Age-related differences in the sequential organization of speech sounds

Stefanie A. Hutka, Claude Alain,[a] and Malcolm A. Binns
Rotman Research Institute, Baycrest Center, 3560 Bathurst Street, Toronto, Ontario M6A 2E1, Canada

Gavin M. Bidelman
Institute for Intelligent Systems, School of Communication Sciences & Disorders, University of Memphis, 807 Jefferson Avenue, Memphis, Tennessee 38105

(Received 30 April 2012; revised 27 March 2013; accepted 2 April 2013)

This study investigated the effects of age on listeners' tendency to group speech tokens into one or two auditory streams. Younger and older adults were presented with sequences of four vowel sounds, which were arranged according to the proximity of first-formant frequencies between adjacent vowels. In Experiment 1, participants were less accurate in identifying the order of the four vowels and more likely to report hearing two streams when the first-formant alternated between low and high frequency and the overall difference between adjacent vowels was large. This effect of first-formant continuity on temporal order judgments and probability of hearing two streams was higher in younger than in older adults. In Experiment 2, participants indicated whether there was rhythm irregularity in an otherwise isochronous sequence of four vowels. Young adults' thresholds were lower when successive first-formants ascended or descended monotonically (condition promoting integration) than when they alternated discontinuously (condition promoting streaming). This effect was not observed in older adults whose thresholds were comparable for both types of vowel sequences. These two experiments provide converging evidence for an age-related deficit in exploiting first-formant information between consecutive vowels, which appear to impede older adults' ability to sequentially group speech sounds over time.

© *2013 Acoustical Society of America*. [http://dx.doi.org/10.1121/1.4802745]

PACS number(s): 43.66.Lj, 43.66.Sr [LD] Pages: 4177–4187

I. INTRODUCTION

Older adults often have trouble listening to a conversation, especially in adverse listening situations such as in the presence of reverberation (Middelweerd *et al.*, 1990; Gordon-Salant and Fitzgibbons, 1993), or when task-relevant speech stimuli are mixed with other task-irrelevant speech sounds (e.g., Duquesnoy, 1983a,b; Pichora-Fuller *et al.*, 1995; Helfer and Freyman, 2008; Schneider *et al.*, 2010). Although age-related declines in hearing sensitivity undoubtedly contribute to speech comprehension deficits (Humes and Roberts, 1990; Humes, 1996; Schneider *et al.*, 2005), there is increasing evidence for other contributing factors to the speech-in-noise problems observed in older adults. These include a failure to segregate mixtures of sounds (e.g., Alain *et al.*, 2006), deficits in filtering out task-irrelevant stimuli (e.g., Hasher and Zacks, 1984; Schneider *et al.*, 2007), deficits in working memory (e.g., Gordon-Salant and Fitzgibbons, 1997), and reduced speed of processing (e.g., Salthouse, 1996). Such deficits are likely to interfere with the integration of acoustic data into a coherent scene.

Alain *et al.* (2006) posited that older adults' difficulty understanding what one person is saying in the presence of background talkers may be related to a failure to effectively organize the mixture of sounds. Though this hypothesis has generally been supported by recent literature, the effects of age on the perceptual organization of sounds may differ for auditory events that occur simultaneously vs those that occur sequentially. Indeed, while older adults typically show deficits parsing concurrent sounds (e.g., Alain *et al.*, 2001; Snyder and Alain, 2005), the literature shows inconsistencies in their ability to parse sequentially presented sounds; some studies report an age-related decline in auditory stream segregation (e.g., Grimault *et al.*, 2001) while others do not (e.g., Trainor and Trehub, 1989; Snyder and Alain, 2007a).

An example of age-related difficulty parsing simultaneous sounds is demonstrated by older adults' higher thresholds for detecting a mistuned harmonic (Alain *et al.*, 2001; Grube *et al.*, 2003; Zendel and Alain, 2012). Low thresholds for mistuned harmonics are understood to indicate effective sound segregation. This age-related deficit in detecting inharmonicity (i.e., mistuning) remains significant even after accounting for differences in audiometric thresholds between age groups (Alain *et al.*, 2001). Further evidence for an age-related decline in concurrent sound segregation comes from the double vowel task, in which participants must identify two different vowels that are presented simultaneously (Summerfield and Assmann, 1989; Culling and Darwin, 1993; Assmann and Summerfield, 1994). While older adults can identify at least one of the two vowels accurately, their accuracy in identifying both vowels is significantly lower than young adults (Snyder and Alain, 2005; Vongpaisal and Pichora-Fuller, 2007). In these studies, older adults were considered to have normal hearing for their age [i.e., $\leq$25 dB

---

[a] Author to whom correspondence should be addressed. Electronic mail: calain@research.baycrest.org

hearing level (HL) for octave frequencies between 250 and 2000 Hz], indicating that age-related deficits in concurrent sound segregation persist in the absence of clinically significant hearing loss.

While aging appears to impair listeners' ability to separate sounds that occur simultaneously, older adults often show no significant deficits in organizing sounds that occur sequentially. Trainor and Trehub (1989) used a temporal order judgment task and found that older adults with normal hearing (i.e., thresholds ≤ 25 dB HL) were less accurate overall than young adults. However, age differences in performance were minimally affected by experimental manipulations aimed at promoting stream segregation. In a related study, Alain *et al.* (1996) used a selective attention task in which younger and older normal hearing adults were told to press a button in response to infrequent pure tone targets embedded in a sequence of pure tone distracters. The frequency separation between the distracters was manipulated either to promote or impede auditory stream segregation. That is, the greater the frequency separation between the tones composing the sequence, the greater the listener's probability of reporting having heard two concurrent streams (for reviews see, Alain and Arnott, 2000; Moore and Gockel, 2002; Snyder and Alain, 2007b). While older adults were generally slower to respond than younger adults, both groups showed a similar improvement in performance in situations that promoted stream segregation. In a subsequent study, Snyder and Alain (2007b) used the more typical pure-tone patterns of "ABA—ABA—" for assessing auditory stream segregation, in which "A" and "B" are tones of different frequencies and "—" is a silent interval. They measured participants' probability of subjectively hearing two streams and found that older adults reported auditory streaming to the same extent as young adults. Together, these findings suggest that sequential auditory stream segregation is largely preserved in elderly adults. However, in a study that employed harmonic complex tones (as opposed to pure tones) in ABA paradigms, young adults typically perceived more streaming than did older adults with either normal (i.e., octave frequency thresholds ≤ 10 dB HL for octave frequency between 250 and 8000 Hz) or impaired (i.e., >30 dB HL at 500, 1000, and 2000 Hz) hearing for their age (Grimault *et al.*, 2001). The above studies suggest that age may have a detrimental effect on the sequential organization of sounds when complex sounds are used instead of pure tones. Spectro-temporally rich sounds, such as those used in spoken communication (e.g., vowels) involve smooth fundamental-frequency (f_0) and formant transitions between adjacent utterances that may play an important role in the perceptual organization of speech sounds. However, despite their higher ecological validity, few studies have used well-controlled speech stimuli to induce stream segregation.

Following many studies on the perception of temporal order, Dorman *et al.* (1975) was among the first to examine whether speech sounds (i.e., consonant-vowel-consonant, or vowel) could be grouped based on the contiguity of first-formants (f_1) between successive vowels. They used four-vowel sequences that were repeated at a high rate. The synthetic vowel sounds shared the same voice pitch (i.e., f_0) but the order of the four vowels was manipulated to promote grouping based on the frequency of the f_1 between adjacent vowels. Young, normal-hearing listeners were asked to write down the perceived order of the vowels. The authors found that the accuracy of reporting the items' temporal order was facilitated when adjacent vowels contained similar (i.e., nearby) f_1 frequencies, which rose monotonically. Difficulty in identifying the sequential order of the vowels occurred with more disjunct formant frequencies. That is, greater deficits in temporal order judgment occurred when the f_1 between adjacent vowels alternated between low and high frequency (i.e., the sequence was discontinuous). The detrimental effects in performance between conditions were explained in terms of differences in stream segregation, triggered by the alternation in f_1 and the larger frequency differences between adjacent vowels. Subsequent studies using repeating three- (Nooteboom *et al.*, 1978) or six-vowel (Gaudrain *et al.*, 2008; Devergie *et al.*, 2011; Gaudrain *et al.*, 2012) sequences have also shown that increasing the f_0 difference (i.e., pitch) of adjacent vowels also promotes the segregation of sequences of vowels into two separate streams.

