Age-related Differences in the Perceptual Organization of Speech Sounds

by

Stefanie Andrea Hutka

A thesis submitted in conformity with the requirements for the degree of Masters of Arts
Graduate Department of Psychology
University of Toronto

© Copyright by Stefanie A. Hutka 2011
Age-related Differences in the Perceptual Organization of Speech Sounds

Stefanie Andrea Hutka

Master of Arts

Graduate Department of Psychology
University of Toronto

2011

Abstract

Aging is associated with a decline in the ability to understand what a person is saying in the presence of other sounds. This study investigated the perceptual organization of speech in young (n=20) and older adults (n=20). Four vowels were arranged into six sequences, defined by either continuous or discontinuous first-formant transitions. Participants first made an objective response (choosing the sequence that best matched the one they just heard from a list), followed by a subjective response (indicating if they heard one or two streams of sound). There were significant interactions between age and sequence-type for both objective and subjective responses, respectively. These results suggest that aging affects the ability to perceptually organize speech-sounds and the ability to perceive sequential streaming of speech. These findings are discussed within the context of further enriching what is known about auditory scene analysis, cognitive aging, and sequential streaming.
Acknowledgment

I would like to thank my supervisor, Dr. Claude Alain, for his support and guidance in the experimental design; Dr. Malcolm Binns, for his guidance in data analysis, and Yu He for technical assistance. I would also like to thank the Natural Sciences and Engineering Research Council of Canada (NSERC) and the Canadian Institutes of Health Research (CIHR) for funding this research.
# Table of Contents

Abstract ........................................................................................................................................ ii
Acknowledgments ......................................................................................................................... iii
Table of Contents .......................................................................................................................... iv
List of Tables ................................................................................................................................. vi
List of Figures ............................................................................................................................... vii

**Chapter 1 Background Literature** .......................................................................................... 1
1 Aging and Auditory Scene Analysis ....................................................................................... 1
   1.1 Aging and the Cocktail Party Problem ............................................................................ 1
   1.2 Concurrent and Sequential Segregation: Definitions and Age-related Differences ....... 2
2 Perceptual Organization of Speech ......................................................................................... 5
3 Purpose, Hypotheses, and Predictions .................................................................................... 7

**Chapter 2** ................................................................................................................................ 10
4 Methods and Materials ............................................................................................................. 10
   4.1 Participants ....................................................................................................................... 10
   4.2 Stimuli ............................................................................................................................... 11
   4.3 Procedure ............................................................................................................................ 13
      4.3.1 Hearing Assessments ............................................................................................... 14
         4.3.1.1 Mistuned-Harmonic Detection Threshold ......................................................... 14
         4.3.1.2 Gap-Detection Threshold ................................................................................ 14
         4.3.1.3 Speech-In-Noise Test .................................................................................... 15
      4.3.2 Statistical Analyses .................................................................................................. 16

**Chapter 3** ................................................................................................................................ 21
5 Results ....................................................................................................................................... 21
   5.1 Preliminary Analyses ......................................................................................................... 21
   5.2 Hearing-Assessments ........................................................................................................ 21
   5.3 Primary Analyses .............................................................................................................. 24
   5.4 Correlations ..................................................................................................................... 26

**Chapter 4** ................................................................................................................................ 30
6 Discussion .................................................................................................................................. 30
7 General Discussion ................................................................................................................... 31
8 Future Directions and Conclusions...............................................................36
References........................................................................................................37
List of Tables

Table 1. Frequencies (Hz) for each vowel for each speaker.

Table 2. Partial correlations between d’ and hearing tests, controlling for age.

Table 3. Partial correlations between the proportion of hearing two streams and hearing tests, controlling for age.

Table 4. Correlations between d’ and hearing tests for young adults only.

Table 5. Correlations between d’ and hearing tests for older adults only.

Table 6. Correlations between streaming judgment and hearing tests for young adults.

Table 7. Correlations between streaming judgment and hearing tests for older adults.
List of Figures

Figure 1. Frequency profiles for discontinuous sequences at $f_0$ (green triangle), the first formant (blue lozenge), and the second formant (red square).

Figure 2. Frequency profiles for continuous sequences at $f_0$ (green triangle), the first formant (blue lozenge), and the second formant (red square).

Figure 3. Audiometric threshold for young and older adults.

Figure 4. Group mean mistuned-harmonic thresholds; unit for the y-axis is the amount of mistuning (Hz).

Figure 5. Group mean gap-detection thresholds; unit for the y-axis is gap-duration (ms).

Figure 6. Group mean performance on the QuickSIN test; unit for the y-axis is SNR loss (dB).

Figure 7. Group mean sensitivity measure (d’) in young and older adults.

Figure 8. Group mean proportion of reporting hearing two streams in young and older adults.
Chapter 1
Background Literature

In this chapter, I will introduce the relationship between aging and auditory scene analysis. Specifically, I will address the relationship between aging and the cocktail party problem. This will be followed by an examination of the two components that comprise auditory scene analysis, namely the concurrent and sequential segregation of sounds. I will then apply this background to examining why speech is a unique stimulus in auditory scene analysis. Finally, I will segue into introducing the current study, including hypotheses and predictions.

1 Aging and Auditory Scene Analysis

1.1 Aging and the Cocktail Party Problem

Aging is associated with a decline in the ability to understand speech in environments in which several acoustic sources are present, such as a cocktail party (Cherry 1953; Committee on Hearing, Bioacoustics, and Biomechanics, 1988). Indeed, many studies have demonstrated an association between age and an impaired ability to segregate speech-in-noise (e.g., Plomp & Mimpen, 1979; Dubno, Dirks, & Morgan, 1984; Tun, 1998). Age-related changes in peripheral and central auditory processing, as well as changes in attention and memory functions, may partially account for the difficulties older adults face in adverse listening situations (Committee on Hearing, Bioacoustics, and Biomechanics, 1988; Schneider, Daneman, & Pichora-Fuller, 2002). Hearing problems hinder communication and may contribute to social isolation (Betlejewski, 2006), and have a profound negative effect on older adults’ autonomy, independence, and general well-being (Salomon, 1986, Betlejewski 2006). It is therefore very important to understand the mechanisms of these age-related changes in auditory perception, as related to the perceptual organization of speech sounds.

The cocktail party problem, first researched by Cherry (1953), poses the following question: How does a person perceive the voice of a particular speaker in an environment in which there are many other individuals simultaneously speaking? The cocktail party problem has often been studied by presenting listeners with two or more messages and asking them to shadow or pay attention to one of the messages in order to detect pre-defined targets (Bregman, 1990). Early studies using such paradigms have shown that observers can more readily follow a particular
speaker when the spoken messages emanate from speakers that differ in gender or location (e.g., left versus right ear) (Cherry, 1953). This acoustic difference likely promotes the segregation of the task-relevant message from the irrelevant one.

Indeed, subsequent research on the perceptual organization of sounds have shown that sounds that are more similar in timbre, pitch, and/or spatial location tend to be perceptually grouped together, compared to sounds that differ in these attributes (Bregman, 1990; Darwin & Carlyon, 1995). Similarly, sounds that move together (often referred to as sharing a common fate), as well as sounds that are continuous, tend to be perceptually grouped together (Bregman, 1990). This process, in which incoming auditory information is organized perceptually, is known as auditory scene analysis (ASA) (Bregman, 1990).

1.2 Concurrent and Sequential Segregation: Definitions and Age-Related Differences

The perceptual organization in ASA occurs along two axes: a frequency axis and a time axis (Alain, Dyson, & Snyder, 2006). In the former, different frequency and harmonic relations are segregated and analyzed on a moment-to-moment basis (Alain et al., 2006). This is often referred to as the segregation of concurrent sounds. An example is two simultaneously-spoken speech sounds that are perceptually segregated and identified (Bregman, 1990). Along the time axis, successive auditory events that occur over several seconds are grouped into one or more streams (Alain et al., 2006). This is often referred to as sequential segregation of successively-occurring sounds (also known as streaming). An example is distinguishing between two melodies with notes that alternate between two musical instruments (Bregman, 1990).

