Research Report

Age-related changes in neural activity associated with concurrent vowel segregation

Joel S. Snyder\textsuperscript{a,}\textsuperscript{*}, Claude Alain\textsuperscript{a,}\textsuperscript{b}

\textsuperscript{a}The Rotman Research Institute, Baycrest Centre for Geriatric Care 3560 Bathurst Street, Toronto, ON, Canada, M6A 2E1
\textsuperscript{b}Department of Psychology, University of Toronto, Canada

Accepted 1 March 2005
Available online 8 April 2005

Abstract

Older adults exhibit degraded speech comprehension in complex sound environments, which may be related to overall age-related declines in low-level sound segregation. This hypothesis was tested by measuring event-related potentials (ERPs) while listeners identified two different vowels presented simultaneously. Older adults were less accurate than young adults at identifying the two vowels, although both groups improved similarly with increasing fundamental frequency differences ($\Delta f_0$) between vowels. Reaction time data showed that older adults took more time to process stimuli, especially those with smaller $\Delta f_0$. A negative ERP wave indexing the automatic registration of $\Delta f_0$ (the object-related negativity) was reduced in older adults. In contrast, young and older adults showed a similar pattern of neural activity indexing attentive processing of $\Delta f_0$. The results suggest that aging affects the ability to automatically segregate speech sounds.

\footnote{Corresponding author. Fax: +1 416 785 2862.
E-mail address: jsnyder@rotman-baycrest.on.ca (J.S. Snyder).}

\textsuperscript{0926-6410/$ - see front matter \textcopyright 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.cogbrainres.2005.03.002}

\section{1. Introduction}

Aging impairs listeners’ ability to identify speech, especially under adverse situations such as in the presence of background noise and acoustic variability \cite{14,15,25,29,30}. Although this age-related change in perception has received considerable attention, there is no consensus about the mechanisms underlying speech perception problems in older individuals. Changes in hearing sensitivity \cite{24}, memory \cite{25}, attention \cite{18}, and speed of processing \cite{28} have all been proposed to account for the difficulties older adults have in processing everyday speech.

Evidence from behavioral studies suggests that impairment in auditory scene analysis may also play an important role in the speech perception problems commonly observed in older adults \cite{3,16,17,22,27}. For example, older adults showed higher thresholds for detecting a mistuned harmonic apart from a complex sound compared to younger adults, even after statistically controlling for audiometric thresholds \cite{3}. This indicates that age-related problems in detecting multiple objects as defined by inharmonicity do not depend solely on peripheral factors and must involve an age-related decline in central auditory functioning. Listeners who find it difficult to detect mistuning may assimilate one or more frequency components coming from secondary sound sources into the target signal. In this context, the mistuned harmonic might be analogous to the presence of a secondary voice related to another fundamental frequency. The inclusion of extraneous frequency components into the target signal could then lead to concomitant errors in perception. Related to this, older adults with hearing loss showed an impaired ability to identify concurrently presented vowel sounds \cite{33}, a task that requires parsing frequencies based on
the fundamental frequency (\(f_0\)) of each vowel. It is unclear, however, whether this impairment arose from peripheral deficits or more central changes because the older participants had hearing loss. Furthermore, performance on behavioral tasks such as the one described above depends on many processes, including accurate peripheral extraction of frequency components from the complex vowel sounds \[23\], segregation of frequency components associated with two vowels, integration of frequency components to form two vowel representations, and comparison of the two representations with stored templates of the vowels. In addition, the double-vowel stimulus must be stored in memory long enough for all of these processes to occur. Thus, performance on this task relies on processing throughout the ascending auditory pathway in addition to cortical areas involved in attention and memory functions.

Recording of event-related potentials (ERPs) is a powerful technique for examining the effects of age on the neural correlates underlying speech separation and identification. ERPs reflect synchronous activity from large neuronal ensembles that are time-locked to sensory or cognitive processes. They are divided into automatic and controlled components. The former refers to brain waves that are stimulus-driven and occur irrespective of the listener’s attentional state whereas the latter occurs only when participants actively process the stimuli. In conjunction with behavioral measures, ERPs can thus help identify the extent to which low-level and high-level factors contribute to age-related differences in sound processing \[5\]. In particular, the effects of age on concurrent sound segregation can be examined by comparing ERPs elicited by two concurrent vowels with either the same or different \(f_0\).