In the present study, we investigated the effects of age on listeners' tendency to group successive vowels according to first-formant frequency. We used a paradigm analogous to that used by Dorman *et al.* (1975), which consisted of presenting a sequence of four vowels arranged according to their f_1 frequency while keeping the f_0 (i.e., voice pitch) constant between vowels. Vowel sequences provide a reliable and useful tool for investigating the perceptual organization of speech sounds that may otherwise be obscured by additional syntactic and semantic information present in sentences (Warren *et al.*, 1996). We used naturally produced vowel sounds from four different speakers (two male and two female) to examine whether findings from prior studies would generalize to different naturally spoken vowels, and to rule out talker and/or stimulus specific effects (i.e., a particular set of vowels). Four vowels were arranged into six different sequences either to promote or impede auditory stream segregation based on the continuity of f_1 frequency (continuous vs alternating or discontinuous). In Experiment 1, we used a forced-choice procedure, in which participants were required to identify the order of the four-vowel sequence as well as indicate whether they perceived one or two streams. This design allowed us to examine the relationship between objective and subjective measures of auditory streaming. In Experiment 2, participants performed a rhythm judgment task requiring them to detect whether there was a change in the temporal regularity of the vowel sequence. According to previous work (e.g., Dorman *et al.*, 1975), accuracy of vowel-order identification should be lower when the f_1 alternates between high and low frequencies, than when it ascends or descends progressively between vowels. Moreover, the probability of hearing two streams should vary with the sequence type, with greater probability of reporting hearing two streams when f_1 alternates and yields large differences in f_1 frequency between adjacent vowels. Additionally, temporal thresholds for detecting a change in rhythm should be higher in conditions promoting stream segregation where one must listen to and monitor multiple streams. If age impairs sequential organization of speech sounds, then we should observe an interaction between the age groups and sequence types.

II. EXPERIMENT 1

A. Methods and materials

1. Participants

Forty-seven volunteers gave their informed consent in compliance with a research protocol approved by the Baycrest Center and the University of Toronto. One young adult was excluded because he/she did not complete the pure-tone threshold test, and one young adult was lost to attrition. Six older adults were excluded because they showed mild hearing loss (threshold >25 dB HL) at 2000 Hz. The final sample included 19 young adults [age range = 19 to 30 yr, mean (M) = 21.60 yr, standard deviation (SD) = 2.79,13 females] and 20 older adults (age range = 65 to 83, M = 70.15 yr, SD = 5.32, 12 females). All participants were screened during a phone interview; only those participants without any self-reported neurological problems and who described their general health as excellent or good were eligible to participate in the study. All participants were non-musicians. Musicianship was defined as regularly practicing an instrument or conducting an ensemble at least once a month, having a degree or diploma in music, or taking music lessons for more than 10 yr at any point in life. This was cumulative; if the participant took five years as a child and five years as an adult, he or she would be considered "musician" and excluded from participation. All participants were native English speakers. All older adults scored 27 or higher on the Mini-Mental State Exam (Folstein *et al.*, 1975). A score greater than or equal to 25 is considered normal cognition.

Participants' hearing ability was assessed prior to beginning the experimental session using pure-tone air conduction thresholds measured at each octave frequency between 250 and 8000 Hz for the left and right ears. Audiometric thresholds at each frequency averaged across ears for the two groups are summarized in Table I. For octave frequencies from 250 to 2000 Hz, older adults had higher pure-tone thresholds than young adults [$F(1, 37) = 48.70$, $p < 0.001$, $\eta_p^2 = 0.568$]. There were no significant differences between the two ears [$F(1,37) < 1$], nor was the interaction between group and ear significant [$F(1,37) < 1$]. The group x frequency interaction was not significant, $F(1,111) = 1.45$, $p = 0.242$, $\eta_p^2 = 0.038$. When considering octave frequencies from 250 to 8000 Hz, we found a significant interaction between group and frequency [$F(5,185) = 35.46$, $p < 0.001$, $\eta_p^2 = 0.489$] and observed greater age differences at higher (i.e., 4000 and 8000 Hz) than lower frequencies (250 to 2000 Hz).

TABLE I. Experiment 1: Mean and standard deviation of audiometric thresholds (dB HL) for young and older adults averaged over the left and the right ears.

	250 Hz	500 Hz	1000 Hz	2000 Hz	4000 Hz	8000 Hz
Young adults	6.18	6.58	5.13	4.61	−0.92	−3.03
Standard deviation	5.07	5.01	5.93	6.33	5.33	6.05
Older adults	20.85	19.25	15.55	17.70	26.48	42.25
Standard deviation	8.97	7.73	8.88	10.55	14.00	19.72

In addition to audiometric thresholds, we also measured speech reception in noise using the Quick Speech-In-Noise test (QuickSIN; version 1.3). The QuickSIN was developed using young normal-hearing listeners and provides an efficient means to measure speech understanding in noise (Killion *et al.*, 2004). In the present study, participants were presented with four lists of six sentences with five key words per sentence embedded in four-talker babble noise. The sentences were presented at a combined amplitude of 70 dB sound pressure level (SPL) using pre-recorded signal-to-noise ratios (SNRs) which decreased in 5 dB steps from 25 dB (very easy) to 0 dB (very difficult). After each sentence presentation, participants repeated the sentence. They were given one point for each correctly repeated key word. "SNR loss" was determined by subtracting the total number of words correct from 25.5. This number represents the SNR required to correctly identify 50% of the key words in the target sentences (Killion *et al.*, 2004). This test was administered half-way through the experiment in order to provide a change of pace from the experimental task. The SNR loss was higher (i.e., poorer) in older [M = 1.86, standard error (SE) = 0.37] than in younger adults (M = 0.04, SE = 0.22), $F(1,37) = 17.37$, $p < 0.001$, $\eta_p^2 = 0.32$. This indicates that older adults' speech recognition was more hindered by noise than young adults.

2. Stimuli

The stimuli consisted of the vowels /ae/ (as in *cat*), /3/ (as in *her*), /i:/ (as in *see*) and /u:/ (as in *moose*), henceforth referred to as "ae," "er," "ee," and "oo," respectively. The vowels were produced by four different native English speakers (two males and two females, aged 18 to 35 yr), who were instructed to generate 10 exemplars of each vowel every two seconds, while trying to maintain the same pitch, and to reduce variation in prosody. Vowel stimuli were recorded in a double-walled sound-attenuated chamber (IAC model 1204A, Electromedical Instruments, Mississauga, ON) using a large-diaphragm Shure KSM44 condenser microphone. Stimuli were digitized using a sample rate of 44.1 kHz via an XLR-to-USB audio interface (Sound Devices USBPre 1.5) and recorded to a PC running Adobe Audition (version 1.5).

For each vowel produced by each speaker, we chose the most typical utterance, defined as the vowel sound that had the least variation in pitch, and was closest to the mean frequency profile for all 10 recordings of that sound. All vowels were edited to be 200 ms in length, matching the length often used for artificially generated vowels (as in Assmann and Summerfield, 1994). They were then adjusted to have the same root mean square (RMS) amplitude using Adobe Audition. To maintain the realistic quality of the vowels, no further manipulations were made to the stimuli. The stimuli were presented binaurally at an average intensity of 75 dB SPL via ER-3A insert earphones (EarTone, Indianapolis, IN).

