SPEECH-IN-SPEECH LISTENING ON THE LISN-S TEST BY OLDER ADULTS WITH GOOD AUDIOWGRAMS DEPENDS ON COGNITION AND HEARING ACUITY AT HIGH FREQUENCIES

OBJECTIVES: The main objective was to investigate age-related differences on the listening in spatialized noise–sentences (LiSN-S) test in adults with normal audiometric thresholds in most of the speech range. A second objective was to examine the contributions of auditory, cognitive, and linguistic abilities to LiSN-S outcomes.

DESIGN: The LiSN-S test was administered to participants in an older group ($M_{\text{Age}} = 72.0$, $SD = 4.3$ years) and a younger group ($M_{\text{Age}} = 21.7$, $SD = 2.6$ years) with $N = 26$ per group. All the participants had clinically normal audiometric thresholds at frequencies up to and including 3000 Hz. The LiSN-S test yields a speech reception threshold (SRT) in each of the four speech-in-speech listening conditions that differ in the availability of voice difference cues and/or spatial separation cues. Based on these four SRTs, the scores were calculated for the talker advantage, the spatial advantage, and the total advantage as a result of both the types of cues. Additionally, the participants completed four auditory temporal-processing tests, a cognitive screening test, a vocabulary test, and tests of linguistic closure for high- and low-context sentences. The contributions of these predictor variables and measures of pure-tone hearing acuity to LiSN-S outcomes were analyzed for both the groups using regression analyses.

Results: Younger listeners outperformed the older listeners on all four LiSN-S SRTs and all the three LiSN-S advantage measures. Age-related differences were larger for conditions involving the use of spatial cues. For the younger group, all LiSN-S SRTs were predicted by the measure of linguistic closure in low-context sentences; in addition, the SRT for the condition with voice difference cues but without spatial separation cues was predicted by vocabulary, and the SRT for the condition with both voice difference cues and spatial separation cues was predicted by temporal resolution at low frequencies. Vocabulary also contributed
to the talker advantage in the younger group, whereas the spatial advantage was predicted by high-frequency pure-tone hearing acuity in the range 6000 to 10,000 Hz (PTA\textsubscript{HIGH}). For the older group, the LiSN-S SRT in the condition with neither voice difference cues nor spatial separation cues was predicted by age; their other three LiSN-S SRTs and all advantage measures were predicted by PTA\textsubscript{HIGH}. In addition, for the older group, cognition predicted LiSN-S SRT outcomes in three of the four conditions. Measures of auditory temporal processing, linguistic abilities, or hearing acuity up to and including 4000 Hz did not predict LiSN-S outcomes in this group.

**Conclusions:** LiSN-S outcomes were poorer for adults aged 65 years or older, even those with good audiograms, compared with younger adults and also compared with people up to the age of 60 years from a previous study. In the present study, regardless of the types of cues, auditory and cognitive interactions were reflected by the combined influences on LiSN-S outcomes of high-frequency hearing acuity and measures of linguistic and cognitive processing. The data also suggest a hierarchy in the deployment of processing resources, which would account for the observed shift from linguistic abilities in the younger group to general cognitive abilities in the older group.