Previous research using harmonic series with a mistuned partial have identified a novel auditory cortical ERP at \(-150\) ms referred to as the object-related negativity (ORN), whose amplitude correlated with listeners’ likelihood of reporting the presence of two simultaneous auditory objects \[2,4\]. The ORN is thought to index low-level sound segregation because it is present even for unattended stimuli \[1,2,4\]. The ORN may thus be viewed as indexing the automatic detection of the mistuned harmonic from a prediction based upon the fundamental frequency of the incoming stimulus. An independent laboratory has observed a similar ERP modulation during perceptual segregation of a narrow frequency band from background noise based on interaural timing information \[19,20\], suggesting that the ORN is not specific to mistuned stimuli.

A more speech-relevant task that is similarly amenable to ERP recording is concurrent vowel identification. In this task, identification rates are larger with increasing fundamental frequency separation (\(\Delta f_0\)) between the two vowels \[7,8,11,12\]. An auditory cortical component with similar latency, amplitude, and topographical distribution as the ORN appeared in the concurrent vowel task \[6\], and may thus provide a useful index of age-related changes in low-level speech segregation. This component likely reflects a similar process as the ORN because they are both present during the segregation and formation of concurrent auditory objects with different \(f_0\) and are both present even when participants ignored the stimuli. An additional response peaking \~250 ms was present in young adults when they actively performed the double-vowel task, but was absent when they ignored the sounds. Thus, this component may index more controlled processing such as the matching between the incoming signal and the stored vowel representations in working memory. These two responses thus reflect the automatic registration of \(\Delta f_0\) and the influence of attention-dependent processes, respectively, related to concurrent vowel segregation and identification based on \(\Delta f_0\).

In the present study, young and older adults with normal hearing were presented with a mixture of two phonetically different vowels with the same or different \(f_0\) and attempted to identify them during ERP recording. We tested the general hypothesis that performance in identifying both vowels would increase as a function of \(\Delta f_0\). We hypothesized that older adults would have more difficulty in identifying concurrent vowels, as assessed by identification rates and reaction times. Specifically, an interaction between age and \(\Delta f_0\) with increasingly lower identification rates or longer reaction times at smaller \(\Delta f_0\) would indicate age-related changes specific to vowel segregation processing. We further hypothesized that changes in processing as indexed by behavioral measures would be paralleled by ERP changes related to either bottom-up or top-down processing of concurrent vowels.

2. Materials and methods

2.1. Participants

Sixteen young adults (8 men and 8 women, age range = 19–34 years, mean age = 24.4 ± 4.6 years) and sixteen older adults (4 men and 12 women, age range = 61–76 years, mean age = 67.5 ± 5.3 years) participated after giving written informed consent according to the guidelines of the Baycrest Centre and the University of Toronto. All participants were compensated for their participation. They were all right-handed, spoke English as a first language, and had normal pure-tone thresholds measured in decibels (dB HL) for their age group at frequencies from 250 to 8000 Hz in both ears (see Table 1). Although our participants were screened for hearing loss, the older adults showed mild elevated hearing thresholds relative to young adults, especially in the high-frequency range. No differences between left and right ears occurred leading us to collapse threshold scores across this factor. A mixed-design analysis of variance (ANOVA) with age group (young and older) and frequency (250, 500, 1000, 2000, 4000, 8000) as factors yielded a main effect of age group, with older adults having higher thresholds than the young adults, \(F(1,30) = 39.71, P < 0.001\). The interaction between age group and
frequency was significant, \( F(5,150) = 9.16, P < 0.001 \), and was mainly due to a greater difference in hearing sensitivity between young and older adults at higher frequencies.