Streaming is often tested using an “ABA” paradigm, in which participants are presented with sequences of “ABA—“ in which A and B are sinusoidal tones of different frequencies, and the “—“ is a silent interval (Alain et al., 2006). Frequency separation between the A and B tones is manipulated to promote the perception of one “galloping” rhythm (e.g., ABA--, ABA--), or the perception of two distinct perceptual streams (A—A—and B—B—) (van Noorden, 1975). Listeners indicate when they start hearing the two tones as a single galloping rhythm, rather than two distinct streams (van Noorden, 1975). The fusion boundary is the frequency separation where participants perceive one stream of sound with a galloping rhythm, while the fission
boundary refers to when participants start to hear the sounds as coming from two distinct sources (van Noorden, 1975).

Alain et al. (2006) posited that older adults’ difficulty in understanding what one person is saying in the presence of background talkers is related to a failure to perceptually organize the mixture of sounds. For effective speech reception to occur, listeners must rely on the perceptual mechanisms that group together those sound elements coming from one source (i.e., one speaker), and segregate those arising from other sources (i.e., another speaker). As problems with speech perception primarily arise in complex listening environments in which there is more than one active sound-source, Alain et al. (2006) reasoned that aging may affect the low-level automatic processes that detect, group, and/or segregate sounds that are similar in physical attributes (e.g., frequency, intensity, and location) and/or higher-level schema-driven processes that reflect listeners’ experience and knowledge of the auditory environment. Though this hypothesis has generally been supported in recent literature, the effects of age on ASA may differ for the segregation of concurrent and sequential sounds.

Indeed, there is evidence to suggest that concurrent sound segregation declines with age (as reviewed in Snyder & Alain, 2007). For instance, older adults have greater difficulty than young adults in identifying two concurrently-presented vowels (Snyder & Alain, 2005). Older adults also have a higher threshold than young adults for detection of mistuning within a harmonic-complex (Alain, McDonald, Ostroff, & Schneider, 2001; Grube, von Cramon, & Rubsamen, 2003), and are more likely to assimilate one or more frequency components coming from secondary sound sources into a target signal (Roberts & Moore, 1990). This may occur because parts of extraneous frequencies may be included in a target signal at an early stage of processing, leading to errors that are paralleled in perception (Snyder & Alain, 2007). This suggests that age-related elevation of detection thresholds for mistuning may be connected to the difficulties older adults face when perceiving speech, as the ability to distinguish between auditory events based on a spectral pattern is common to both the mistuned harmonic task and the cocktail party problem (Snyder & Alain, 2007).

In contrast to these findings, several studies suggest that there are no age-related differences in streaming. Studies, such as Trainor and Trehub (1989) and Alain, Ogawa, and Woods (1996), have used indirect objective measures of sequential sound segregation, such as detecting a
change to a stimulus that is difficult without perceptual segregation, to study aging and sequential sound segregation. These studies, which both used non-verbal stimuli, yielded no evidence for an effect of aging on streaming.

As Snyder and Alain (2007) point out, indirect, objective measures – though beneficial in that they do not rely on self-report - do not always clearly relate to the subjective perception of two streams of sound. Stainsby, Moore, and Glasberg (2004) used both indirect (objective) and direct (subjective) measures of streaming, and found that older adults with hearing impairments were able to segregate sequential patterns of harmonic complexes using temporal cues. However, these participants’ performance was not compared to a control group, obscuring whether any impairments based on age, hearing loss, or neither, were at play.

Grimault, Micheyl, Carlyon, Arthaud, and Collet (2001) investigated the influence of hearing loss and aging on perceptual organization of sound sequences in young and older adults with normal hearing, and in older adults with hearing impairments. The ability to form perceptual auditory streams from sequences of harmonic complexes as a function of differences in fundamental frequency ($f_0$) was tested (Grimault et al., 2001). When the $f_0$s of the stimuli were so low that the harmonics could not be individually resolved, there were no differences between streaming performance in all participant-groups (Grimault et al., 2001). However, when the $f_0$s of the stimuli were high enough for the harmonics to be resolved, young adults perceived more streaming than older adults (Grimault et al., 2001). When the older adults were divided according to hearing impairment, those with impairment perceived more streaming for the resolved tones than did the older adults without impairment (Grimault et al., 2001). These findings suggest that streaming differences between young and older adults may simply be due to hearing loss - specifically, reduced peripheral frequency selectivity (Grimault et al., 2001). This conclusion is similar to that of Rose and Moore (1997), who demonstrated that listeners with a bilateral hearing-impairment required greater frequency separation to facilitate the streaming of sinusoidal tone-bursts, as compared to those with intact-hearing.

The studies above suggest that age has little impact on auditory stream segregation. However, Mackersie, Prida, and Stiles (2001) and Hong and Turner (2006) have demonstrated that there is an important relationship between age-related declines in perceiving speech-in-noise, and the sequential segregation of sound. Mackersie et al. (2001) investigated the role of frequency
selectivity and streaming in the perception of simultaneous sentences by younger and older adults, without and with hearing impairment, respectively. Even after partialling out the effects of age, Mackersie et al. (2001) found a strong relationship between the fusion threshold and simultaneous-sentence perception. Higher sentence scores were associated with smaller frequency differences at fusion thresholds, while there was no relationship between frequency selectivity and simultaneous sentence perception (Mackersie et al., 2001). In summary, there was a strong association between the ability to separate a mixture of two sentences into independent linguistic streams, and the ability to separate alternating tones into independent tonal streams (Mackersie et al., 2001).

Hong and Turner (2006) examined the ability of cochlear implant users (ages 39 to 78) and normal-hearing participants (ages 21 to 35) to perform streaming of pure tones. The hearing-intact participants’ auditory stream segregation increased with increasing frequency separation (Hong & Turner, 2006). In the cochlear implant users, variable pure-tone streaming performance was observed, and correlated with speech perception in noise (Hong & Turner, 2006). Overall, greater streaming was associated with better understanding of speech-in-noise, suggesting that auditory streaming is a contributing factor in the ability to understand speech in background noise (Hong & Turner, 2006). In addition, the findings of Hong and Turner (2006) demonstrate the link between stream segregation and everyday communication, as cochlear implant users who demonstrated improved streaming ability also had better perceptions of speech in background noise.

2 Perceptual Organization of Speech

The findings discussed earlier raise the question why age does not appear to impact the sequential segregation of speech sounds, when streaming appears to play an important role in the ability to understand speech amongst background noise - an ability that declines with age. Furthermore, why are there age-related differences found in literature on concurrent sound segregation and not in the sequential streaming literature, when both concurrent and sequential sound segregation are components of ASA?

These questions may be answered by examining a critical difference in the stimuli used in the concurrent and sequential streaming literature, respectively. In the concurrent-sound-segregation
literature discussed above, both pure tones and complex stimuli, such as vowel sounds, were used. In the sequential streaming literature, only pure tones or harmonic complexes have been used. This is a relevant distinction, as there is evidence that the perceptual organization of speech sounds may not always follow the same grouping principles as those used to commonly account for the perceptual organization of non-verbal material (Remez, Rubin, Berns, Pardo, and Lang, 1994).