Stimulus intensity was measured using a Larson-Davis SPL meter (model 824, Provo, UT) with the transducer's plastic tubes attached to a 2-cc coupler (Model AEC100l). Separate calibrations were made for left and right ear channels.

We analyzed the frequency content of each vowel using Praat (version 5.2.17), which generated f_0 and four formant values. Table II displays the frequency-profiles for each vowel sound from the f_0 through the fourth formant. Values refer to the mean frequency over the entire vowel duration. Vowels in the lower frequency range were used to minimize the effects of high-frequency hearing loss (e.g., presbycusis) typically observed in older individuals, including those in our subject pool (see Table I).

3. Procedure

Following hearing assessment, participants took part in a single 5-h session that comprised the experimental tasks. Participants were seated in a chair, ~85 cm from a computer monitor, within the sound-attenuated chamber. Prior to the experiment, participants completed a brief, self-paced vowel identification task where each vowel was presented 500 ms after a button press. This task was used to ensure that all participants could accurately identify the speech sounds. The task comprised 16 trials including the four vowel sounds spoken by all four voices. The stimuli were presented in a random order and participants were asked to press one of the four buttons corresponding to the heard vowel. We used a custom MATLAB program (Version 5.3, The Mathworks, Natick, MA) on a Pentium 4 PC (soundcard: SoundBlaster Live! Wave Device) to run the experiment. Participants were required to correctly identify all 16 sounds before proceeding to the experimental task.

TABLE II. Voice fundamental (f_0) and formant frequencies (f_1 to f_4) for each vowel per speaker. All values are reported in Hz.

Speaker	Formant	ee	ae	er	oo
Female 1	f_0	156	156	148	162
	f_1	262	1040	519	316
	f_2	2461	1648	1339	600
	f_3	3596	2814	1542	2438
	f_4	4039	4331	3943	3847
Female 2	f_0	262	234	244	263
	f_1	361	1043	636	429
	f_2	2871	1490	1575	1345
	f_3	3482	1815	2052	2613
	f_4	4362	3079	3914	3841
Male 1	f_0	158	151	136	151
	f_1	307	860	532	327
	f_2	2351	1629	1286	1098
	f_3	2847	2714	1762	2493
	f_4	4819	4540	3374	3364
Male 2	f_0	156	125	134	137
	f_1	276	823	527	284
	f_2	2011	1329	1422	1582
	f_3	2940	2435	1690	1928
	f_4	3532	3600	3221	3255

Before beginning the experimental task, participants were provided with written instruction, as well as exemplars of the various sequences. They were also familiarized with the response buttons and paradigm prior to initiating the task. In the experimental paradigm, participants were presented with six different sequences per speaker containing the four vowel sounds arranged in a specific order (i.e., ee-ae-er-oo, ee-er-ae-oo, ee-oo-ae-er, ee-oo-er-ae, ee-ae-oo-er, or ee-er-oo-ae). For the first four sequences, hereafter referred to as "continuous," the f_1 frequency increased or decreased progressively between two or three consecutive vowels; the mean difference in f_1 (i.e., the difference in Hz between two adjacent vowels) averaged across all four voices was 320 Hz. In the last two sequences (i.e., ee-ae-oo-er, or ee-er-oo-ae), hereafter refer to as "discontinuous," the f_1 frequency alternated (low-high; high-low) between consecutive vowels with the mean difference in adjacent f_1 being 427.25 Hz. Spectrograms of exemplar "continuous" and "discontinuous" vowel conditions are shown in Fig. 1.

The vowel "ee" was arbitrarily selected as the first vowel in each sequence, as per Dorman *et al.* (1975). The inter-stimulus interval between successive vowels within a four-vowel pattern and between the last and the first vowel of the four-vowel patterns within the sequence was 12 ms. Each trial included 24 repetitions of a four-vowel sequence. The total duration of the sequence was 20.35 s, including 2.5 s of rise and fall time (i.e., linear ramps, fade in/out). This sequence duration was chosen to allow streaming to build up and stabilize. The slow rise and fall times were required to prevent participants from using the first vowel sound as an anchor to process, and thus, easily memorize, the sequence (Dorman *et al.*, 1975).

At the end of each sequence, participants were asked to identify the heard sequence from among six alternative choices visually presented on the computer screen (the *objective* task). After choosing the sequence that best matched the one presented, participants were also asked to subjectively indicate whether they heard one or two streams of sound via a keyboard prompt. They were told that there could be some change in their perception throughout the sequence. Participants were instructed that if their perception flipped back and forth between hearing 1- and 2-streams, to make their response based on what they felt best represented the percept of the entire sequence. Participants' subjective responses were defined as the percentage of trials heard as two streams of sound, and ranged from 0% (participants always responded that they heard a single stream of sound) to 100% (participants always responded that they heard two streams of sound).

Each participant completed six blocks of 36 trials (each trial involved listening to one sequence and then answering the two questions about the sequence). Within a block of trials, each sequence type was presented six times in a random order. Participants were given short breaks (~5 min) between blocks. Each participant completed three blocks of one of two female speakers (Female 1 or Female 2), and three blocks of one of two male speakers (Male 1 or Male 2). The order of speaker gender was counterbalanced between participants.

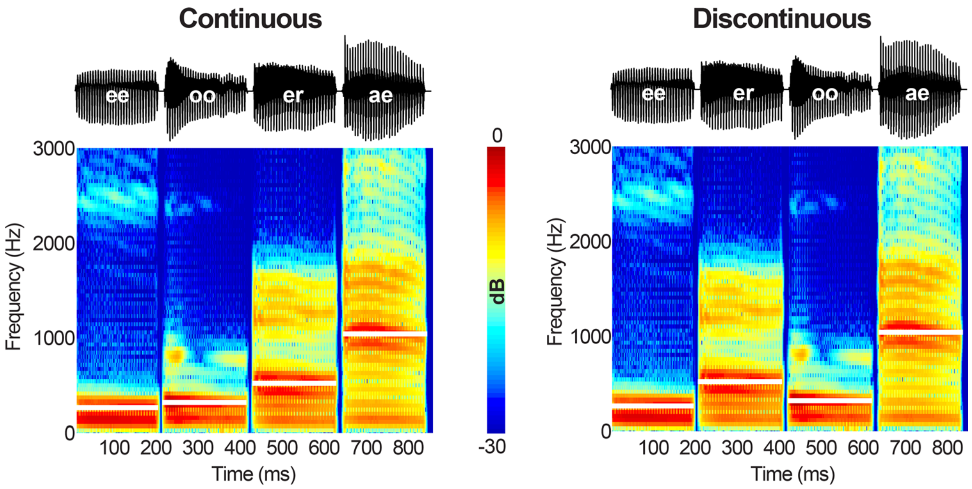

FIG. 1. (Color online) Spectrograms for an exemplar of the continuous sequence-type (ee-oo-er-ae), and an exemplar of the discontinuous sequence-type (ee-er-oo-ae). White lines mark the frequencies of vowel first formants.

4. *Statistical analyses*

Participants' responses to the objective task were converted to *d*-prime (d') scores. For each stimulus type, d' was calculated in three steps. First, for a given sequence type, hit rate was computed as the proportion of trials correctly identified when actually presented [e.g., Pr(respond "Seq_1"/ Seq_1 presented)]; false alarm rate was computed as the proportion of trials presenting one of the other sequence orders that was incorrectly identified as the index stimulus [e.g., Pr(respond "Seq_1"/$Seq_{2,3,4,5,\text{ or }6}$ presented)]. The second step was to transform the hit rate and false alarm rate using the inverse standard normal cumulative distribution function to amplify differences in proportions close to zero or one and reduce ceiling/floor effects. Last, the difference between the two transformed proportions yielded the d' accuracy score.