### 2.2. Stimuli and procedure

Stimuli were five steady-state American English vowels: ‘EE’ as in /i/, ‘AH’ as in /a/, ‘AE’ as in /æ/, ‘OO’ as in /u/, and ‘ER’ as in /ɜ/ [8]. Each vowel was 200 ms in duration (2000 samples at a 10-kHz sample rate). \( f_0 \) and formant frequencies were held constant. Formant frequencies were patterned after a male speaker from the North Texas region. The five vowels were less regionally variable than other American English vowels (P. Assmann, personal communication, 2004), allowing us to test Canadian participants. The source signal was the same in all five vowels, simulating ‘equal vocal effort’ (for more details of the stimuli, see [8]).

To create the double vowels, we summed the digital waveforms of two phonetically different vowels. Each pair contained one vowel with \( f_0 \) set at 100 Hz; the other vowel’s \( f_0 \) was 0, 0.25, 0.5, 1, 2, or 4 semitones higher (1 semitone = 1/12 octave). Each vowel was paired with every other vowel, giving a total of 120 different pairs. Stimuli were randomized in blocks of 120 trials. Participants were presented with five blocks of trials. This was a self-paced task and took \( \sim 1 \) h. The stimuli were presented binaurally through Sennheiser HD 265 headphones (Sennheiser Electronic Corporation, Old Lyme, CT) at 80 dB SPL. The same sound level was used for young and older adults to minimize effects of loudness recruitment due to elevated hearing thresholds in older adults.

Listeners were instructed to identify both vowels in the pair. They registered their response by sequentially pressing two of five keys, labeled ‘EE’, ‘AH’, ‘AE’, ‘OO’, and ‘ER’. All responses were made with the right index finger. Participants were told that a pair of non-identical vowels would be presented in each trial, with the inter-trial interval fixed at 1500 ms. No performance feedback was provided.

Prior to the experiment, each participant was presented with each stimulus individually (30 trials; five vowels by six \( f_0 \) levels) and was asked to identify the vowel by pressing the corresponding key on the keyboard. This ensured that our participants could accurately identify each vowel when presented individually. None of the participants had any difficulty in identifying the single vowels, and all reached a level of 83% correct or better (all but one participant were well over 90%). Most of the mistakes were confusions between ‘AH’ and ‘AE’. In cases of these confusions, participants were instructed on the difference between these two vowel sounds.

### 2.3. Electrophysiological recording and analysis

While participants performed the double-vowel task, electrophysiological responses were continuously collected, digitized, and filtered (250 Hz sampling rate; bandpass 0.05–50 Hz) from an array of 64 electrodes using NeuroScan SynAmps and Scan version 4.1 software (Compumedics USA, El Paso, TX) and stored for offline analysis. Eye movements were monitored with electrodes placed at the outer canthi and at the superior and inferior orbit. During the recording, all electrodes were referenced to the midline central electrode (i.e., Cz). For off-line data analysis, they were re-referenced to an average reference.

The analysis epoch included 200 ms of pre-stimulus activity and 1200 ms of post-stimulus activity. Trials contaminated by excessive peak-to-peak deflection (±120 \( \mu \)V) at the channels not adjacent to the eyes were automatically rejected before averaging. For each participant, ERPs were then averaged separately for each level of \( \Delta f_0 \) and electrode site. The ERPs were collapsed over correct and incorrect responses to facilitate comparisons between the various levels of \( \Delta f_0 \) and the two age groups. For each individual average, the ocular artifacts (e.g., blinks, saccades, and lateral movements) were corrected by means of ocular source components using the brain electrical source analysis software (BESA 3.0) [9,26]. ERPs were digitally bandpass filtered to attenuate frequencies outside 0.5–20 Hz, using BESA 3.0.