Research on the basic principles of ASA, such as perceptual organization via pitch, timbre, location, and good continuation, have typically focused on the perception of relatively simple sounds (Darwin & Carlyon, 1995). Nonetheless, these principles have often been applied to the perceptual organization of speech – a very unique, familiar, and complex sound. For instance, good continuation cues provided by formant transitions between phonetic segments (Cole & Scott, 1973; Dorman, Cutting, & Raphael, 1975) and pitch contour (Darwin & Bethell-Fox, 1977) reduce the tendency to perceive streaming in a repeating sequence of syllables that consist of synthetic speech sounds (namely, consonant-vowel and vowel-vowel sequences). Differences in onset time and $f_0$ have been shown to influence the grouping of synthetic consonant-vowel syllables (e.g., Darwin, 1981). Brokx and Nootboom (1982) found that differences in $f_0$ between a target speaker’s voice and interfering speech improved intelligibility of the target. Bird and Darwin (1998) also found that this improvement in perception depends on across-formant grouping, according to common $f_0$.

However, in some cases, speech does not appear to be governed by the basic principles of ASA. For example, speech produced by a single speaker often violates grouping principles such as similarity and good continuation. This is due to aspects of vocal production, including vocal articulators (i.e., closures when articulating plosives and affricates) (Roberts, Summers, & Bailey, 2010). Furthermore, Remez, Rubin, Berns, Pardo, and Lang (1994) found that the perceptual, sequential organization of speech appears to be governed by sensitivity to time-varying acoustic patterns, which are not applicable to non-verbal material. Specifically, ASA and phonetic organization occur independently when the components of an acoustic display simultaneously yield impressions of a single phonetic source and multiple auditory sources (Mattingly & Liberman, 1990). Thus, there appears to be a difference between the organization of speech sounds versus non-speech sounds.
This calls into question the applicability of findings based on non-verbal stimuli to understanding the relationship between the cocktail party problem, streaming, speech, and aging. As there is a difference between the principles of streaming verbal versus non-verbal material, this may explain why no age-related differences in streaming are found when only pure-tone or harmonic complexes are used as stimuli. Therefore, it is possible that age-related differences in streaming may be found if natural speech stimuli are used.

This possibility is supported by evidence that older adults, when presented with complex sounds and listening situations, have more difficulty with streaming than young adults. For example, older adults experience greater difficulty understanding speech in noisy listening conditions than young adults, particularly when there is reverberation (Duquesnoy & Plomp, 1980; Nabelek & Robinson, 1982), or when the competing signal is speech (Duquesnoy, 1983). In addition, older adults have difficulties in following target speech embedded in multi-speaker babble, as compared to young adults (e.g., Wilson & Weakley, 2005; Wingfield & Stine, 1992). Thus, there may be age-related differences in auditory stream segregation when natural speech sounds are used as stimuli, as opposed to pure tones or harmonic complexes.

The current behavioural study addresses this point by examining different aspects of sequential speech segregation using natural vowel stimuli in young and older adults, with pure-tone thresholds that correspond to sub-clinical or mild hearing loss. The individual speech sounds included: /æ/ (e.g., “ae” in cat), /ɜ/ (e.g., “er” in her), /iː/ (e.g., “ee”, in see) and /uː/ (e.g., “oo” as in moose). Please note that these vowels will henceforth be referred to by their phonetic names (“ae”, “er”, “ee”, and “oo”). To our knowledge, this is the first study on sequential streaming to use natural speech stimuli. Participants heard stimuli spoken by female and male speakers, respectively. This use of natural speech stimuli increased the ecological validity of our study, as the results were based on speech stimuli that are experienced in real-world listening situations.

3 Purpose, Hypotheses, and Predictions

The purpose of this study was to examine how the perceptual organization of natural speech is affected by normal aging. This was accomplished by investigating the role of formant transition between adjacent vowels on auditory stream segregation by manipulating vowel-order to promote auditory stream segregation, while keeping the $f_0$ constant.
Stream segregation is promoted by discontinuous first-formant transitions between adjacent vowels (Dorman et al., 1975). Warren, Healy, and Chalikia (1996) have also demonstrated that vowel sequences provide a reliable and useful tool for investigating the perceptual organization of speech sounds that may otherwise be obscured by additional linguistic processing. In addition, vowels with similar first formats, such as “ee” and “oo”, tend to be perceptually grouped together into a separate stream (Dorman et al., 1975), distinct from the “ae” and “er” vowels. Thus, participants were presented with six different combinations of these four vowels: “ee-ae-er-oo”, “ee-er-ae-oo”, “ee-oo-ae-er”, and “ee-oo-er-ae” (continuous formant transitions), and “ee-ae-oo-er” and “ee-er-oo-ae” (discontinuous formant transitions).

There were two dependent measures in the current study, namely accuracy (identifying the speech sounds in each sequence), and whether participants heard streaming in each sequence. These two measures targeted a different aspect of streaming, namely objective and subjective perceptions, respectively.

We predicted that discontinuous sequences were more likely to be heard as two streams as compared to continuous sequences, based on Dorman et al. (1975). It follows that perceiving two streams of sound would make the identification of vowel-order more difficult than when perceiving one stream of sound, as it is more difficult to track, and thus identify individual vowels in a discontinuous sequence. We therefore predicted a significant effect of sequence type for the objective responses, in which the vowel sounds within the discontinuous formant transitions were more difficult to identify than those within continuous formant transitions.

We also hypothesized that, since natural speech stimuli may not follow the same grouping principles as non-verbal material, there would be a significant interaction of age and stimuli type for both the objective and subjective response. Specifically, we predicted that the effects of formant transition on objective and subjective measures of streaming would be greater in younger than older adults.

Measures of pure-tone thresholds, speech-in-noise, mistuned-harmonic threshold, and gap detection threshold were also collected in order to assess multiple aspects of auditory processing. These measures provided insight into peripheral auditory functioning (via pure-tone thresholds, Mazelova, Popelarm & Skaya, 2003; Nelson & Hinojosa, 2006), and central auditory
functioning. Specifically, the speech-in-noise score measured the ability to detect a target speaker amongst an increasing number of background speakers, thus mimicking a realistic cocktail-party environment. The mistuned harmonic threshold measured acuity to spectral detail, while gap detection measured acuity to temporal detail. We hypothesized that aging is associated with the broadening of one’s auditory filters, yielding negative correlations between these three measures and older adults’ objective and subjective measures, respectively.
Chapter 2

4 Methods and Materials

4.1 Participants

Twenty healthy, young adults (13 females, age range = 19 – 30 years, M = 21.60 years) and twenty healthy, older adults (12 females, age range = 65 – 83, M = 70.15 years) participated, after giving written informed consent according to the guidelines of the Baycrest Centre for Geriatric Care and the University of Toronto. All participants were native English speakers; all older adults were administered the Mini-Mental State Exam (MMSE; Folstein, Folstein, & McHugh, 1975) to screen for their cognitive status (Tombaugh & McIntyre, 1992; all scores ≥ 27/30).

Pure-tone thresholds of each participant were measured (250 Hz to 8000 Hz), and hearing loss was defined as per the World Health Organization’s (2011) definition, namely as a speech-frequency pure-tone average of thresholds at 500, 1,000, 2,000, and 4, 000 Hz in the better-hearing ear. These frequencies encapsulated the target frequency range of the experimental stimuli (262 Hz - 1043 Hz). Only one participant (young adult) did not complete the pure-tone threshold test, due to technical difficulties. This participant reported no problems with hearing during the screening phase of recruitment.

All participants who were tested had speech-frequency pure-tone averages that were considered to be either subclinical (≤25 dB hearing level) or mild hearing-loss (26-40 dB; applicable to two older-adult participants), based on Clark’s (1981) definition. In the range of frequencies tested (250-8000Hz), there were no significant differences between the two ears (F(1,37) = .311, p = .580, η² = .008). There were no interaural differences greater than 15 dB at no more than two frequencies. There was a significant age-related increase in hearing thresholds (F(1, 37) = 103.793, p <.001, η² = .737), which is considered a normal aspect of aging (Committee on Hearing, Bioacoustics, and Biomechanics, 1988). All participants were non-musicians, as the ability to perceptually organize many streams of sound becomes specialized with extensive musical training (i.e., distinguishing between many orchestral parts) (Bregman, 1990). Musicianship was defined as regularly practicing an instrument or conducting an ensemble at
least once a month, having a degree or diploma in music, and/or taking music lessons for more than 10 years at any point in life.