<table>
<thead>
<tr>
<th>Acronym</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>4FAHL</td>
<td>4-frequency average hearing loss (PTA for the frequencies 500, 1000, 2000, 4000 Hz)</td>
</tr>
<tr>
<td>ANOVA</td>
<td>analysis of variance</td>
</tr>
<tr>
<td>ASA</td>
<td>auditory scene analysis</td>
</tr>
<tr>
<td>DVO</td>
<td>SRT for different voices at ±0°</td>
</tr>
<tr>
<td>DV90</td>
<td>SRT for different voices at ±90°</td>
</tr>
<tr>
<td>$F_0$DL</td>
<td>$F_0$ difference limen</td>
</tr>
<tr>
<td>$F_0$DL$_{10}$</td>
<td>log-transformed $F_0$DL scores</td>
</tr>
<tr>
<td>GAP$_{5U}$</td>
<td>gap detection threshold</td>
</tr>
<tr>
<td>GAP$_{5U-LG10}$</td>
<td>log-transformed GAP$_{5U-LG10}$ scores</td>
</tr>
<tr>
<td>LISN-S</td>
<td>listening in spatialized noise–sentences</td>
</tr>
<tr>
<td>MOCA</td>
<td>Montreal cognitive assessment</td>
</tr>
<tr>
<td>MOCA$_{RLG10}$</td>
<td>reversed and log-transformed MoCA scores</td>
</tr>
<tr>
<td>O</td>
<td>older</td>
</tr>
<tr>
<td>PTA</td>
<td>pure-tone average</td>
</tr>
<tr>
<td>PTA$_{HIGH}$</td>
<td>PTA for the frequencies 6000, 8000, 9000, 10,000 Hz</td>
</tr>
<tr>
<td>SPIN</td>
<td>speech perception in noise</td>
</tr>
<tr>
<td>SNR</td>
<td>signal-to-noise ratio</td>
</tr>
<tr>
<td>SRM</td>
<td>spatial release from masking</td>
</tr>
<tr>
<td>SRT</td>
<td>speech reception threshold</td>
</tr>
<tr>
<td>SVO</td>
<td>SRT for same voice at ±0°</td>
</tr>
<tr>
<td>SV90</td>
<td>SRT for same voice at ±90°</td>
</tr>
<tr>
<td>TFS</td>
<td>temporal fine structure</td>
</tr>
<tr>
<td>TFS$_{750HZ}$</td>
<td>TFS sensitivity at 750 Hz</td>
</tr>
<tr>
<td>TFS$_{3000HZ}$</td>
<td>TFS sensitivity at 3000 Hz</td>
</tr>
<tr>
<td>TING</td>
<td>detection of tones in noise gaps</td>
</tr>
<tr>
<td>TING$_{750HZ}$</td>
<td>detection of 750-Hz tones in noise gaps</td>
</tr>
<tr>
<td>TING$_{3000HZ}$</td>
<td>detection of 3000-Hz tones in noise gaps</td>
</tr>
<tr>
<td>TRT</td>
<td>text reception threshold</td>
</tr>
<tr>
<td>TRT$_{500}$</td>
<td>score from modified TRT test with 500 ms full-sentence presentation</td>
</tr>
<tr>
<td>TRT$_{LOW}$</td>
<td>TRT$_{500}$ scores for low-context sentences</td>
</tr>
<tr>
<td>TRT$_{HIGH}$</td>
<td>TRT$_{500}$ scores for high-context sentences</td>
</tr>
<tr>
<td>WIN</td>
<td>words in noise</td>
</tr>
<tr>
<td>Y</td>
<td>younger</td>
</tr>
<tr>
<td>Z-TRT</td>
<td>mean of z-transformed TRT$<em>{LOW}$ and TRT$</em>{HIGH}$ scores</td>
</tr>
</tbody>
</table>
4.1 INTRODUCTION

Study objectives

The purpose of the present study was twofold. The first objective was to investigate age-related differences on the listening in spatialized noise–sentences (LiSN-S; Cameron & Dillon, 2007) test in adults with normal audiometric thresholds in most of the speech range. The second objective was to examine the contributions of auditory, cognitive, and linguistic abilities to LiSN-S outcomes.

4.1.1 Age-related Differences on the LiSN-S Test

Having a conversation in adverse listening conditions is difficult for most listeners, and it is especially hard for older people. According to the 1988 report of the Committee on Hearing Bioacoustics and Biomechanics, the origin of the speech-understanding difficulties of older adults could be declines in auditory peripheral and/or central auditory, and/or cognitive processing. It is widely accepted that deficits in the auditory periphery resulting in elevated hearing thresholds are the most apparent source of the difficulties experienced by older people (e.g., van Rooij & Plomp, 1992; Divenyi et al., 2005; Schmiedt, 2010). However, even older adults with normal or near-normal audiograms who have little difficulty in quiet situations can experience more difficulties than younger listeners in challenging realistic listening conditions (e.g., Dubno et al., 1984; Pichora-Fuller & Souza, 2003; Tun et al., 2009). The nature and degree of their difficulties in everyday situations may depend on the particular suprathreshold auditory declines associated with different subtypes of presbycusis (for a review see Schmiedt, 2010) and/or age-related changes in cognitive processing (for reviews see Schneider et al., 2002, 2010). In audiology clinics, speech understanding is typically evaluated with tests of word recognition in quiet or noise (for a review see Wilson & McAr-dle, 2005). However, the listening conditions tested in the clinic do not approximate the challenges of everyday listening and typical speech-in-noise thresholds do not enable benefit from the different cues to be evaluated. The recently developed LiSN-S test (Cameron & Dillon, 2007) offers a promising new way to measure the contribution of specific cues to older adults’ speech-understanding difficulties in more realistic listening conditions.