The effects of \( \Delta f_0 \) on ERP amplitude were quantified at nine fronto-central electrodes (i.e., Fz, F1, F2, FCz, FC1, FC2, Cz, C1, C2). To control for age-related differences in ERP amplitude, we examined the effect of age on the difference waves between ERPs elicited by double-vowel stimuli that share the same \( f_0 \) from those that have a different \( f_0 \). Five difference waves were thus calculated for young and older adults, respectively: 4 vs. 0, 2 vs. 0, 1 vs. 0, 0.5 vs. 0, and 0.25 vs. 0 semitone \( \Delta f_0 \). The difference waves were calculated based on the logic that increased brain activity as a function of \( \Delta f_0 \) reflects \( f_0 \)-based sound segregation. Furthermore, the difference wave allowed us to examine the interaction between age and \( \Delta f_0 \) on neural activity while controlling for general age-related differences in absolute amplitude of brain responses. We identified two sustained negative deflections associated with increased \( \Delta f_0 \) between the two vowel constituents, one from 130–270 ms and the other from 540–640 ms. The early negative deflection from baseline was further divided into two sub-peaks (130–170 ms and 230–270 ms) based

<table>
<thead>
<tr>
<th>Frequency (Hz)</th>
<th>Young Right ear</th>
<th>Young Left ear</th>
<th>Older Right ear</th>
<th>Older Left ear</th>
</tr>
</thead>
<tbody>
<tr>
<td>250</td>
<td>6.6 (5.4)</td>
<td>4.7 (7.4)</td>
<td>17.2 (9.3)</td>
<td>21.3 (16.1)</td>
</tr>
<tr>
<td>500</td>
<td>3.8 (7.0)</td>
<td>4.7 (6.2)</td>
<td>16.0 (10.4)</td>
<td>20.6 (12.4)</td>
</tr>
<tr>
<td>1000</td>
<td>2.8 (5.8)</td>
<td>3.8 (6.2)</td>
<td>15.0 (8.4)</td>
<td>20.6 (16.0)</td>
</tr>
<tr>
<td>2000</td>
<td>0.6 (5.7)</td>
<td>0.9 (7.6)</td>
<td>16.9 (14.1)</td>
<td>22.8 (22.1)</td>
</tr>
<tr>
<td>4000</td>
<td>1.3 (6.5)</td>
<td>2.8 (13.4)</td>
<td>23.1 (23.1)</td>
<td>24.7 (27.0)</td>
</tr>
<tr>
<td>8000</td>
<td>3.0 (4.1)</td>
<td>1.3 (6.7)</td>
<td>41.3 (23.0)</td>
<td>43.8 (28.6)</td>
</tr>
</tbody>
</table>
on previous studies of the mistuned harmonic paradigm [1,2,4] and the double-vowel paradigm [6].

2.4. Statistical analysis

The proportion of trials in which both vowels were correctly identified and the response time (RT) on correct trials taken to make the two button presses were analyzed in a mixed-design ANOVA with age group (young and older) as the between-subject factor and \( \Delta f_0 \) (0, 0.25, 0.5, 1, 2, or 4 semitones) as the within-subject factor. We excluded RTs that were 2 standard deviations (SD) faster or slower than the mean. To test for effects on ERP difference waves, we used a mixed design ANOVA with age group and \( \Delta f_0 \) as factors. Whenever appropriate, the degrees of freedom were adjusted with the Greenhouse–Geisser epsilon (\( \varepsilon \)). All reported probability estimates are based on the reduced degrees of freedom but the original degrees of freedom are reported. ERP activations for the three time spans defined above (130–170, 230–270, 540–640 ms) were related to double-vowel identification by simple correlations for individual participants across all six \( \Delta f_0 \) conditions. These simple correlations were submitted to one-sample \( t \) tests to test for differences from 0.

3. Results

3.1. Behavioral data

To demonstrate that our participants were able to perceive the synthetic vowels used in this study and to rule out general age-related perceptual or cognitive deficits, we analyzed the proportion of trials that participants were able to identify at least one of the vowels correctly. This measure does not reflect segregation processes as much as identifying both vowels because one vowel is usually dominant. Both groups of participants were proficient in identifying correctly at least one of the two vowels (young: mean = 94.90%, SE = 0.42; older: mean = 94.94, SE = 0.54).