4.2 Stimuli

Participants were presented with six different sequences of four natural vowel sounds (ee, ae, oo, and er), spoken by a single speaker (there were four speakers: two males and two females). When recording the stimuli, native English female and male speakers (aged 18-35) were instructed to pronounce all vowels at the same pitch, and to reduce variation in prosody (Gaudrain, Grimault, Healy, & Bera, 2008). Vowel stimuli were recorded by a large-diaphragm Shure KSM44 condenser microphone on Adobe Audition v1.5. The microphone was connected to a laptop computer running on Windows Vista, via an XLR-to-USB adapter. The sampling rate was set at 44,100 Hz.

The most paradigmatic vowel sound of each group of ten was chosen for inclusion in the auditory stream-stimuli. This was defined as the vowel sound that had the least variation in pitch, and was closest to the mean frequency profile for all ten sounds. All vowels were then adjusted to have the same RMS power. No further manipulations were made to the stimuli, to maintain the realistic quality of the vowels. Table 1 displays the frequency-profiles for each vowel sound used in the study, from the fundamental frequency through the fourth formant. Vowels in the lower frequency range were used because presbycusis tends to affect higher frequencies.

The six combinations of the four vowel sounds that participants heard were: “ee-ae-er-oo”, “ee-ae-oo-er”, “ee-er-ae-oo”, “ee-er-oo-ae”, “ee-oo-ae-er”, and “ee-oo-er-ae.” The vowel “ee” was arbitrarily selected as the first vowel in each sequence, as per Dorman et al. (1975). The duration of each vowel sound was 200ms, matching the length often used for artificially-simulated vowels (as in Scheffers, 1982, and Assmann & Summerfield, 1994). The inter-stimulus interval was 12 ms. Each trial included 24 randomly-ordered repetitions of each of the four-vowels in the six combinations listed above, lasting 20.4 seconds, including 2.5 seconds of rise and fall time. This sequence duration was chosen to allow streaming to build up and stabilize (Bregman, 1990; Warren, Obusek, Farmer, & Warren, 1969). 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.
Table 1.

*Frequencies (Hz) for each vowel for each speaker.*

<table>
<thead>
<tr>
<th>Speaker</th>
<th>Formant</th>
<th>ee</th>
<th>ae</th>
<th>er</th>
<th>oo</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female 1</td>
<td>$f_0$</td>
<td>156</td>
<td>156</td>
<td>148</td>
<td>162</td>
</tr>
<tr>
<td></td>
<td>F1</td>
<td>262</td>
<td>1040</td>
<td>519</td>
<td>316</td>
</tr>
<tr>
<td></td>
<td>F2</td>
<td>2461</td>
<td>1648</td>
<td>1339</td>
<td>600</td>
</tr>
<tr>
<td></td>
<td>F3</td>
<td>3596</td>
<td>2814</td>
<td>1542</td>
<td>2438</td>
</tr>
<tr>
<td></td>
<td>F4</td>
<td>4039</td>
<td>434</td>
<td>3943</td>
<td>3847</td>
</tr>
<tr>
<td>Female 2</td>
<td>$f_0$</td>
<td>262</td>
<td>234</td>
<td>244</td>
<td>263</td>
</tr>
<tr>
<td></td>
<td>F1</td>
<td>361</td>
<td>1043</td>
<td>636</td>
<td>429</td>
</tr>
<tr>
<td></td>
<td>F2</td>
<td>2871</td>
<td>1490</td>
<td>1575</td>
<td>1345</td>
</tr>
<tr>
<td></td>
<td>F3</td>
<td>3482</td>
<td>1815</td>
<td>2052</td>
<td>2613</td>
</tr>
<tr>
<td></td>
<td>F4</td>
<td>4362</td>
<td>3079</td>
<td>3914</td>
<td>3841</td>
</tr>
<tr>
<td>Male 1</td>
<td>$f_0$</td>
<td>158</td>
<td>151</td>
<td>136</td>
<td>151</td>
</tr>
<tr>
<td></td>
<td>F1</td>
<td>307</td>
<td>860</td>
<td>532</td>
<td>327</td>
</tr>
<tr>
<td></td>
<td>F2</td>
<td>2351</td>
<td>1629</td>
<td>1286</td>
<td>1098</td>
</tr>
<tr>
<td></td>
<td>F3</td>
<td>2847</td>
<td>2714</td>
<td>1762</td>
<td>2493</td>
</tr>
<tr>
<td></td>
<td>F4</td>
<td>4819</td>
<td>4540</td>
<td>3374</td>
<td>3364</td>
</tr>
<tr>
<td>Male 2</td>
<td>$f_0$</td>
<td>156</td>
<td>125</td>
<td>134</td>
<td>137</td>
</tr>
<tr>
<td></td>
<td>F1</td>
<td>276</td>
<td>823</td>
<td>527</td>
<td>284</td>
</tr>
<tr>
<td></td>
<td>F2</td>
<td>2011</td>
<td>1329</td>
<td>1422</td>
<td>1582</td>
</tr>
<tr>
<td></td>
<td>F3</td>
<td>2940</td>
<td>2435</td>
<td>1690</td>
<td>1928</td>
</tr>
<tr>
<td></td>
<td>F4</td>
<td>3532</td>
<td>3600</td>
<td>3221</td>
<td>3255</td>
</tr>
</tbody>
</table>
4.3 Procedure

The entire testing session lasted five hours per participant. Participants were seated in a sound-proof booth, approximately 85 cm from a computer monitor. First, each participant completed a training task, during which they were presented with sixteen individual vowel sounds, one-at-a-time. The presentations were run through a custom Matlab program (Version 5.3, The Mathworks, Natick, MA) on a Pentium 4 PC. Each vowel sound was spoken by male and female speakers, and were each 200 ms in length, presented at an average intensity of 75 dB SPL via 3A Insert EarTone Headphones (Indianapolis, IN). After the participant initiated the training task via a computer keyboard, each speech sound was presented randomly. After each presentation, participants had to match the sound they heard with one of four keys, labelled “ee”, “ae”, “er”, and “oo”. Participants had to correctly identify all sixteen sounds before proceeding to the experimental task.

In the experimental paradigm, participants were presented with one of the six sequences containing the four vowel sounds (ee-ae-er-oo, ee-er-ae-oo, ee-oo-ae-er, ee-oo-er-ae, ee-ae-oo-er, and ee-er-oo-ae). After each sequence was heard, participants were asked two questions. First, they were asked to select the stream of vowels that corresponded to the one that they just heard (the objective task). The six possible vowel streams were written out and listed on the computer screen; participants entered a number (1 through 6) to indicate which one they just heard.

Secondly, participants were asked a subjective question on the computer screen: to rate whether they heard one or two streams of sound, by entering either “a” or “b” on a computer keyboard. Participants’ subjective response was defined as the proportion of trials participants responded that they heard two streams of sound. The subjective response was scored from 0 (participants responded that they heard a single stream of sound) to 1 (participants responded that they heard two streams of sound). This design allowed the relationship between objective and subjective measures of streaming to be observed. Each participant completed three blocks of 36 trials of sequences per block (each trial involved listening to one sequence, and then answering the two questions about the sequence). Each participant completed three blocks (108 trials) of one of two female speakers (Female 1 or Female 2), and three blocks (108 trials) of one of two male speakers (Male 1 or Male 2). The order of the set of blocks for each speaker was
counterbalanced across participants. Participants were given breaks of approximately five minutes between blocks.

