2076–2087.

Shafiro, V., Sheft, S., & Risley, R. (2016). The intelligibility of interrupted and temporally altered speech: Effects of context, age, and hearing loss. *Journal of the Acoustical Society of America*, *139*, 455–465.

Shafiro, V., Sheft, S., Risley, R., & Gygi, B. (2015). Effects of age and hearing loss on the intelligibility of interrupted speech. *Journal of the Acoustical Society of America*, *137*, 745–756.

Spehar, B. P., Tye-Murray, N., & Sommers, M. S. (2008). Intra-versus intermodal integration in young and older adults. *Journal of the Acoustical Society of America*, *123*, 2858–2866.

Stewart, R., Ethan, Y., & Wingfield, A. (2008). Perception of alternated speech operates similarly in young and older adults with age-normal hearing. *Perception & Psychophysics*, *70*, 337–345.

Studebaker, G. A. (1985). A "rationalized" arcsine transform. *Journal of Speech Language Hearing Research*, *28*, 455–462.

Warren, R. M., Riener, K. R., Bashford, J. A. Jr., & Brubaker, B. S. (1995). Spectral redundancy: Intelligibility of sentences heard through narrow spectral slits. *Perception & Psychophysics*, *57*, 175–182.

Wingfield, A. (1977). The perception of alternated speech. *Brain and Language*, *4*, 219–230.

Wingfield, A., & Tun, P. A. (2007). Cognitive supports and cognitive constraints on comprehension of spoken language. *Journal of the American Academy of Audiology*, *18*, 548–558.