Performance was not at ceiling, however, as shown by the significant effect of \( \Delta f_0 \) on correctly identifying at least one vowel, \( F(5,150) = 4.67, P < 0.005 \). There was no significant difference between young and older adults on this measure, \( F(1,30) < 1, P = 0.98 \), suggesting that there was no effect of age on the perception and identification of the dominant vowel. There was also no significant interaction between age group and \( \Delta f_0 \), \( F(5,150) = 0.51, P = 0.70 \).

Fig. 1 shows the group mean accuracy in identifying both vowels correctly for young and older participants at each level of \( \Delta f_0 \). Overall performance level for the young adults is somewhat lower than previous studies using this paradigm [7,8,12]. This is likely due in part to our use of relatively naive participants compared to past investigations that used highly trained individuals. In agreement with earlier studies, participants showed an increase in performance with increasing \( \Delta f_0 \), \( F(5,150) = 30.17, P < 0.001 \). Pairwise comparisons indicated that even the smallest \( \Delta f_0 \) between the two vowel constituents led to significant improvement in vowel identification. Performance improved gradually with increasing \( \Delta f_0 \) up to 0.5 semitone \( (P < 0.005) \) after which there was no significant improvement in performance from 0.5 semitone to 4 semitones \( \Delta f_0 \). Young adults showed an overall higher level of accuracy compared to older adults, \( F(1,30) = 5.59, P < 0.025 \). There was no significant interaction between age group and \( \Delta f_0 \), \( F(5,150) = 1.04, P = 0.39 \), suggesting that young and older adults took advantage of the \( \Delta f_0 \) cue for vowel segregation to a similar extent.

If older adults had more difficulty extracting \( \Delta f_0 \) compared to young adults, one might expect them to take more time to identify the second vowel, especially for small \( \Delta f_0 \). We tested this hypothesis by comparing the time needed to generate the first response (RT1) and the time needed to generate the second response (RT2) when both vowels were correctly identified in young and older adults. Fig. 2 shows the group mean RT1 and RT2 values at each \( \Delta f_0 \). Overall, older adults tended to take more time than young adults in producing their first response, although the effect did not reach significance, \( F(1,30) = 3.87, P = 0.06 \). However, older adults did take more time than young adults in generating their second response, \( F(1,30) = 5.30, P < 0.05 \), consistent with general age-related slowing (Salhoushe, 1996). RT1 did not show a main effect of \( \Delta f_0 \), \( F(5,150) = 1.37, P = 0.26 \), whereas RT2 was significantly longer at smaller values of \( \Delta f_0 \), \( F(5,150) = 4.66, P < 0.01 \). No significant interaction between age group and \( \Delta f_0 \) occurred for RT1, \( F(5,150) = 1.66, P = 0.18 \), whereas RT2 showed a significant interaction, \( F(5,150) = 3.28, P < 0.05 \), with older adults taking more time compared to young adults to respond as \( \Delta f_0 \) became smaller. Along with the identification data presented above, the RT data suggest that extracting the dominant vowel was relatively easy across all levels of
DC0, whereas extracting the second vowel from the signal was particularly difficult for older adults, especially at small values of DC0.

3.2. Electrophysiological data

Fig. 3 shows the group mean ERPs elicited by the double-vowel stimuli for young and older adults, separately. Both groups showed a characteristic N1-P2 response following stimulus onset, with larger responses in older adults. This amplitude difference could be in part due to longer response-stimulus intervals in older adults as a result of slower responding to the two vowels in the self-paced task (see Fig. 2), in addition to any age-related changes in auditory evoked response generators. To control for age-related ERP amplitude differences, we simply examined difference ERPs between double-vowel stimuli with different DC0 levels.

To isolate bottom–up and top–down neuroelectric activity associated with DC0 processing, we took the difference ERPs between double-vowel stimuli with different DC0 levels.

To isolate bottom–up and top–down neuroelectric activity associated with DC0 processing, we took the difference ERPs between double-vowel stimuli with different DC0 levels.