4.3.1 Hearing Assessments

Participants also completed a mistuned harmonic threshold test, a gap detection test, and a speech-in-noise test. These tests were administered between the two sets of three blocks in order to provide a change of pace from the previous task.

4.3.1.1 Mistuned-Harmonic Detection Threshold

Stimuli were harmonic complexes, made up of ten harmonics at equal intensity levels with a $f_0$ of 200 Hz at 65 dB SPL. Stimulus duration was 200 ms including a 10 ms rise and fall time. The initial value of mistuning was set at 16%. In each trial, two stimuli were presented sequentially. One stimulus contained the mistuned harmonic (order was varied randomly). Participants indicated which sound contained the mistuned-harmonic by pressing a button on a response box. Participants were given feedback on whether they answered correctly after each trial. This was done so each individual could reach their lowest mistuned harmonic threshold. After three correct responses, the amount of mistuning was reduced by 50% (e.g., reduced from 16% of the original value to 8% of the original value). If one incorrect response was made, mistuning was increased by 32% on the next trial. Following the first two reversals, the amount of mistuning decreased by 24% after three correct responses, and the final thresholds were an average of the last eight reversals. A total of three blocks of trials were run, and the best two of three thresholds were then averaged together. One young adult and one older adult did not complete this test due to technical difficulties.

4.3.1.2 Gap-Detection Threshold

Stimuli were tone pips produced by multiplying a 1-kHz pure tone by a temporal window created by summing a series of Gaussian envelopes spaced 0.5 ms apart (Schneider, Pichora-Fuller, Kowalchuk, & Lamb, 1994). Two 10-ms tones marked the beginning and end of the gap. The duration of the gap, $\Delta t$, was defined as the time between the last Gaussian in the leading marker and the first Gaussian of the lagging marker. The comparison stimulus (a tone whose duration
and energy are equal to that of the two markers defining the gap) was created by filling in the missing Gaussians between the two markers. The initial gap size was set at 31ms. Stimuli were presented in a two alternative forced-choice paradigm, with a 3 down 1 up tracking procedure to determine the 79.4% accuracy point on the psychoacoustic curve (Levitt, 1971). All stimuli were presented at 65 dB SPL.

In each trial, two stimuli were presented sequentially in random order. The listener identified which of the two sounds contained a gap by pressing a button on a response box. For the first reversal, the size of the gap was reduced by 8ms after three correct responses, or increased by 8ms after one incorrect response. The amount of change in gap size was reduced by 50% after each reversal to a minimum of 0.5ms. Each block of trials lasted until there were 12 reversals. The threshold was determined by averaging the last 8 reversals. Participants were given feedback on whether they answered correctly after each trial. This was done so each individual could reach their lowest gap detection threshold. A total of three blocks of trials were run, and the best two of three thresholds were then averaged together. Two young adults and one older adult did not complete this test, due to technical difficulties.

### 4.3.1.3 Speech-In-Noise Test

The effects of age on a participant’s ability to process speech in noise was assessed using the QuickSIN (Speech-In-Noise) test (Version 1.3; Etymotic Research, 2001). 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 combined amplitude of 70 db SPL using pre-recorded signal-to-noise ratios (SNRs) which decreased in 5-dB steps from 25 dB (very easy) to 0 (very difficult). Thresholds were defined as the SNR needed to identify 50% of the target words for the four lists. Participants were instructed to repeat back each target sentence, and were given one point for each of five target words in each sentence, for a possible total of 30 points per list. The SNR loss was calculated by subtracting the total number of correct words from 25.5 - the SNR the participant needed to correctly identify 50% of the key words in the target sentences (Killion, 1997). Killion, Niquette, Gudmundsen, Revit, & Banergee (2004) offer a description of the standardized procedure of the QuickSIN. One young adult did not complete this test, due to technical difficulties.
4.3.2 Statistical Analyses

Participants’ responses to the objective task were converted into d-prime (d’) scores. As per signal detection theory as outlined in Kaplan, MacMillan, & Creelman (1978), d’ was the difference between the z-transformation of the hit rate (the proportion of trials participants gave the ‘same’ response, given the same stimuli) and the z-transformation of the false alarm rate (the proportion of trials participants gave the ‘same’ response, given different stimuli).

Participants’ response to the subjective task was converted to a proportion from 0 to 1, based on the mean number of times a participant entered that they heard two streams of sound for each sequence.

For the primary analyses, linear mixed effects modeling was used to examine the differences between groups (age) and within groups (sequence-type). The six stimuli were classed as either “discontinuous” (ee-ae-oo-er, and ee-er-oo-ae) or “continuous” (ee-ae-er-oo, ee-er-ae-oo, ee-oo-ae-er, ee-oo-er-ae), based on their frequency profiles (Figure 1 and Figure 2, respectively). These sequence classifications originate from the findings of Dorman et al. (1975), in which stream segregation was promoted by discontinuous formant transitions between adjacent vowels. Since each participant had a score for each of six sequences spoken by either the Female or Male speaker, sequence type was considered a nested variable. Thus, twelve scores were nested within the sequence-type variable. Sequence-type remained a nested variable throughout all analyses.

The denominator degrees-of-freedom for effects involving sequence-type were based on twelve rows of scores nested within six sequences, for forty participants. It is important to note that the model-dimension output reported that all analyses were based on forty participants. Thus, the denominator degrees-of-freedom accounted for the nested variables, rather than “interpreting” the number of rows as 480 individual participants.
Figure 1.

Frequency profiles for discontinuous sequences at $f_0$ (green triangle), the first formant (blue lozenge), and the second formant (red square).
Note: There was discontinuity between the first formants of each of these speech sounds, in which there was very rapid alternation between high and low frequencies. The low-to-high, high-to-low, and low-to-high pattern of the frequency-gradients is evident in all four speakers.
Figure 2.

Frequency profiles for continuous sequences at $f_0$ (green triangle), the first formant (blue lozenge), and the second formant (red square).
Note: There was continuity between the first formants of each of these speech sounds, in which there was a smooth formant transition between a minimum of two consecutive vowels.
Chapter 3

5 Results

It is of note that the two participants who were classified as having mild hearing loss were initially removed from the above analyses. All of the above analyses were run with and without these participants. No difference in the pattern of results was found. Thus, these two participants were included in the analysis.

5.1 Preliminary Analyses

Accuracy and subjective report did not differ significantly between the two female speakers (F(1, 38) < 1 in both cases) or between the two male speakers (F(1, 38) < 2.747, p ≥ .106). Hence, the objective and subjective dependent measures for the two female and two male speakers were averaged together, respectively. The averaged data will henceforth be referred to as “Female speaker” and “Male speaker”.

5.2 Hearing-Assessment Analyses

Figure 3 shows the group mean audiometric thresholds, while figures 4 through 6 show the group mean thresholds for the mistuned-harmonic threshold, gap detection threshold, and speech in noise (QuickSIN), respectively. There were no significant differences between young and older participants for mistuned harmonic thresholds, F(1, 36) = .511, p = .480, ηp² = .014, or for gap-detection thresholds F(1, 35) = 3.200, p = .082, ηp² = .084. There was a significant differences between young and older adults on QuickSIN scores, F(1,37) = 17.373, p < .001, ηp² = .320. Older adults had a higher threshold, defined as the SNR needed to identify 50% of the target words for the four lists (M = 1.863, SD = 1.636), than younger adults (M = .040, SD = 1.003).
Figure 3.

Audiometric threshold for young and older adults.

Figure 4.

Group mean mistuned-harmonic thresholds; unit for the y-axis is the amount of mistuning (Hz).
Figure 5.

*Group mean gap-detection thresholds; unit for the y-axis is gap-duration (ms).*

Figure 6.