The LiSN-S test measures four speech recognition thresholds (SRTs) for sentences presented in a meaningful two-talker background. The four test conditions separately manipulate the availability of spatial separation cues and voice difference cues. Based on the four outcome SRTs, it is possible to calculate the advantage in SRT that listeners gain when either voice differences (talker advantage), spatial separation (spatial advantage), or both types of cues (total advantage) are provided. For the Australian version of the LiSN-S test, normative
data are available for individuals up to the age of 60 years who have normal audiometric thresholds (Cameron, Glyde & Dillon, 2011). For the North-American version of the test, normative data are available for individuals up to the age of 30 years (Brown, Cameron, Martin, Watson & Dillon, 2010). Cameron et al. (2011) found the advantage measures to be largely unaffected by age in listeners younger than 60 years of age, but that all of the LiSN-S SRTs became poorer in the 6th decade.

Recently, Glyde, Cameron, Dillon, Hickson en Seeto (2013) investigated the effect of age, pure-tone hearing loss, and cognitive ability on LiSN-S outcomes in participants spanning a greater range of ages (7–89 years) and hearing losses. Average hearing acuity at the frequencies 500, 1000, 2000, and 4000 Hz (4FAHL [4-frequency average hearing loss]) accounted for more than half of the variance in three of the five reported LiSN-S measures and it was an especially good predictor of the ability of listeners to benefit from spatial separation cues. Controlling for the 4FAHL, age accounted for some SRT variance in the high-cue condition, the condition that provided both spatial separation and voice difference cues, but not for any other LiSN-S measure. The cognitive measure (COGNISTAT; Mueller, Kiernan & Langston, 2001) that Glyde, Cameron et al. (2013) used did not explain variance in any of the LiSN-S outcomes. Notably, age and hearing loss were confounded for the adults in their study sample (N = 65). Specifically, there were only five participants, 65 years of age or older who did not have a 4FAHL of 25 dB HL or more. Thus, it is unclear whether or not the LiSN-S measures would be reduced for listeners of older ages, if age were not confounded with hearing loss. Therefore, the first aim of the present study was to investigate age-related differences in LiSN-S SRTs and advantage measures for listeners 65 years or older who have clinically normal audiometric thresholds at frequencies up to and including 3000 Hz.

4.1.2 Factors Affecting LiSN-S Outcomes

Given that many previous research studies have found that age and higher-level linguistic and cognitive factors influence speech understanding in various adverse listening situations (Mattys et al., 2012), it is somewhat surprising that Glyde, Cameron et al. (2013) found that age was related only to the LiSN-S condition where both the cue types were available and that cognition was unrelated to all LiSN-S outcomes. Our second aim was to investigate the contributions of a wider range of measures to performance on the LiSN-S test. In addition to age and standard pure-tone thresholds, we also tested higher-frequency pure-tone thresholds and a set of suprathreshold auditory temporal processing, linguistic, and cognitive abilities. Higher-frequency audiometric thresholds were tested to assess variations

Previous LiSN-S research

Study aim: age-related differences in LiSN-S scores

Study aim: prediction of LiSN-S scores
in early signs of peripheral hearing loss. We selected measures of suprathreshold auditory processing for which there was evidence of age-related differences and evidence of a relationship to performance on speech-in-noise tests. In addition, two linguistic measures were selected, one a measure of lexical knowledge, and the other a measure of sentence-processing abilities. Finally, a cognitive screening test was selected that included measures of language, memory, and attention that could contribute to speech understanding.

4.1.2.1 Auditory Suprathreshold Processing

For individuals with sensori-neural hearing loss, performance on speech-in-noise tests has been attributed to a variety of changes in suprathreshold auditory processing, associated with elevated pure-tone thresholds; for example, reduced frequency selectivity as a result of broadened auditory filters. However, age alone does not seem to alter auditory filtering (Fitzgibbons & Gordon-Salant, 2010). In contrast, age-related declines in auditory temporal processing can occur independently of elevations in pure-tone thresholds and they have been observed using a variety of behavioral (e.g., Fitzgibbons & Gordon-Salant, 2010) and physiological measures (e.g., Frisina & Frisina, 1997; Anderson, Parbery-Clark, White-Schwoch & Kraus, 2012; Anderson et al., 2013). Importantly, speech conveys information in its temporal fine structure (TFS) and its temporal envelope fluctuations (Rosen, 1992; Greenberg, 1996). Speech understanding in noise relies on the processing of both of these types of temporal information and there is ample evidence that the ability to use these cues changes with age. Furthermore, sensitivity to both the types of temporal speech cues can be useful for segregating speech from a masker, especially when the masker is speech.