A 2×2 mixed design analysis of variance (ANOVA) was used to examine components of variability in accuracy on the objective task. The model included age group as a between-subject factor, and sequence type as a within-subject factor. Twelve replications—six sequences by speaker gender—were observed for each participant. Statistical analyses were implemented using SPSS v20.

B. Results

1. *Effects of speaker voice on accuracy and subjective reports*

First, we tested whether performance varied as a function of the talker voices. Separate ANOVAs for the two male and two female voices did not reveal significant differences in accuracy nor subjective reports [female speakers: $F(1, 38) < 1$; male speakers: $F(1, 38) = 2.747$, $p = 0.106$]. The main effect of speaker gender (i.e., male vs female) was not significant [$F(1,38) < 1$], nor did speaker voice interact with any of the other factors. Consequently, the results for the female and male speakers were combined together and subsequent analyses focus on within and between-group differences in processing continuous and discontinuous streaming conditions.

2. *Effects of age and first formant transition on accuracy and subjective reports*

Figure 2 shows the group mean d' measure for each of the six vowel sequences. Overall, young adults (M = 2.25, SE = 0.220) were more accurate than older adults (M = 1.14, SE = 0.198) in identifying the order of the vowels in the sequences, $F(1, 37) = 15.12$, $p < 0.001$. Moreover, participants were more accurate in the continuous (M = 2.09, SE = 0.210) than in the discontinuous (M = 0.89, SE = 0.722) condition, $F(1, 419) = 191.94$, $p < 0.001$. There was a significant interaction between age and condition, $F(1, 419) = 10.57$, $p < 0.001$, which was due to greater age-related differences in the continuous than in the discontinuous condition. The main effect of age on accuracy could be partly related to age differences in audiometric thresholds. To assess this possibility, the effects of age on accuracy were re-analyzed using the mean audiometric thresholds for octave frequencies between 250 and 2000 Hz from both ears as a covariate in an analysis of covariance. This analysis yielded a main effect of age, $F(1,36) = 6.90$, $p = 0.013$, indicating that the effects of age on accuracy remained even after controlling for differences in hearing sensitivity.

Figure 3 shows the mean subjective response for the six vowel sequences in young and older adults. Participants were more likely to report hearing two streams in the

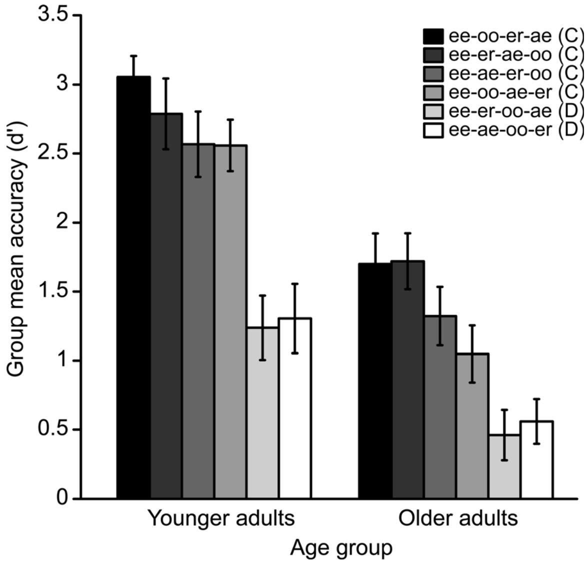

FIG. 2. Mean sensitivity measure (d') in young and older adults. The letters C and D refer to continuous and discontinuous conditions, respectively. Error bars show standard error of the mean.

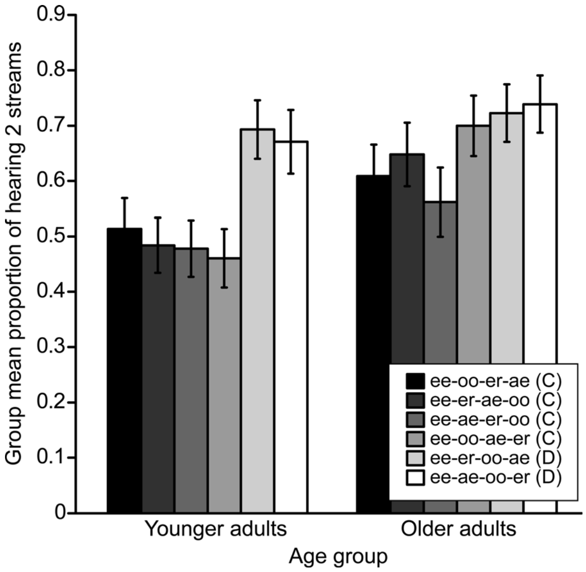

FIG. 3. Mean proportion of reporting hearing two streams in young and older adults. The letters C and D refer to continuous and discontinuous conditions, respectively. Error bars show standard error of the mean.

discontinuous ($\mathrm{M} = 0.710$, $\mathrm{SE} = 0.054$) compared to the continuous condition ($\mathrm{M} = 0.56$, $\mathrm{SE} = 0.055$), $F(1,419) = 12.51$, $p < 0.001$. While the main effect of age was not significant [$F(1, 37) = 1.03$, $p = 0.316$]; the interaction between age and condition was significant, [$F(1, 419) = 5.38$, $p = 0.021$]. Examination of this interaction revealed that young adults reported greater stream segregation in the discontinuous than in the continuous condition [$F(1, 226) = 8.32$, $p = 0.004$], whereas older adults' subjective judgment was not found to differ between the two conditions, $F(1, 238) < 1$.

3. Correlations

Bivariate correlations between QuickSIN score, d', the probability of hearing two streams ("streaming"), and mean audiometric thresholds from 250 to 2000 Hz were examined to explore the relationship between the perceptual organization of speech sounds, hearing sensitivity, and speech-in-noise comprehension. First, we observed a significant positive relationship between mean audiometric thresholds and QuickSIN, $r = 0.504$, $p = 0.001$. There was also a significant negative relationship between mean audiometric threshold and d', $r = -0.445$, $p = 0.005$. The correlation between mean audiometric threshold and streaming was not significant, $r = -0.026$, $p = 0.874$. Moreover, we also observed a significant negative relationship between QuickSIN score and d', $r = -0.421$, $p = 0.008$. The correlation between QuickSIN and the subjective measure of streaming was not significant ($r = 0.07$, $p = 0.652$), nor was the correlation between d' and streaming, $r = -0.147$, $p = 0.373$.

Given that mean audiometric thresholds were correlated with QuickSIN and d', we used a partial correlation to test for a linear relationship between these latter two variables while controlling for audiometric thresholds. The coefficient for the association between QuickSIN and d' became smaller following the partial correlation analysis ($r_p = -0.255$, $p = 0.123$), and was no longer significant, suggesting that this relationship was partly driven by audiometric thresholds. The associations between QuickSIN and streaming ($r_p = 0.102$, $p = 0.543$), and streaming and d' ($r_p = -0.177$, $p = 0.289$) were only modestly affected when taking into account mean audiometric thresholds in the partial correlation analysis.

C. Discussion

Young and older adults were less accurate in identifying the order of the four vowels when the f_1 frequency alternated and the difference between adjacent vowels was large, that is, during conditions which promote hearing multiple auditory steams. Our findings replicate and extend those of prior studies (Dorman *et al.*, 1975; Nooteboom *et al.*, 1978; Gaudrain *et al.*, 2008) by showing a similar effect of vowel frequency difference for naturally spoken vowels from four different talkers. The role of f_1 continuity in the perceptual organization of speech is further supported by past literature that posited formant transitions carry phonetic information (Dorman *et al.*, 1975) and bind together phonetic segments so that the temporal order of speech is preserved at rapid rates of transmission (Cole and Scott, 1973; Dorman *et al.*, 1975). In the present study, listeners' performance was little affected by the speaker voice, demonstrating that perceptual grouping based on f_1 is not talker specific nor is it modulated by the overall height of voice pitch, as there was no difference in f_0 between our male and female talker stimuli.