Interactions between age group and DC0 indicate differences in DC0 processing between young and older adults. The ORN had a maximum amplitude at 157 ms and was larger in young adults, F(1,30) = 5.25, P < 0.05. The effect of DC0 on amplitude was significant, F(4,120) = 5.59, P < 0.005, and the linear contrast demonstrated that amplitude increased with increasing DC0 in both groups, F(1,30) = 11.65, P < 0.005 (see Fig. 5). As with identification performance, there was no interaction between age group and DC0 on ORN amplitude, F(4,120) = 1.05, P = 0.38. The ORN amplitude correlated significantly with double-vowel identification performance in young adults, r(4) = 0.26, t(15) =
2.33, \( P < 0.05 \), but not older adults, \( r(4) = 0.16, t(15) = 1.35, P = 0.20 \).

In contrast to the ORN, the N2b (250 ms after sound onset) had similar amplitude between the two age groups, \( F(1,30) = 0.31, P = 0.58 \). The effect of \( \Delta f_0 \) on amplitude was marginally significant, \( F(4,120) = 2.51, P = 0.06 \). N2b amplitude did not correlate significantly with performance in young adults, \( r(4) = 0.06, t(15) = 0.54, P = 0.60 \), but was significant for older adults, \( r(4) = 0.26, t(15) = 2.16, P < 0.05 \). Despite the similar amplitude of the N2b in young and older adults, the scalp topography of the response showed a more frontal distribution in older adults, although this was not statistically reliable.

The processing of \( \Delta f_0 \) was also associated with a sustained late negativity in the difference wave (LN) that was maximal at fronto-central sites. The LN was only present in young adults (see Fig. 4). We quantified peak amplitude in the time span of 540–640 ms for both young and older adults. The rightmost plot in Fig. 5 shows the amplitude dependence of LN on \( \Delta f_0 \). The peak had a maximal amplitude at 593 ms and was larger in young than older adults, \( F(1,30) = 8.48, P < 0.01 \). There was no significant effect of \( \Delta f_0 \) on LN, \( F(4,120) = 1.96, P = 0.12 \), and no interaction between age group and \( \Delta f_0 \), \( F(4,120) = 0.80, P = 0.50 \). LN amplitude did not correlate significantly with performance in young adults, \( r(4) = 0.19, t(15) = 1.58, P = 0.14 \), or older adults, \( r(4) = -0.22, t(15) = -1.85, P = 0.08 \).

4. Discussion

The ability to correctly identify two vowels presented simultaneously improved with increasing \( \Delta f_0 \), consistent with previous studies [7,8,11,12]. In the present study, older adults were less accurate in identifying both vowels, consistent with the proposal that speech perception problems commonly observed in older adults [14,25,29] may be in part related to deficits in parsing concurrent sounds [3,17]. It is unlikely that the age effects on accuracy during the task were due to general perceptual and cognitive declines because the young and older adults did not differ in identifying at least one of the two vowels. Age effects on performance manifested themselves only after the initial identification of the dominant vowel while attempting to extract the non-dominant vowel.

The current results extended previous work by showing that although older adults with normal hearing exhibited an overall decrease in identification rate, they were able to use \( \Delta f_0 \) between the two vowels to the same extent as young adults. This is in contrast to effects of hearing impairment on vowel segregation in which substantial deficits in using the \( \Delta f_0 \) cue were observed [33]. However, the analysis of RT data suggested that, despite their ability to use \( \Delta f_0 \) to the same extent as young adults, normal hearing older adults required more processing time at small values of \( \Delta f_0 \).

Together, the accuracy and RT data suggest that older adults may allocate more processing time on the stimuli, especially for the most difficult levels of \( \Delta f_0 \). One possibility is that older adults may rely more on slower controlled processes because of age-related changes in the cochlea that are likely to impoverish the quality of auditory information reaching higher auditory centers. This initial distortion of the acoustic waveform may lead to further reductions in fidelity as the signal is submitted to further analyses, resulting in the ultimate disruption of higher-order processes such as speech comprehension. In other words, older adults may have recruited more top–down resources during this additional processing time compared to young adults for small \( \Delta f_0 \) stimuli.