*Group mean performance on the QuickSIN test; unit for the y-axis is SNR loss (dB).*
5.3 Primary Analyses

Figure 7 shows the group mean d’-measure for the continuous and discontinuous formant transitions. Overall, young adults (M = 2.251, SD = 1.565) were more accurate than older adults (M = 1.135, SD = 1.341) in identifying the order of the vowels in the sequences, F (1, 38) = 10.930, p = .002. Moreover, participants were more accurate when the transition between the first formant was continuous (M = 2.094, SD = 1.493) than discontinuous (M = .891, SD = 1.372), F (1, 430) = 191.239, p < .001. There was a significant interaction between age and type of sequence, F(1, 430) = 9.362, p = .002 (Figure 7). A closer examination of this interaction revealed greater age-related differences when the sequence was continuous (F(1,38) = 14.118, p = .001) than discontinuous (F(1, 38) = 5.753, p = .021). Young adults were more accurate in identifying continuous sequences (M = 2.741, SD = 1.345) than older adults (M = 1.271, SD = 1.521). Similarly, young adults were more accurate in identifying discontinuous sequences (M = 1.271, SD = 1.521) than older adults (M = .510, SD = 1.088). These means also demonstrate that overall accuracy for both age-groups was higher for continuous (M = 2.094, SD = 1.493) than discontinuous sequences (M = .891, SD = 1.372). The main effect of speaker voice (i.e., male versus female) was not significant (F<1), nor did speaker voice interact with any of the other factors.

Figure 8 shows the group mean subjective responses for continuous and discontinuous formant transitions. Participants reported that they heard two streams of sound more frequently when the sequence was discontinuous (M = .662, SD = .350) than continuous (M = .578, SD = .357), F (1,430) = 12.149, p = .001. There was no significant main effect of age (F(1, 38) = 1.372, p = .249); however, the interaction between age and sequence-type was significant, (F(1, 430) = 5.052, p = .025) (Figure 8). A closer examination of this interaction revealed than an effect of sequence type was significant in the young (F(1, 238) = 8.290, p = .004) but not the older-adult group (F(1, 238) = .354, p = .552). Specifically, young adults perceived more streaming for discontinuous (M = .640, SD = .342) rather than continuous sequences (M = .505, SD = .342). Older adults perceived similar amounts of streaming for discontinuous (M = .683, SD = .359) and continuous sequences (M = .654, SD = .356). The main effect of speaker voice (i.e., male versus female) was not significant (F<1), nor did speaker voice interact with any of the other factors.
Figure 7.

*Group mean sensitivity measure (d’) in young and older adults.*

![Bar chart showing group mean sensitivity measure (d’) in young and older adults.]

Figure 8.

*Group mean proportion of reporting hearing two streams in young and older adults.*

![Bar chart showing group mean proportion of reporting hearing two streams in young and older adults.]

25
5.4 Correlations

Partial correlations between sequence type, mistuned harmonic threshold, gap detection threshold, QuickSIN scores, and d’ scores can be found in Table 2. Partial correlations between these three measures and the proportion of times participants heard sequences as two streams can be found in Table 3. Age was controlled for in both partial correlation calculations.

Table 2.

*Partial correlations between d’ and hearing tests, controlling for age.*

<table>
<thead>
<tr>
<th></th>
<th>Mistuned</th>
<th>Gap</th>
<th>QuickSIN</th>
</tr>
</thead>
<tbody>
<tr>
<td>d’</td>
<td>-.185</td>
<td>-.166</td>
<td>-.142</td>
</tr>
<tr>
<td></td>
<td>.279</td>
<td>.335</td>
<td>.410</td>
</tr>
<tr>
<td></td>
<td>34</td>
<td>34</td>
<td>34</td>
</tr>
<tr>
<td>Mistuned</td>
<td></td>
<td>.562</td>
<td>.185</td>
</tr>
<tr>
<td></td>
<td></td>
<td>.000**</td>
<td>.280</td>
</tr>
<tr>
<td></td>
<td></td>
<td>34</td>
<td>34</td>
</tr>
<tr>
<td>Gap</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>-.095</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>.582</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>34</td>
<td></td>
</tr>
</tbody>
</table>

*Note: **. Correlation is significant at the 0.01 level (2-tailed).*
Table 3.

Partial correlations between the proportion of hearing two streams and hearing tests, controlling for age.

<table>
<thead>
<tr>
<th></th>
<th>Mistuned</th>
<th>Gap</th>
<th>QuickSIN</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Streaming</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>-.104</td>
<td>-.337</td>
<td>.015</td>
</tr>
<tr>
<td></td>
<td>.546</td>
<td>.045</td>
<td>.931</td>
</tr>
<tr>
<td></td>
<td>34</td>
<td>34</td>
<td>34</td>
</tr>
<tr>
<td><strong>Mistuned</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>.583</td>
<td>.240</td>
<td></td>
</tr>
<tr>
<td></td>
<td>.000**</td>
<td>.159</td>
<td></td>
</tr>
<tr>
<td></td>
<td>34</td>
<td>34</td>
<td></td>
</tr>
<tr>
<td><strong>Gap</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>-.022</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>.900</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>34</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Note:** **. Correlation is significant at the 0.01 level (2-tailed).

Pearson product-moment correlations for d’ for young (table 4) and older adults (table 5) yielded no significant differences, with the exception of a significant correlation between the gap-detection threshold and mistuned-harmonic threshold score, \( r = .654, p = .001 \) in older adults. Pearson product-moment correlations for the frequency of reporting two streams was significantly correlated with gap detection threshold (\( r = -.448, p = .047 \)) only in young adults (table 6). Mistuned harmonic-detection threshold and gap-detection threshold were significantly correlated in young adults, \( r = .681, p = .001 \) (table 6). There were no significant correlations between early hearing-assessment measure and streaming in older adults (table 7). There were also no significant correlation between the two dependent measures (d’ and the perception of two streams), \( r = .014, p = .934 \).
Table 4.

*Correlations between d’ and hearing tests for young adults only.*

<table>
<thead>
<tr>
<th></th>
<th>Mistuned</th>
<th>Gap</th>
<th>QuickSIN</th>
</tr>
</thead>
<tbody>
<tr>
<td>d’</td>
<td>.144</td>
<td>-.177</td>
<td>.173</td>
</tr>
<tr>
<td></td>
<td>.556</td>
<td>.482</td>
<td>.479</td>
</tr>
<tr>
<td></td>
<td>19</td>
<td>18</td>
<td>19</td>
</tr>
</tbody>
</table>

| Mistuned  | .215 | .062 |
|           | .391 | .801 |
|           | 18   | 19   |

| Gap       |       | -.074 |
|           |       | .769  |
|           |       | 18    |

*Note: p > .05 for all correlations.*

Table 5.

*Correlations between d’ and hearing tests for older adults only.*

<table>
<thead>
<tr>
<th></th>
<th>Mistuned</th>
<th>Gap</th>
<th>QuickSIN</th>
</tr>
</thead>
<tbody>
<tr>
<td>d’</td>
<td>-.442</td>
<td>-.224</td>
<td>-.344</td>
</tr>
<tr>
<td></td>
<td>.058</td>
<td>.356</td>
<td>.138</td>
</tr>
<tr>
<td></td>
<td>19</td>
<td>19</td>
<td>20</td>
</tr>
</tbody>
</table>

| Mistuned  | .654** | .234 |
|           | .002   | .335 |
|           | 19     | 19   |

| Gap       |       | -.108 |
|           |       | .659  |
|           |       | 19    |

*Note: ** Correlation is significant at the 0.01 level (2-tailed).*
Table 6.