The effects of f_1 continuity on accuracy were greater in young than in older adults. This is consistent with the hypothesis that age impairs listeners' ability to sequentially group speech sounds based on f_1 and exploit differences in time-varying acoustic information to facilitate speech recognition. This age difference in accuracy was accompanied by changes in the probability of reporting hearing two concurrent streams, with older adults consistently reporting two concurrent streams of sounds independent of condition. Our results imply that older adults may have difficulty tracking and monitoring changes in f_1 over time. The objective measure of streaming used in the present study supports the hypothesis that perceiving two streams of sound makes the identification of vowel-order more difficult than when perceiving one stream of sound (Bregman and Campbell, 1971; Fullgrabe and Moore, 2012). The fact that older adults often reported more than one stream for the continuous condition suggests a problem in sequential integration and/or a tendency toward hearing multiple streams. Further research is needed to better understand the link between objective and subjective measures of streaming.

Overall, older adults were less accurate than younger adults in identifying the correct order of a given vowel sequence. This finding is consistent with literature demonstrating impairment in temporal order processing in older adults (Trainor and Trehub, 1989; Gordon-Salant and Fitzgibbons, 1999; Szymaszek *et al.*, 2009; Ulbrich *et al.*, 2009; Fogerty *et al.*, 2010). There is evidence that this age-related decline in temporal order judgment may reflect impaired auditory sensory memory (Alain and Woods, 1999; Fogerty *et al.*, 2010) and/or a general decline in attention (e.g., Fogerty *et al.*, 2010) or memory (Ulbrich *et al.*, 2009).

Using regression analyses, Ulbrich *et al.* (2009) showed that performance on temporal order tasks, similar to the one used in the present study, can be predicted by cognitive factors as measured by standardized tests of reasoning, short-term memory, and attention. Consequently, the findings from the temporal judgment task may be partly confounded with an age-related decline in cognitive functions, which may have exacerbated or interacted with age-related changes in the perceptual judgment of sounds. Hence, it is important to further investigate whether age impairs the perceptual organization of speech sounds using measures that are less sensitive to attention and memory.

One way to more objectively measure stream segregation while minimizing attentional and memory demands is to use a task that requires the detection of a change in temporal coherence or rhythm in the sequence (Vliegen *et al.*, 1999; Cusack and Roberts, 2000; Boehnke and Phillips, 2005; Micheyl and Oxenham, 2010; Fullgrabe and Moore, 2012; Richards *et al.*, 2012). In such a task, the stimulus onset asynchrony (SOA) between stimuli is increased or reduced, thereby inducing a disruption in the ongoing rhythm. The main finding from these types of paradigms is that changes in SOA are easier to detect when successive sounds are grouped together than when they are segregated into separate streams. This paradigm measures primitive/obligatory stream segregation (Fullgrabe and Moore, 2012). That is, lower thresholds are typically associated with sequential integration, while higher thresholds are indicative of stream segregation. In a second experiment, we used a subset of sequences from Experiment 1 and introduced a delay in SOA between successive vowels, thereby disrupting the ongoing rhythm of the sequence. An adaptive procedure was utilized to measure listeners' thresholds in detecting a temporal irregularity in the rhythm of the vowel stream. We hypothesized that thresholds for detecting asynchrony within the stimulus sequence would be lower in the continuous than in the discontinuous condition. We also anticipated, based on Experiment 1, that older adults would show deficits in using f_1 frequency between successive vowels; as a consequence, we anticipated that thresholds for detecting a temporal irregularity in the rhythm of a vowel stream would be little affected by the sequence type in older adults.

III. EXPERIMENT 2

A. Methods and materials

1. *Participants*

Twenty-two young adults and 25 older adults (who did not participate in Experiment 1) provided written informed consent according to the guidelines of the Baycrest Center and the University of Toronto. Participants' hearing abilities were assessed as in Experiment 1 and three older adults were excluded because they showed mild hearing loss. Five young and five older adults were excluded from data analysis because they had two or more anisochrony detection threshold values exceeding 100 ms. That is, they could not reliably detect a change in temporal irregularity in the rhythm of the vowel stream even for the largest values used at the beginning of the block of trials. The final sample comprised 17 young (age range = 19 to 35 yr; M = 25.29, SD = 4.34; 9 females), and 17 older adults (age range = 65 to 74 yr; M = 68.65, SD = 2.89; 10 females). All participants were non-musicians. All older adults scored 28 or higher on the Mini-Mental State Exam (Folstein *et al.*, 1975).

Table III shows the mean audiometric thresholds for young and older adults averaged across ears. For octave frequencies between 250 and 2000 Hz, older adults had higher pure tone thresholds than young adults [$F(1, 32) = 69.489$, $p < 0.001$, $\eta_p^2 = 0.685$]. There were no significant differences between the two ears [$F(1,32) < 1$], nor was the interaction between age and ear significant [$F(1, 32) < 1$].

Speech reception thresholds were again assessed using the QuickSIN test. As in Experiment 1, the SNR loss was higher (i.e., poorer) in older adults (M = 1.588, SE = 0.389) than in younger adults (M = −0.3476, SE = 0.132), $F(1, 32) = 22.234$, $p < 0.001$.

2. *Stimuli*

Two combinations of the four vowels sequences from Experiment 1were used, namely a continuous sequence, *ee-oo-er-ae*, and a discontinuous sequence, *ee-er-oo-ae*. We chose *ee-oo-er-ae* as the exemplar continuous sequence because its f_1 frequencies rose continuously between successive vowels, and overall it was most often reported as "one stream" by listeners in Experiment 1. We chose *ee-er-oo-ae* as the exemplar discontinuous sequence because its f_1 alternated between high and low frequencies between successive vowels, and because it was the sequence most often heard as "two streams" in Experiment 1.

As in Experiment 1, the vowel "ee" was arbitrarily selected as the first vowel in each sequence and the ISI was kept at 12 ms. Each trial included eight repetitions of one of the four-vowel sequences listed above, and lasted 6.78 s. Unlike Experiment 1, these sequences did not have rise and fall times (i.e., onset/offset fades), as the experimental task did not require participants to temporally order the vowel sounds (and thus, it was irrelevant whether or not participants could easily hear and/or identify the first vowel in the sequence). The sequence duration was shortened in Experiment 2 to increase the number of observation per participant, in order to obtain reliable threshold measurements within a reasonable amount time, and avoid fatigue and/or decline in motivation that may occur during a longer testing session.

TABLE III. Experiment 2: Mean and standard deviation of audiometric thresholds (dB HL) averaged over the left and the right ears for young and older adults.

	250 Hz	500 Hz	1000 Hz	2000 Hz	4000 Hz	8000 Hz
Young adults	7.79	6.62	2.50	1.77	−4.12	−1.47
Standard deviation	5.80	3.64	5.81	5.21	4.84	6.91
Older adults	17.06	14.71	12.06	15.85	28.24	47.21
Standard deviation	4.46	4.92	5.92	7.60	15.319	22.03

3. *Procedure*

The experimental setup was similar to Experiment 1. Each participant was first familiarized with the stimuli and listening task. Following a brief familiarization phase, thresholds for detecting rhythmic irregularity in the vowel sequence were measured for each participant using a single-interval "yes"/"no" paradigm. On each trial, listeners heard a sequence that comprised eight repetitions of the four-vowel pattern. The rhythmic regularity was disrupted by inserting an additional silence between the second and third vowel of the four-vowel pattern thereby increasing the SOA. The disruption in SOA occurred after at least three unaltered repetitions of the four-vowel pattern so as to establish the rhythm in the sequence. After that point, the disruption in SOA could occur in any of the subsequent four-vowel repetitions in the trial. The participants' task was to identify whether or not the sequence sounded "regular" (i.e., isochronous) or "irregular" (i.e., anisochronous).