Fine-structure cues provide information that can be used to identify and segregate voices based on their fundamental frequency and harmonic structure and age-related differences have been found in numerous experiments assessing the use of TFS in speech understanding (e.g., Summers & Leek, 1998; Vongpaisal & Pichora-Fuller, 2007; Hopkins & Moore, 2010; Jackson & Moore, 2013). For example, for periodicity cues related to the F0, several studies have shown reductions in the accuracy of periodicity coding in older listeners that are unrelated to cochlear hearing loss (Summers & Leek, 1998; Vongpaisal & Pichora-Fuller, 2007). Results from physiological (Frisina et al., 2001; Anderson et al., 2012), psychoacoustic (Pichora-Fuller & Schneider, 1992), and computer modeling work (Saremi & Stenfelt, 2013) suggest that reduced periodicity coding in older listeners is a consequence of decreased neural synchrony in the aging auditory system. Furthermore, regardless of audiometric threshold elevations, age-related differences have also been observed for other measures of TFS sensitivity that are not related specifically to F0 discrimina-
tion (e.g., Grose & Mamo, 2010; Hopkins & Moore, 2011; Neher et al., 2011). In general, TFS cues seem to support speech understanding for listeners who have relatively good audiograms, particularly in fluctuating maskers (e.g., Lorenzi, Gilbert, Carn, Garnier & Moore, 2006; Hopkins & Moore, 2009; Eaves, Kitterick & Summerfield, 2011). TFS information is not only used to identify the target speech in the masker’s amplitude dips, but it contributes to speech understanding over a wide dynamic range of masking (Drullman, 1995; Stone, Moore & Füllgrabe, 2011) and can be used specifically in the amplitude peaks of the target signal for voice segregation in speech-in-speech listening (Stone et al., 2011).

Envelope cues can provide information about phonemic contrasts, vowel identity, and syllabic structure (Rosen, 1992). Studies demonstrating successful understanding of vocoded speech confirm the relevance of the ability to process the signal envelope in speech understanding (e.g., Shannon, Zeng, Kamath, Wygonski & Ekelid, 1995). In addition to the ability to make use of the temporal envelope of the target speech, the ability to make use of envelope fluctuations in the masker has also been found to enhance speech understanding (e.g., George et al., 2006). Age-related decreases in the ability of those with good audiograms to make use of temporal envelope information have been found repeatedly in studies of speech understanding (e.g., Grose, Mamo & Hall, 2009) and in research relating speech perception to psychoacoustic measures such as gap detection (e.g., Gordon-Salant, Yeni-Komshian, Fitzgibbons & Barrett, 2006; Pichora-Fuller, Schneider, Benson, Hamstra & Storzer, 2006; Harris, Eckert, Ahlstrom & Dubno, 2010).

4.1.2.2 Cognitive and Linguistic Abilities

The importance of cognitive processing during speech understanding in noise is well known (e.g., Committee on Hearing Bioacoustics and Biomechanics, 1988; van Rooij & Plomp, 1990; Pichora-Fuller et al., 1995). Indeed, it is a central topic in the field of cognitive hearing science (Arlinger et al., 2009). The influences of cognitive abilities are prominent when the listening conditions are adverse, whether or not listeners have normal audiometric thresholds (e.g., Lunner, 2003; Besser et al., 2013; Rönnberg et al., 2013). Meaningful speech produced by competing talkers introduces high levels of semantic interference with the target talker, especially for single- and two-talker maskers and such maskers can induce high cognitive processing load (Koelewijn et al., 2012; Ng et al., 2013). Not surprisingly, susceptibility to semantic interference depends on the listener’s linguistic and cognitive processing abilities (e.g., Carhart, Tillman & Greetis, 1969; Shinn-cunningham & Ihlefeld, 2004; Mattys, Brooks & Cooke, 2009; Ruggles, Bharadwaj & Shinn-Cunningham, 2011; Brouwer, van Engen, Calandruccio & Bradlow, 2012; Koelewijn et al., 2012). Age-
related declines in auditory processing can aggravate age-related differences in various aspects of cognitive processing, such as speed of processing, working memory, and divided attention (for a review see Schneider et al., 2010).