Current models of auditory scene analysis postulate both low-level automatic processes and higher-level controlled or schema-based processes [10]. Whereas the automatic processes use basic stimulus properties to segregate the incoming sounds, controlled processes use previously learned criteria to group the acoustic input into meaningful sources. The use of prior knowledge is particularly evident in adverse listening situations such as a cocktail party scenario. In an analogous laboratory situation, a sentence’s final word embedded in noise was more easily detected when it was contextually predictable [25], and older adults benefited more than young from contextual cues in identifying the sentence’s final word embedded in noise [25]. Thus, schema-driven processes provide a way to resolve perceptual ambiguity in complex listening situations and older adults appear to rely more heavily on controlled processing.

The electrophysiological results provide converging evidence for age-related declines in low-level sound processing that are compensated by top–down processing [5] such as focusing attention on the stimuli or inhibiting irrelevant material [21,32]. Overall ERP amplitude was larger in older adults, an effect at least partly attributable to larger response-stimulus intervals in older adults who responded slower in the self-paced task. To control for this, we examined difference ERPs related to \( \Delta f_0 \) processing. Young adults showed a larger early negative wave (ORN) at ~150 ms compared to older adults. Furthermore, double-vowel identification correlated significantly with ORN amplitude in young but not older adults.
In the present study, the age-related reduction in ORN amplitude suggests impairment in the early bottom-up parsing of concurrent vowels. This interpretation is based on the fact that the ORN becomes larger as a function of \( \Delta f_0 \), and appears in young adults whether or not they paid attention to concurrent vowel sounds [6]. The reduced ORN amplitude may be related to declines in frequency selectivity [31], which might impact the ability to form accurate \( f_0 \) representations for concurrent vowels. For example, decreased frequency selectivity would make frequency components from different vowels more likely to be integrated into the same auditory filter. This would imply that older adults might compensate for decreased frequency selectivity by using other cues for segregation [13]. Utilization of other cues would not be reflected in the difference ERPs, however, because they are only designed to isolate \( f_0 \)-based processing.

The N2b was preserved in older adults in the present study and was previously found to occur only when young adults actively processed the vowel stimuli [6]. The N2b also correlated significantly with performance in older adults but not young adults. These findings suggest that older adults have impaired bottom-up processing (ORN) of double-vowel stimuli but preserved top-down processing (N2b). These ERP data along with the RT data might explain why older adults did not demonstrate impaired behavioral utilization of \( \Delta f_0 \), as indexed by identification accuracy for both vowels. This is consistent with a recent study of gap detection in young, middle-aged, and older adults, demonstrating preserved neural responses related to controlled gap detection but age-related impairments of neural responses related to automatic gap detection [5]. Future experiments that either use passive listening conditions or stimuli that yield equivalent performance in young and older adults could provide additional evidence for compensatory mechanisms in older adults during concurrent sound segregation.

In summary, the current study tested young and older adults’ ability to identify concurrently presented vowels, providing evidence for age-related changes in auditory scene analysis and its neural underpinnings. Older adults had poorer overall performance although they benefited to the same extent as young adults from \( \Delta f_0 \) between the two vowels by taking more processing time for particularly difficult stimuli. This lack of impairment was likely facilitated by an increase in processing time by older adults, especially at small \( \Delta f_0 \). A neural process related to bottom-up parsing of concurrent vowels using \( \Delta f_0 \) was smaller in older adults, suggesting deficiencies in low-level auditory scene analysis. In contrast, a neural process related to top-down parsing was not affected by age suggesting preserved controlled processing of double-vowel sounds. Age-related deficits in concurrent sound segregation observed in the current study likely contribute to speech processing difficulties encountered by older adults in complex acoustic environments.

Acknowledgments

The authors thank Ben Dyson, Cheryl Grady, Erin Hannon, and Tony Herdman for comments on a previous draft of the manuscript, and Yu He, Kelly McDonald, and Chenghua Wang for technical assistance. We are particularly indebted to Peter Assmann and Quentin Summerfield for generously providing the vowel stimuli. This research was funded by grants from the Canadian Institute for Health Research, the Natural Sciences and Engineering Research Council of Canada, and the Premier’s Research Excellence Award.

References