*Correlations between streaming judgment and hearing tests for young adults.*

<table>
<thead>
<tr>
<th></th>
<th>Mistuned</th>
<th>Gap</th>
<th>QuickSIN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Streaming</td>
<td>-.089</td>
<td>-.448*</td>
<td>-.021</td>
</tr>
<tr>
<td></td>
<td>.709</td>
<td>.047</td>
<td>.929</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>20</td>
<td>20</td>
</tr>
<tr>
<td>Mistuned</td>
<td></td>
<td>.681**</td>
<td>.346</td>
</tr>
<tr>
<td></td>
<td>.001</td>
<td>.135</td>
<td></td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>20</td>
<td></td>
</tr>
<tr>
<td>Gap</td>
<td></td>
<td></td>
<td>-.055</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>.816</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>20</td>
</tr>
</tbody>
</table>

*Note:* *Correlation is significant at the 0.05 level (2-tailed).*

**Correlation is significant at the 0.01 level (2-tailed).**

Table 7.

*Correlations between streaming judgment and hearing tests for older adults.*

<table>
<thead>
<tr>
<th></th>
<th>Mistuned</th>
<th>Gap</th>
<th>QuickSIN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Streaming</td>
<td>.036</td>
<td>.464</td>
<td>.128</td>
</tr>
<tr>
<td></td>
<td>.887</td>
<td>.061</td>
<td>.602</td>
</tr>
<tr>
<td></td>
<td>18</td>
<td>17</td>
<td>19</td>
</tr>
<tr>
<td>Mistuned</td>
<td>-.131</td>
<td>-.080</td>
<td></td>
</tr>
<tr>
<td></td>
<td>.617</td>
<td>.751</td>
<td></td>
</tr>
<tr>
<td></td>
<td>17</td>
<td>18</td>
<td></td>
</tr>
<tr>
<td>Gap</td>
<td></td>
<td></td>
<td>.378</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>.134</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>17</td>
</tr>
</tbody>
</table>

*Note:* *p > .05 for all correlations.*
Chapter 4

6 Discussion

The main effects of age on d’ show an age-related decline in performance, in which older adults had more difficulty in accurately identifying the speech sounds than young adults. The age by sequence-type interaction (Figure 7) demonstrates that there was a larger difference between young adult’s d’ scores for each sequence type, as compared to older adults. This interaction suggests that older adults are less sensitive to the continuity of first-formant transitions, as compared to young adults. These findings support the hypothesis that aging impacts the ability to perceptually organize speech sounds.

For the subjective response, participants were more likely to report two streams when the sequence was discontinuous, accounting for the significant main effect of sequence-type. The significant interaction between age and sequence-type was characterized by older adults indicating that they heard two streams of sound for both sequence-types, unlike young adults. Young adults reported more streaming for the discontinuous sequence, as compared to the continuous sequence. It is important to interpret these findings with the consideration that this was a purely subjective perceptual measure. There is a limitation in this type of measure, in that one must consider participant’s response biases. Conversely, the inclusion of this measure was also able to give important insight into each participant’s experience of the sequences, justifying its inclusion in the current study.

We found a significant effect of sequence type on the subjective response, with discontinuous formant transitions rated as “two streams” more frequently than continuous formant transitions. This finding confirms our prediction that discontinuous formant transitions were perceived as two streams of sound more often than continuous formant transitions, as per Dorman et al. (1975). A significant effect of sequence type on d’ scores was also found. Specifically, participants were more accurate in identifying speech sounds when formant transitions between adjacent vowels were smooth (continuous), rather than abrupt (discontinuous). This is consistent with the findings of Cullinan, Erdos, Schaefer, and Tekieli (1977), which suggested that listeners can better identify the correct order of vowel sounds when repeating sequences more closely resemble connected, rather than disconnected, speech. Together, these findings support the
hypothesis that perceiving two streams of sound makes the identification of vowel-order more difficult than when perceiving one stream of sound.

In order to see if we replicated the general patterns of identification in Dorman et al. (1975), we examined our raw accuracy scores for each sequence for young adults, and compared them to the raw accuracy scores for the young adults in Dorman et al. (1975). The measure of accuracy was defined as the average “hit-rate” - the count of correctly-identified sequences divided by the count of stimulus presentations for each stimulus for each sequence expressed as a percentage. The discontinuous streams in our study (ee-er-oo-ae and ee-ae-oo-er) had average hit rates of 54% and 51%, respectively. The corresponding streams in Dorman et al. (1975) had average hit rates of 63% and 51%, respectively. For the continuous streams in our study (ee-oo-er-ae, ee-oo-ae-er, ee-er-ae-oo, ee-ae-er-oo), the average hit rates were 87%, 89%, 75%, and 74%, respectively. The corresponding streams in Dorman et al. (1975) had average hit rates of 74%, 64%, 76%, and 83%, respectively. With the exception of ee-oo-ae-er, the mean hit rates are similar. Dorman et al.’s (1975) overall pattern of discontinuous sequences having lower hit rates than the continuous sequences was replicated in the current study.

7 General Discussion

These results collectively suggest that there is a relationship between aging and the sequential streaming of speech. This is a novel finding, compared to the previous literature on sequential streaming which did not find age-effects. A central difference between the current study and previous literature is the type of stimuli – namely, the use of natural speech sounds, rather than non-verbal material. Perhaps the use of natural stimuli is related to the age effects found in the present research. This interpretation is supported by research that suggests that speech is a unique and complex stimulus, perceptually processed in a different manner than non-speech stimuli (Remez et al., 1994; Mattingly & Liberman, 1990; Roberts, Summers, & Bailey, 2010).

These findings also speak to the importance of formant transitions in studying the perceptual organization of speech. In the current study, manipulating the continuity of the first-formant transitions had significant main effects on both objective and subjective measures. This suggests that the manipulation of first-formant transitions is a reliable method of studying the perceptual streaming of speech. The role of first-formant transitions in the perceptual organization of
speech is supported by past literature that posited formant transitions both 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 & Scott, 1973; Dorman et al., 1975).

The results of our partial correlations differ from the patterns found in Mackersie et al. (2001). In the current study, there were no significant correlations between performance on either objective or subjective measures and the three hearing-assessment measures, controlling for age. In Mackersie et al. (2001), the relationship between sentence-perception and fusion thresholds remained significant even after controlling for ages. It is notable that the age-groups in Mackersie et al. (2001) were very broad, with one group aged 23 to 41, and another from 51 to 87. Perhaps when groups are polarized into young and older adults, such as in the current study, controlling for age may yield a different pattern of results in Mackersie et al. (2001). Furthermore, in Mackersie et al.’s (2001) sample of sixteen participants, eleven had mild-moderate sensorineural hearing loss. It does not appear that Mackersie et al. (2001) controlled for hearing loss. It is possible that this is what was driving their significant correlation, contrasting with our non-significant correlations in hearing-intact participants.

We did, however, find a significant age effect on QuickSIN score, suggesting that older adults did worse on detecting speech in noise than young adults in the current study. Furthermore, we know from the current study that older adults differ in their objective and subjective performance, respectively, from young adults. Perhaps there is another variable that moderates a link between speech-in-noise detection and task-performance. It is possible that performance is related to a combination of abilities, such as spectral detail (measured by the mistuned-harmonic threshold), temporal detail (gap-detection threshold), and the ability to detect speech-in-noise (measured by the QuickSIN test).

The results of the hearing-assessments in the current study suggest that young and older adults had similar levels of spectral acuity, as measured by the mistuned-harmonic threshold, as well as seemingly-similar levels of temporal acuity, as measured by the gap-detection threshold. It is important to note that the high variance amongst older adults’ gap-detection thresholds, as compared to the variance for young adults’ thresholds, may indicate an underlying influence of age on temporal acuity. These results, in addition to older adults’ impaired performance on the
QuickSIN test, suggest that older adults’ difficulties with the perceptual organization of speech may be influenced by a factor that is inherent to both temporal acuity and SNR-discrimination.