4.1.2.3 Age-Related Differences in Spatial Listening

As described earlier, the influences of auditory temporal and cognitive processing abilities on speech understanding in noise, especially competing speech, are relatively well documented. However, the spatial separation of a target talker from a competing talker can confer a substantial advantage in everyday listening situations. Specifically, speech understanding is better when the target talker and interfering talkers are spatially separated from each other compared with when the voices are collocated, a phenomenon known as spatial release from masking (SRM; Litovsky, 2012). Older adults have generally poorer SRM (e.g., Murphy, Daneman & Schneider, 2006; Francis, 2010; Glyde, Hickson, Cameron & Dillon, 2011; Gallun, Diedesch, Kampel & Jakien, 2013), but it should be noted that poorer spatial processing in older adults may be confounded with declines in pure-tone hearing acuity (e.g., Marrone, Mason & Kidd, 2008; Neher et al., 2009; Glyde et al., 2011). In a study of the effect of age on SRM in which only older adults with clinically normal audiometric thresholds up to 3000 Hz were tested, although younger adults had better thresholds in noise than older adults, both the age groups demonstrated comparable SRM (Li, Daneman, Qi & Schneider, 2004). On the other hand, (Gallun et al., 2013) found spatial processing deficits for older listeners when controlling for the audibility of the stimuli. A possible explanation for this discrepancy in the findings is that (Li et al., 2004) used the precedence effect to simulate spatial separation, which minimizes signal-to-noise ratio (SNR) cues, whereas Gallun et al. (2013) used real spatial separation.

The reductions in SRM observed for older listeners when SNR cues are available may be related to the mechanisms of binaural processing. The listener has to be able to process and integrate binaural cues of time and level differences to establish the source’s spatial location and benefit from the spatial separation of multiple sources (Neher et al., 2009; Goverts & Houtgast, 2010; Glyde et al., 2011). In addition, it is known that the listeners need to deploy cognitive mechanisms to selectively attend to the target sound source and filter out interfering signals (Francis, 2010). Singh, Pichora-Fuller en Schneider (2008) found that with spatially separated target and competing talkers, older listeners with good audiograms benefited as much as the younger listeners from expectations about the spatial location of the target. In a follow-up study, however, Singh, Pichora-Fuller en Schneider (2013) did find age-related differences when the instructions were modified to require listeners to redirect attention. Thus, it seems that
age-related differences in spatial listening are highly related to attentional mechanisms that become more important as the nature of the task increases the cognitive demands.

4.1.3 Everyday Listening From a Theoretical Perspective

An early two-component model of speech understanding in noise is the attenuation-distortion model introduced by Plomp (1978). The attenuation component of Plomp’s model accounts for the effects of reduced audibility because of elevated audiometric thresholds, while the distortion component recognizes that deficits in suprathreshold aspects of auditory processing also play a role. Subsequently, the Committee on Hearing Bioacoustics and Biomechanics (1988) report hypothesized that the difficulties experienced by older adults in understanding speech in noise could be because of peripheral auditory, central auditory, or cognitive deficits. Researchers in cognitive hearing science (Arlinger et al., 2009) have tried to understand the interactions of the different processing stages in various adverse listening conditions and have developed models describing these interactions for specific cognitive resources, such as working memory (Rönnberg et al., 2013). Generally, these models suggest a hierarchical organization of the processing stages, where higher-order auditory and cognitive processing accounts for variability in the speech understanding performance when audibility is sufficient for successful signal identification (e.g., Humes, 2007; Humes & Dubno, 2010; Humes, Kidd & Lentz, 2013).

Auditory scene analysis (ASA) (Bregman, 1990) is another approach to describing the factors involved in listening in realistic situations. This approach does not explicitly distinguish different processing stages or consider the effects of aging or hearing loss, but it does incorporate the role of attentional mechanisms and the processing of the spatial dimension of everyday listening. ASA research investigates the ability of the listeners to navigate complex auditory scenes, including three-dimensional displays, by means of segregating the acoustic scene into auditory “objects” and “streams” (see also Shinn-Cunningham & Best, 2008). Behavioral research in ASA has identified several auditory and nonauditory suprathreshold processing factors that influence the ability of the listeners to form auditory streams, thereby supporting speech understanding in the presence of a competing sound source. Domains of influence include temporal processing (e.g., binaural TFS sensitivity) different attentional mechanisms (Neher et al., 2011; Snyder, Gregg, Weintraub & Alain, 2012). Furthermore, recent findings from neuroscience suggest that the extraction of auditory objects from auditory scenes is facilitated by a cortical system of ventral and dorsal pathways; whereas the ventral network encodes information about the signal content (“what”), the dorsal net-
work encodes information about the signal location ("where") (Hickok & Poeppel, 2007; Goll, Crutch & Warren, 2010). These studies also implicate the involvement of top-down acting cortical networks beyond primary auditory cortex, supporting the notion of a hierarchical, but highly integrated neural processing network for analyzing auditory scenes. Moreover, this notion of a hierarchical but highly integrated system is consistent with research combining behavioral and neural measures of central and cognitive processing to predict speech understanding; for example, Anderson et al. (2013) found that the cognitive abilities influence speech understanding not only directly, but also indirectly by mediating central auditory neural processing through top-down modulation.