Detection thresholds were measured using a 2-down, 1-up adaptive tracking paradigm targeting 71% correct performance on the psychometric function (Levitt, 1971). On each run, the initial silence (i.e., gap) was set at 104 ms, which was well above the expected thresholds based on prior studies using a similar experimental design (Vliegen *et al.*, 1999; Cusack and Roberts, 2000; Boehnke and Phillips, 2005; Micheyl and Oxenham, 2010; Richards *et al.*, 2012). Following two consecutive correct responses, the gap duration was decreased for the subsequent trial (i.e., made more difficult) and increased following a single incorrect response (i.e., made easier). Gap duration was varied using a geometric step size factor of two for the first four reversals and then $\sqrt{2}$ thereafter. Fourteen reversals were measured and the mean of the last six were used to compute each individual's temporal threshold for the run. There were no catch trials. A total of three detection thresholds were obtained for each type of streaming condition (continuous vs discontinuous) per listener. The order of the streaming condition was counterbalanced between participants.

4. *Statistical Analyses*

The three threshold runs per sequence type were each averaged, to form a single score per condition. Descriptive analysis of the threshold values revealed a skewed distribution, which was adjusted by applying a log transformation to the data to normalize the distribution prior to employing parametric tests. A 2×2 repeated-measures ANOVA was then conducted, comparing young and older adults' thresholds for the two sequence types.

B. Results

1. *Effects of age and sequence type*

Figure 4 shows the group mean thresholds as a function of sequence type. The main effect of age was not significant [$F(1,32) = 1.001$, $p = 0.325$], nor was the main effect of sequence type, $F(1, 32) = 1.092$, $p = 0.304$. However, there was a significant group x sequence type interaction, $F(1, 32) = 5.339$, $p = 0.027$, $\eta_p^2 = 0.143$. While young adults were better at detecting rhythmic irregularities in the continuous than in the discontinuous condition [$F(1, 16) = 7.269$, $p = 0.016$, $\eta_p^2 = 0.312$], older adults showed comparable thresholds in both continuous and discontinuous cases [$F(1,16) < 1$]. While comparing the effect of age as a function of condition, we found lower thresholds in young ($M = 31.810$ ms, $SE = 4.438$) than in older ($M = 44.931$ ms, $SE = 4.928$) adults for the continuous condition [$F(1,32) = 4.579$, $p = 0.040$, $\eta_p^2 = 0.125$]. No age differences were observed for thresholds obtained from the discontinuous sequences [$F(1,32) < 1$].

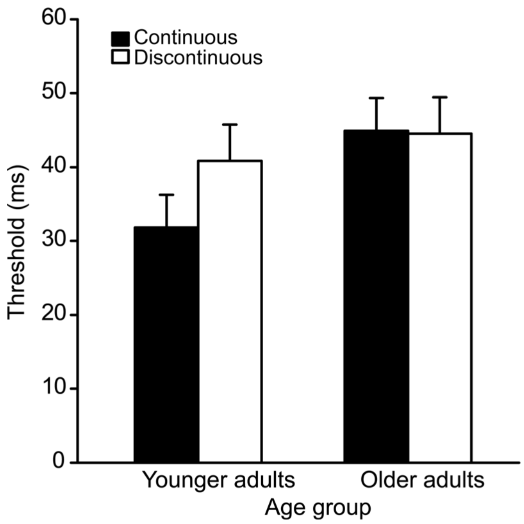

FIG. 4. Mean temporal thresholds (ms) for detecting rhythmic irregularity in the speech stream. Error bars show standard error of the mean.

C. Discussion

Young adults were worse (i.e., higher thresholds) in detecting a change in rhythm within vowel sequences with larger, alternating f_1 between consecutive vowels. This finding is consistent with the proposal that f_1 continuity helps promote stream segregation. Smaller differences in f_1 across tokens (continuous condition) seemed to increase the perceptual fusion of speech tokens into a single auditory stream thereby facilitating the detection of anisynchrony within the sequence. In contrast, disjointed cues between adjacent speech sounds (discontinuous condition) promote multiple streams and hence, hinder one's ability to monitor changes in the ongoing sequence (Fig. 4). Our results are also consistent with previous studies showing that, indeed, stream segregation does impede the ability to detect a rhythmic irregularity in an otherwise isochronous sequence (Vliegen *et al.*, 1999; Cusack and Roberts, 2000; Boehnke and Phillips, 2005; Micheyl and Oxenham, 2010; Richards *et al.*, 2012) and that it is more difficult to process the timing of auditory stimuli that are perceptually separated into more than one stream. Our findings also extend previous work by showing that objective measures, such as the rhythmic judgment task used here, are sensitive enough to assess age-related effects and the role of relatively subtle cues in the perceptual organization of speech sounds.

In older adults, the ability to detect changes in tempo was little affected by f_1 continuity. This finding is consistent with the notion that older adults have difficulty using f_1 over time in grouping speech sounds. Age differences in temporal thresholds were specific to the continuous condition. This suggests that older adults have difficulty integrating the sequence as a whole and appear to automatically segregate the speech stimuli into different streams. This seems consistent with both accuracy and subjective reports from Experiment 1, which revealed greater age difference for the continuous than the discontinuous conditions. Together with the age difference in thresholds observed in Experiment 2, this age effect in accuracy and subjective report provides converging evidence for age-related deficits in processing sequential first-formant cues.

That said, it is possible that the absence of a streaming effect (difference between continuous and discontinuous conditions) in the older group could be related to age difference in the build-up period, with older adults needing more time to perceive streaming (Snyder and Alain, 2007a). Moreover, it is important to note that the averaged thresholds reported in Experiment 2 are relatively elevated in young and older adults even for the conditions thought to promote integration rather than segregation. Indeed, the level of performances observed in young adults for the continuous conditions is more similar with the thresholds that are typically reported in the literature for segregated conditions. For example, for rhythmic discrimination with sequences of vowels, Devergie *et al.* (2011) reported thresholds around 16 ms for integrated conditions and around 28 ms for segregated conditions. While some methodological differences could explain this apparent discrepancy, it remains possible that the low level of performances in Experiment 2 could be partly accounted for by the fact that the vowel sequences were at least partially segregated even in the condition designed to promote integration. The subjective reports in Experiment 1, where the lowest proportion of hearing two streams always remains fairly high (around 50%), is consistent with this hypothesis. In Experiment 1 and 2, the transition between successive, naturally spoken vowels was not smooth but rather included some discontinuity, which could have promoted the segregation of one or two vowels in a separate stream.

IV. GENERAL DISCUSSION

The aim of this study was to assess whether age impairs a listener's ability to integrate/segregate speech sounds across time. This work was motivated by the fact that prior research on aging and sequential integration using pure tone stimuli have found little evidence for an age-related decline in stream segregation abilities (Trainor and Trehub, 1989; Alain *et al.*, 1996; Snyder and Alain, 2007a). Yet, somewhat contradictorily, other studies that have focused on concurrent speech segregation of single speech tokens (Snyder and Alain, 2005; Vongpaisal and Pichora-Fuller, 2007) or speech-in-noise (e.g., Pichora-Fuller *et al.*, 1995; George *et al.*, 2007; Anderson Gosselin and Gagne, 2011) have consistently observed an age-related decline in performance, suggesting that aging may impair the perceptual organization of speech sounds. Here, we showed in two separate experiments, using different participants and methodology, that older adults have difficulties in processing f_1 continuity, which has been shown to play an important role in the perceptual organization of speech sounds. The results of the current study collectively suggest a relationship between aging and the sequential streaming of speech.