Examining the trends in the separate correlation tables for young (tables 4 and 6) and older adults (tables 5 and 7) provide insight into the links between our hearing assessment measures, and the objective and subjective measures used in this study. In older adults, there was a clear pattern that accuracy tended to increase as thresholds on all three tests became lower (table 5), unlike for young adults (table 4), who had showed a weak association between increased accuracy and higher thresholds. This demonstrates that a greater ability to detect spectral and temporal detail, as well as a heightened ability to detect speech-in-noise, was related to older adults correctly identifying sequences. This association was not present in young adults.

Examining correlations with the subjective measure, we observed a significant relationship between hearing two streams of sound and having a small gap-detection threshold in young adults (table 6). This indicates that the ability to better discriminate fine temporal detail was closely associated with perceiving two streams of sound in young adults. The mistuned-harmonic threshold and QuickSIN score were weakly associated with streaming (table 6). Specifically, streaming was associated with a higher threshold and score, respectively. These findings suggest that the gap detection test, and thus, detecting temporal detail, is important to the perception of streaming in young adults.

In older adults, hearing two streams was related to larger thresholds for the mistuned-harmonic and gap-detection test, as well as higher QuickSIN scores. This association was strongest for the gap-detection threshold (the relationship with each of the other two measures was very weak). These findings suggest that poor ability to detect temporal detail is related to older adults perceiving two streams of sound. This is the opposite of the finding for young adults.

These trends collectively suggest that accuracy is related to more acute discrimination of spectral and temporal detail in older adults, as well as the ability to detect speech in noise. Detecting temporal detail appears to be most closely related to perceiving two streams of the three hearing assessments. Younger adults seem to detect more streaming as they get better able to detect temporal detail, while older adults seem to detect more streaming as they get worse at detecting temporal detail.
It is also notable that there does not appear to be any unique association between the QuickSIN score (which simulates a realistic cocktail party scenario) and either the objective or subjective responses. This suggests that QuickSIN may not be the best measure of accuracy or streaming, perhaps testing only a restricted range of both measures. There may be overlap between QuickSIN and the two measures, respectively. However, each may measure different cognitive domains. Future research could examine alternate measures of speech-in-noise detection, such as the Connected Speech Test (CST) (Cox, Alexander, & Gilmor, 1987), the Speech Perception in Noise test (SPIN) (Bilger, Nuetzel, Rabinowicz, & Rzeczkowski, 1979), or Hearing in Noise Test (HINT) (Nilsson, Soli, & Sullivan, 1994). Perhaps one of these tests might better represent the link between real-world ability in detecting speech-in-noise, and the objective and subjective responses in the current study.

It is of note that older adults perceived more streaming for both types of sequence, as compared to young adults. In contrast, young adults heard streaming more frequently for the discontinuous sequences, as predicted. These results suggest that older adults are indeed capable of perceptually-perceiving sequential speech sounds. In fact, they are comparable to young adults in this regard (hence, no significant main effect of age on the subjective response). It is on the objective measure that age-related decline in performance is evident.

These subjective and objective-response results are complimentary, in that perceiving two streams of sound makes tracking the individual vowel sounds in a sequence more difficult than when only one stream is perceived. This may explain why older adults had decreased d’ performance on the objective task, as compared to young adults.

It is also important to consider possible alternative explanations for the effects on the subjective and objective responses in the current study. For example, the significant interaction between age and sequence-type for both the objective and subjective measures in the study may be related to temporal-order-processing impairments faced by older adults. Such impairments have been linked to decreased speech-in-noise understanding in older adults in numerous studies (e.g., Gordon-Salant, & Fitzgibbons, 1993, 1999; Schneider and Pichora-Fuller, 2000, and Pichora-Fuller, 2003, as reviewed in Fogerty, Humes, and Kewley-Port, 2010). In addition, Trainor and Trehub (1989) found evidence of temporal-order impairment in elderly listeners, without age differences in the perceptual ability to stream sounds.
However, the current study also demonstrated that there was a difference in the way older adults perceive streaming, as compared to younger adults - findings thus contrast with that of Trainor and Trehub (1989). Perhaps the use of natural speech stimuli contributed to this difference. Nonetheless, it is important to consider the role of impaired temporal-order processing in the significant effects seen in the current study. Future studies could employ a measure of temporal-order processing, such as the temporal-order processing subtest of the Test of Basic Auditory Capabilities (TBAC), to closer examine the link between this impairment and the sequential streaming of speech sounds. Such a link may elucidate the contributions of more cognitive processes (e.g., temporal-order processing) to the perceptual phenomenon of streaming.

One might also argue that age-differences in short-term memory span could moderate the significant interaction between age, accounting for why older adults had lower accuracy than young adults. This possibility is unlikely, based on what is known about aging, short-term memory, and recognition. In the current study, the objective question involved participants listening to a target sequence consisting of four sounds, and then recognizing the target sequence amongst similar distractors moments later. As reviewed by Grady and Craik (2000), only minor age-related impairments are seen in short-term memory span tasks (e.g., repeating back a string of words or letters, or numbers). Furthermore, Grady and Craik (2000) summarized that there are minor age-related declines in performance on recognition memory tasks, where older adults identify if a previously encountered stimulus (e.g., a word) among new, similar distractor stimuli. Some specific studies that demonstrate these findings are McIntosh, Sekuler, Penpeci, Rajah, & Grady (1999), and Sekuler, McLaughlin, and Kahana (2006). Therefore, it is unlikely that memory span had an impact on the age effects seen for the objective responses in the current study. Future studies could use a measure such as digit span to ensure that memory-span differences do not contribute to these age effects.

Finally, speed of processing might be related to objective-task performance in the current study. Caplan, Dede, Waters, Michaud, and Tripodis (2011) found that older adults had a slower speed of processing than young adults, when presented with a series of sentences with relative clauses. Furthermore, older adults exhibited poor comprehension of structurally-complex sentences (Caplan et al., 2011). Processing time was also slower for these sentences, suggesting a link between slow and inefficient processing in older adults (Caplan et al., 2011). Perhaps differences in speed of processing contributed to both the age effect and interaction with the d’-
measure in the current study. Namely, young adults may have been able to process sequences more quickly and efficiently than older adults. This could explain why young adults were more sensitive to the differences between sequence-types, as compared to older adults. Future studies could measure processing speech by having participants complete tasks such as digit copying, pattern comparison, and letter comparison (as in Caplan et al., 2011). Furthermore, reaction-time could be incorporated as a dependent measure to reflect processing time for each sequence.

8 Future Directions and Conclusions

By examining an essential component of sequential streaming using both objective and subjective behavioural measures, as well as natural speech stimuli, this study has laid important groundwork for future investigations on aging and streaming. Such future investigations could include studying a middle-aged population, which may provide insight to the age at which performance on the organization of sequential speech sounds starts to be impacted by age.

In addition, future studies could use evoked response potential (ERP) to examine neural activity during the perception of the different vowel sequences used in the current study. The proposed research can also serve as the foundation for future studies that examine how different types of training (e.g., music lessons) can affect sequential streaming of speech in young and older adults. This would complement research by Zendel and Alain (2009), which examined the enhancement of concurrent sound segregation in musicians.

In conclusion, the findings from this study have helped to better conceptualize the impact of normal aging on the perceptual organization of speech sounds. The current novel finding that the perceptual organization of natural speech is affected by normal aging provides evidence for the importance of using realistic stimuli when studying sequential streaming. This supports the idea that both sequential and concurrent streaming are impacted by age. Continuing this line of research will help build an integrated theory of auditory scene analysis and cognitive aging, as related to sequential streaming. This research could eventually be applied to interventions than can help preserve older adults’ ability to effectively communicate in a variety of everyday acoustic environments.
References