The purpose of the present study was not to test any of the described models. Rather, the described theoretical insights served as a basis for the selection of our predictor variables, which were chosen to behaviorally measure processing corresponding to the hierarchy of abilities suggested by the theoretical work. Taken together, previous research suggests that speech-in-speech and spatial listening rely on a hierarchy of abilities beginning with the audibility of the speech signal, but extending to processing at multiple higher-order levels including auditory temporal, cognitive, and linguistic processing. Importantly, abilities in these higher-order processing domains have been shown to interact with age even when audibility is controlled.

4.1.4 Hypotheses

To answer our research question regarding whether or not there are age-related differences in LiSN-S outcomes for individuals with good audiometric thresholds in the speech range, we compared the LiSN-S outcomes of a group of listeners aged 65 years or older to the results we obtained for a group of adults 27 years or younger and to the normative data reported previously by Cameron et al. (2011) for listeners up to the age of 60 years. We hypothesized that unlike the earlier results of Glyde, Cameron et al. (2013), LiSN-S outcomes for listeners with good audiograms would be influenced by age and higher-order processing abilities, especially for older listeners. In addition to testing for age-related differences on LiSN-S outcomes, we further investigated the nature of age-related differences by examining contributions to LiSN-S outcomes of (1) audiometric thresholds for typical speech frequencies and for higher frequencies, (2) two tests of ability to detect temporal envelope modulations, (3) two tests of ability to detect differences in periodicity or TFS cues, (4) linguistic tests for vocabulary and sentence-level linguistic closure abilities, (5) a test for cognitive ability comprising items for memory, attention, language fluency, and executive ability, and (6) age. We hypothesized that the
selected measures would be relevant in all LiSN-S conditions, but that their weighting might change depending on the availability of specific cues. For example, cognitive influences might be most pronounced in the condition with the least acoustic cues (i.e., no spatial separation and no voice differences) and influences of F0 discrimination abilities might be more pronounced in conditions that include voice difference cues.

4.2 MATERIALS AND METHODS

4.2.1 Participants

The experiment included two groups of participants, each consisting of 26 adults, either younger (18–27 years, $M_{\text{Age}} = 21.7$ years, $SD = 2.6$ years) or older (66–82 years, $M_{\text{Age}} = 72.0$ years, $SD = 4.3$ years). The younger group had a mean of 15.6 ($SD = 2.1$) years of education and the older group had a mean of 17.2 ($SD = 3.6$) years ($t = 1.9$, $p = 0.06$). Self-reported health was significantly better ($t = -2.5$, $p = 0.02$) for the younger group ($M_{\text{Health}} = 3.4; SD = 0.6$) than for the older group ($M_{\text{Health}} = 2.9; SD = 0.7$) on a 4-point scale (1 = poor, 2 = average, 3 = good, and 4 = excellent). Hearing acuity was tested up to 20,000 Hz. To be eligible for the study, the participants were required to have pure-tone air conduction audiometric thresholds of 25 dB HL or less at frequencies from 250 Hz to 3000 Hz in both the ears, with interaural threshold differences not exceeding 10 dB at more than two adjacent standard octave frequencies from 250 Hz to 8000 Hz; i.e., the participants had no clinically significant interaural asymmetry in hearing acuity. Note that, given the eligibility criteria, the older group in the present study had better hearing than would be considered normal for their age range (ISO, 2000; Echt, Smith, Backscheider Burridge & Spiro III, 2010). All the participants were fluent English speakers and had begun learning English in an English speaking country before the age of 5 years. The participants received monetary compensation of $10 CAD per hour or course credits and provided informed consent before testing. The study was conducted in accordance with the human ethics standards and received approval from the research ethics board of the University of Toronto.