In the present study, perceptual grouping of speech sounds was promoted by manipulating the order of four vowels according to their first-formant frequency. The f_1 difference between successive vowels was relatively small in comparison to the more typical frequency difference used in pure tone, "ABA"-like sequences (e.g., Trainor and Trehub, 1989; Alain *et al.*, 1996; Snyder and Alain, 2007a). Evidence suggests that older adults are particularly at a disadvantage in tasks that involve processing of the temporal-fine-structure, which are independent of peripheral hearing status (Moore *et al.*, 2012). Hence, the absence of an f_1-based streaming effect in the older adults could be a consequence of an age-dependent change in temporal-fine-structure processing as opposed to deficits in streaming per se. In other words, it remains to be determined whether the age difference in the perceptual organization of speech sounds is due to lower sensitivity to first-formant difference and/or a general deficit in the "streaming mechanism." Future studies may help determine whether the age difference is related to the material complexity or the magnitude/salience of the cue used to induce stream segregation. Indeed, it is possible that age-related effects on sequential auditory streaming may only manifest when considering the more subtle, yet equally critical, cues of the speech signal.

In the current study, we did not find an association between performance during the speech-in-noise test and either objective or subjective measures of stream segregation. This result differs from those reported by Mackersie *et al.* (2001), who found a significant correspondence between streaming judgment and simultaneous sentence perception. It is notable that Mackersie *et al.* (2001) used a broader age range for their young and older adults, as well as a broader range of hearing ability as measured with pure-tone thresholds. In other words, their sample of young and older adults was more heterogeneous than the one used in the present study. Indeed, our samples of younger and older adults were fairly homogenous with respect to hearing ability and age, thereby reducing the variability in responses and likely our ability to observe non-zero correlations between tasks. Our findings also differ from those of Gaudrain *et al.* (2012), who observed a significant correlation between listeners' performance in an order-naming task on vowel sequence and their ability to identify monosyllabic words embedded in time reverse speech from a single talker. There are several factors that could account for this discrepancy. These include the method used to infer streaming and the task used to assess speech-in-noise reception and comprehension. In the present study, the lack of relationship between QuickSIN score and the objective or subjective responses suggests that different perceptual and/or cognitive processes were engaged during the experimental tasks and the speech-in-noise test.

For example, the QuickSIN may rely more on cognitive (rather than perceptual) processes, such as attention and working memory, while the objective and subjective measures used in the present study are more perceptually based. Future studies could incorporate measures of attention and working memory to explore this relationship further.

In both experiments, the continuous and discontinuous conditions differed along two dimensions: the direction of the first-formant progression (monotonically increasing or decreasing vs up-down-up-down) and the amount of first-formant change between successive vowels. Consequently, it remains to be determined whether the difference between sequence types was primarily due to the alternating low and high f_1 frequency or the actual magnitude of changes in f_1 frequency between successive vowels. Further research is needed to assess the contribution of each factor separately.

In conclusion, the findings from the present study suggest that older adults are impaired in their ability to perceptually organize sequential speech sounds. Our findings with natural vowels contrast with earlier studies that previously indicated little age-related difference in stream segregation using simpler stimuli and paradigms (pure-tone ABA-sequences). This study thus highlights the importance of using realistic stimuli when studying sequential streaming. Continuing this line of research will help to build an integrated theory of auditory scene analysis and cognitive aging as related to sequential streaming. To this aim, future studies could examine analogues between the ABA-paradigm and those involving naturalistic vowel sounds to understand the interesting relationship between aging, sequential auditory streaming, and the perception of speech.

ACKNOWLEDGMENTS

The research was supported by grants from the Canadian Institutes of Health Research, the Natural Sciences and Engineering Research Council of Canada (NSERC), and NSERC-Create: Training in Auditory Cognitive Neuroscience. We would like to thank Yu He for technical assistance, Claire Salloum for careful reading and suggestions of earlier versions, and Nicolas Grimault for suggesting the rhythmic task.

Alain, C., and Arnott, S. R. (**2000**). "Selectively attending to auditory objects," Front. Biosci. **5**, D202-212.

Alain, C., Dyson, B. J., and Snyder, J. S. (**2006**). "Aging and the perceptual organization of sounds: a change of scene?," in *Handbook of Models for the Study of Human Aging*, edited by M. Conn (Elsevier, Amsterdam), pp. 759–769.

Alain, C., McDonald, K. L., Ostroff, J. M., and Schneider, B. (**2001**). "Age-related changes in detecting a mistuned harmonic," J. Acoust. Soc. Am. **109**, 2211–2216.

Alain, C., Ogawa, K. H., and Woods, D. L. (**1996**). "Aging and the segregation of auditory stimulus sequences," J. Gerontol. B Psychol. Sci. Soc. Sci. **51**, P91–P93.

Alain, C., and Woods, D. L. (**1999**). "Age-related changes in processing auditory stimuli during visual attention: evidence for deficits in inhibitory control and sensory memory," Psychol. Aging **14**, 507–519.

Anderson Gosselin, P., and Gagne, J. P. (**2011**). "Older adults expend more listening effort than young adults recognizing speech in noise," J. Speech Lang. Hear. Res. **54**, 944–958.

Assmann, P. F., and Summerfield, Q. (**1994**). "The contribution of waveform interactions to the perception of concurrent vowels," J. Acoust. Soc. Am. **95**, 471–484.

Boehnke, S. E., and Phillips, D. P. (**2005**). "The relation between auditory temporal interval processing and sequential stream segregation examined with stimulus laterality differences," Percept. Psychophys. **67**, 1088–1101.

Bregman, A. S., and Campbell, J. (**1971**). "Primary auditory stream segregation and perception of order in rapid sequences of tones," J. Exp. Psychol. **89**, 244–249.

Cole, R. A., and Scott, B. (**1973**). "Perception of temporal order in speech: the role of vowel transitions," Can. J. Psychol. **27**, 441–449.

Culling, J. F., and Darwin, C. J. (**1993**). "Perceptual separation of simultaneous vowels: within and across-formant grouping by F0," J. Acoust. Soc. Am. **93**, 3454–3467.

Cusack, R., and Roberts, B. (**2000**). "Effects of differences in timbre on sequential grouping," Percept. Psychophys. **62**, 1112–1120.

Devergie, A., Grimault, N., Gaudrain, E., Healy, E. W., and Berthommier, F. (**2011**). "The effect of lip-reading on primary stream segregation," J. Acoust. Soc. Am. **130**, 283–291.

Dorman, M. F., Cutting, J. E., and Raphael, L. J. (**1975**). "Perception of temporal order in vowel sequences with and without formant transitions," J. Exp. Psychol. Hum. Percept. Perform. **104**, 147–153.

Duquesnoy, A. J. (**1983a**). "Effect of a single interfering noise or speech source upon the binaural sentence intelligibility of aged persons," J. Acoust. Soc. Am. **74**, 739–743.

Duquesnoy, A. J. (**1983b**). "The intelligibility of sentences in quiet and in noise in aged listeners," J. Acoust. Soc. Am. **74**, 1136–1144.

Fogerty, D., Humes, L. E., and Kewley-Port, D. (**2010**). "Auditory temporal-order processing of vowel sequences by young and elderly listeners," J. Acoust. Soc. Am. **127**, 2509–2520.

Folstein, M. F., Folstein, S. E., and McHugh, P. R. (**1975**). " 'Mini-mental state.' A practical method for grading the cognitive state of patients for the clinician," J. Psychiatr. Res. **12**, 189–198.