4.2.2 Outcome Measures

The main outcome measures of the study were scores on the different LiSN-S test conditions. First, the LiSN-S test will be described in detail, followed by descriptions of the potential explanatory measures: audiometric measures, measures of auditory temporal processing, and finally, the measures of cognitive and linguistic abilities. All testing was performed in a 3.3 m single-walled sound-attenuating
Industrial Acoustics Co. (International Acoustic Chamber [IAC]) sound booth. The WIN test was administered only to the older group, but all the other tests were administered to both the groups. All the tests of auditory temporal processing were conducted monaurally in the right ear except the GapSU test for gap detection in speech markers (see description later), which was conducted binaurally.

4.2.2.1 Listening in Spatialized Noise–Sentences (LiSN-S) Test

We administered the commercial version of the North-American LiSN-S test (Cameron & Dillon, 2009; Cameron et al., 2009). The test manipulates the availability of spatial separation cues and voice difference cues in four test conditions. In all the conditions, the participants listened to sentences spoken by a female target talker in the presence of two female interfering talkers. The interfering talkers formed a continuous speech stream that was not interrupted between target-sentence presentations. The stimuli were presented binaurally under headphones (Sennheiser HD215). Head-related transfer functions were used to simulate the spatial locations of the talkers such that the target talker was perceived to be located in front of the listener (at 0° azimuth) in all the test conditions. The speech of the interfering talkers was presented such that it would be perceived to be located at the same (0° azimuth) or different (+90° and −90° azimuth) spatial locations and in the same voice (SV) or a different voice (DV) relative to the target talker. The four test conditions were presented in a fixed order, according to the recommendations of Cameron and Dillon (2009): (1) different voices from different locations (DV90), providing both spatial separation and voice difference cues, (2) same voice from different locations (SV90), providing spatial separation but no voice difference cues, (3) different voices at the same spatial location (DV0), providing voice differences but no spatial separation cues, (4) same voice at the same spatial location (SV0), providing neither spatial separation nor voice difference cues.

The listeners were asked to repeat the sentences spoken by the target talker. A 200 ms 1000 Hz tone burst alerted the listener to the presentation of the next target sentence 500 msec before its onset. The unamplified initial speech level was 62 dB SPL and the masker level was fixed at 55 dB SPL. When the listeners correctly repeated more than 50% of the words in the target sentence, the SNR of the next sentence was decreased by 2 dB. When less than half of the words in the target sentence were repeated correctly, the SNR was increased by 2 dB. When the listeners repeated exactly half of the words correctly, the SNR did not change. Changes in SNR were achieved by adjusting the level of the target speech. Because many of the older participants had elevated pure-tone thresholds at high frequencies, we used the prescribed National Acoustics Laboratories–Revised Profound (NAL-

---

Speech in speech

Four test conditions with spatial-separation and/or voice-difference cues

Adaptive procedure

---

1 Phonak Communications AG, Switzerland, product-version number 2.002.
RP) gain amplifier, which is built into the LiSN-S software, to guarantee stimulus audibility for all the participants. The NAL-RP calculates the signal output levels for both the ears separately based on the formula described in Dillon (2001). Amplification was applied equivalently to the speech of the target and interfering talkers.

Four SRTs were calculated, one for each test condition (DV90, SV90, DV0, SV0), as the mean SNR of the presented sentences. However, the first sentences in each condition—up to the first upward reversal in SNR (<50% of the words repeated) or at least five sentences—were considered practice sentences and not used for calculating the outcome SRT. A condition was completed when either the maximum of 30 test sentences had been presented, or the practice sentences plus at least 17 test sentences had been presented and the standard error of the presentation SNRs in the ongoing test condition was less than 1 dB. In addition to the SRT measures, the spatial advantage was calculated by subtracting the SRT in the SV90 condition from the SRT in the SV0 condition. The talker advantage was calculated by subtracting the SRT in the DV0 condition from the SRT in the SV0 condition. The total advantage was calculated by subtracting the SRT in the DV90 condition from the SRT in the SV0 condition.