Fullgrabe, C., and Moore, B. C. (**2012**). "Objective and subjective measures of pure-tone stream segregation based on interaural time differences," Hear. Res. **291**, 24–33.

Gaudrain, E., Grimault, N., Healy, E. W., and Bera, J. C. (**2008**). "Streaming of vowel sequences based on fundamental frequency in a cochlear-implant simulation," J. Acoust. Soc. Am. **124**, 3076–3087.

Gaudrain, E., Grimault, N., Healy, E. W., and Bera, J. C. (**2012**). "The relationship between concurrent speech segregation, pitch-based streaming of vowel sequences, and frequency selectivity," Acta Acust. Acust. **98**, 317–327.

George, E. L., Zekveld, A. A., Kramer, S. E., Goverts, S. T., Festen, J. M., and Houtgast, T. (**2007**). "Auditory and nonauditory factors affecting speech reception in noise by older listeners," J. Acoust. Soc. Am. **121**, 2362–2375.

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

Gordon-Salant, S., and Fitzgibbons, P. J. (**1997**). "Selected cognitive factors and speech recognition performance among young and elderly listeners," J. Speech Lang. Hear. Res. **40**, 423–431.

Gordon-Salant, S., and Fitzgibbons, P. J. (**1999**). "Profile of auditory temporal processing in older listeners," J. Speech Lang. Hear. Res. **42**, 300–311.

Grimault, N., Micheyl, C., Carlyon, R. P., Arthaud, P., and Collet, L. (**2001**). "Perceptual auditory stream segregation of sequences of complex sounds in subjects with normal and impaired hearing," Br. J. Audiol. **35**, 173–182.

Grube, M., von Cramon, D. Y., and Rubsamen, R. (**2003**). "Inharmonicity detection. Effects of age and contralateral distractor sounds," Exp. Brain Res. **153**, 637–642.

Hasher, L., and Zacks, R. T. (**1984**). "Automatic processing of fundamental information: the case of frequency of occurrence," Am. Psychol. **39**, 1372–1388.

Helfer, K. S., and Freyman, R. L. (**2008**). "Aging and speech-on-speech masking," Ear. Hear. **29**, 87–98.

Humes, L. E. (**1996**). "Speech understanding in the elderly," J. Am. Acad. Audiol. **7**, 161–167.

Humes, L. E., and Roberts, L. (**1990**). "Speech-recognition difficulties of the hearing-impaired elderly: the contributions of audibility," J. Speech Hear. Res. **33**, 726–735.

Killion, M. C., Niquette, P. A., Gudmundsen, G. I., Revit, L. J., and Banerjee, S. (**2004**). "Development of a quick speech-in-noise test for measuring signal-to-noise ratio loss in normal-hearing and hearing-impaired listeners," J. Acoust. Soc. Am. **116**, 2395–2405.

Levitt, H. (**1971**). "Transformed up-down methods in psychoacoustics," J. Acoust. Soc. Am. **49** Suppl. 2, 467–477.

Mackersie, C. L., Prida, T. L., and Stiles, D. (**2001**). "The role of sequential stream segregation and frequency selectivity in the perception of simultaneous sentences by listeners with sensorineural hearing loss," J. Speech Lang. Hear. Res. **44**, 19–28.

Micheyl, C., and Oxenham, A. J. (**2010**). "Objective and subjective psychophysical measures of auditory stream integration and segregation," J. Assoc. Res. Otolaryngol. **11**, 709–724.

Middelweerd, M. J., Festen, J. M., and Plomp, R. (**1990**). "Difficulties with speech intelligibility in noise in spite of a normal pure-tone audiogram," Audiology **29**, 1–7.

Moore, B. C., Glasberg, B. R., Stoev, M., Fullgrabe, C., and Hopkins, K. (**2012**). "The influence of age and high-frequency hearing loss on sensitivity to temporal fine structure at low frequencies (L)," J. Acoust. Soc. Am. **131**, 1003–1006.

Moore, B. C., and Gockel, H. (**2002**). "Factors influencing sequential stream segregation," Acta Acust. Acust. **88**, 320–333.

Nooteboom, S. G., Brokx, J. P. L., and de Rooij, J. J. (**1978**). "Contributions of prosody to speech perception," in *Studies in the Perception of Language*, edited by W. J. M. Levelt and G. B. F. d'Arcais (Wiley, New York), pp. 75–107.

Pichora-Fuller, M. K., Schneider, B. A., and Daneman, M. (**1995**). "How young and old adults listen to and remember speech in noise," J. Acoust. Soc. Am. **97**, 593–608.

Richards, V. M., Carreira, E. M., and Shen, Y. (**2012**). "Toward an objective measure for a 'stream segregation' task," J. Acoust. Soc. Am. **131**, EL8–13.

Salthouse, T. A. (**1996**). "The processing-speed theory of adult age differences in cognition," Psychol. Rev. **103**, 403–428.

Schneider, B. A., Daneman, M., and Murphy, D. R. (**2005**). "Speech comprehension difficulties in older adults: cognitive slowing or age-related changes in hearing?," Psychol. Aging **20**, 261–271.

Schneider, B. A., Li, L., and Daneman, M. (**2007**). "How competing speech interferes with speech comprehension in everyday listening situations," J. Am. Acad. Audiol. **18**, 559–572.

Schneider, B. A., Pichora-Fuller, M. K., and Daneman, M. (**2010**). "Effects of senescent changes in audition and cognition on spoken language comprehension," in *Handbook of Auditory Research: The Aging Auditory System*, edited by S. Gordon-Salant, R. D. Frisina, A. N. Popper, and R. R. Fay (Springer, New York), pp. 167–210.

Snyder, J. S., and Alain, C. (**2005**). "Age-related changes in neural activity associated with concurrent vowel segregation," Brain Res. Cognit. Brain Res. **24**, 492–499.

Snyder, J. S., and Alain, C. (**2007a**). "Sequential auditory scene analysis is preserved in normal aging adults," Cereb. Cortex **17**, 501–512.

Snyder, J. S., and Alain, C. (**2007b**). "Toward a neurophysiological theory of auditory stream segregation," Psychol. Bull. **133**, 780–799.

Summerfield, Q., and Assmann, P. F. (**1989**). "Auditory enhancement and the perception of concurrent vowels," Percept. Psychophys. **45**, 529–536.

Szymaszek, A., Sereda, M., Poppel, E., and Szelag, E. (**2009**). "Individual differences in the perception of temporal order: the effect of age and cognition," Cognit. Neuropsychol. **26**, 135–147.

Trainor, L. J., and Trehub, S. E. (**1989**). "Aging and auditory temporal sequencing: ordering the elements of repeating tone patterns," Percept. Psychophys. **45**, 417–426.

Ulbrich, P., Churan, J., Fink, M., and Wittmann, M. (**2009**). "Perception of temporal order: the effects of age, sex, and cognitive factors," Neuropsychol. Dev. Cognit. B Aging Neuropsychol. Cognit. **16**, 183–202.

Vliegen, J., Moore, B. C., and Oxenham, A. J. (**1999**). "The role of spectral and periodicity cues in auditory stream segregation, measured using a temporal discrimination task," J. Acoust. Soc. Am. **106**, 938–945.

Vongpaisal, T., and Pichora-Fuller, M. K. (**2007**). "Effect of age on F0 difference limen and concurrent vowel identification," J. Speech Lang. Hear. Res. **50**, 1139–1156.

Warren, R. M., Healy, E. W., and Chalikia, M. H. (**1996**). "The vowel-sequence illusion: intrasubject stability and intersubject agreement of syllabic forms," J. Acoust. Soc. Am. **100**, 2452–2461.

Zendel, B. R., and Alain, C. (**2012**). "The Influence of lifelong musicianship on neurophysiological measures of concurrent sound segregation," J. Cognit. Neurosci. **25**, 503–516.