### 4.2.2 Auditory Temporal Processing

Auditory temporal processing ability was tested with respect to both envelope (gap detection and tones in the noise gaps) and fine-structure cues ($F_0$ difference limen and TFS sensitivity). For the envelope and fine-structure measures, respectively, one test used speech material as test items (gap detection and $F_0$ difference limen), whereas the other test used nonspeech material (tones in noise gaps and TFS sensitivity).
4.2.3.1  *Gap Detection (GapSU)*

We tested the ability of the participants to detect gaps in 40 ms speech markers using the same procedure used by Pichora-Fuller et al. (2006). Recordings of the syllable [su] spoken by an adult female speaker were presented to the participants with a gap of varying durations inserted between the consonant and the vowel at a zero crossing of the waveform of the syllable [su]. A two-alternative forced-choice staircase procedure was used to determine the shortest gap duration that the participant could detect. In each trial, the listener heard one stimulus containing a gap and one stimulus without a gap and had to indicate which of the stimuli contained the gap by pressing a button on a button box. The interstimulus interval was 1 s. Feedback was provided by illuminating the light above the button corresponding to the correct answer. We used a 3-down-1-up procedure to find the 79.9% point on the psychometric function (Levitt, 1971); the initial gap duration was 350 msec. The starting step size was 32 msec and it changed by a factor of 0.5 with each reversal. The gap size decreased after three correct responses and increased after one incorrect response. The stimuli were presented binaurally at 75 dB SPL. The test runs were completed after 12 reversals, and the average of the last eight reversals represented the gap detection threshold for each run. All the participants completed at least three runs. If the thresholds for the participant were continuing to improve on the third run, then up to three more runs were completed until a plateau in performance was observed. The gap detection threshold (in msec) was calculated as the mean threshold of the best two runs. A smaller outcome score indicates better performance.

4.2.3.2  *Tones in Noise Gaps (TiNG)*

Auditory temporal resolution was measured with a test for tonedetection thresholds in noise gaps based on the procedure used by Larsby en Arlinger (1998). Different from Larsby en Arlinger, we used a three-alternative forced-choice paradigm rather than a Békésy tracking procedure to estimate the detection thresholds. Each trial consisted of three stimulus intervals, separated by 500 ms of silence. Two test conditions were administered. The first condition was a reference condition to determine tone detection thresholds in a continuous speech spectrum noise masker. In this condition, each stimulus interval contained an 850 ms long noise burst. The second condition was used to determine tone detection thresholds when the tone was presented in a gap in a masker. In the second condition, each interval consisted of two successive 400 ms noise bursts, separated by a 50 ms gap. In both the conditions, one of the three stimulus intervals contained a pure tone (175 ms) presented at the temporal center of the gap or of the continuous noise; the other intervals contained only the masker. The test was conducted for pure tones at 750 Hz (TiNG750Hz).
and at 3000 Hz (TiNG\textsubscript{3000Hz}). The participant was asked to indicate the interval containing the tone by pressing the corresponding button on a button box. Feedback was provided by the illumination of a light above the correct button.

The speech spectrum noise masker was presented at 65 dB A. The presentation level of the tone was adjusted for each trial in a 2-down-1-up fashion to estimate the 70.7\% correct point on the psychometric function. The starting SNR in the reference condition was 10 dB. The starting SNR in the comparison condition was the participant’s mean outcome SNR of the test runs in the reference condition. Until the first reversal, the level of the tone was decreased in steps of 4 dB. Thereafter, after two correct responses, the level of the tone was decreased by 2 dB. After an incorrect response, the level was increased by 2 dB. The test stopped after six reversals. The tone detection threshold was calculated as the mean SNR over all of the last four reversals. For each test frequency, the participants performed three test runs in each condition and the outcome scores were averaged over all runs per condition. For each frequency, the final TiNG score was calculated by subtracting the pure-tone threshold in gaps from the threshold in continuous noise. Thus, a higher score indicates better performance. Per test frequency, we examined practice effects for TiNG scores on the three test runs with repeated measures analyses of variance, applying Bonferroni corrections for multiple comparisons. No practice effects were observed for the TiNG\textsubscript{750Hz} or for the TiNG\textsubscript{3000Hz} test. Averages of the three test runs per condition were used for the data analyses.

### 4.2.3.3 $F_0$ Difference Limen ($F_0$DL)

This test was used to measure the ability to detect $F_0$ differences in synthesized tokens of the vowel /a/ with five fixed formant frequencies (750, 1050, 2950, 3350, and 3850 Hz) using the same procedure as that used by Vongpaisal en Pichora-Fuller (2007). The stimuli were presented at a level of 80 dB SPL. A three-alternative forced-choice procedure was used. The first token of each trial was used as a reference token. It had a $F_0$ of 120 Hz. One of the two remaining tokens also had a $F_0$ of 120 Hz, whereas the other one had a higher $F_0$. The task of the participant was to press a button on a button box to identify the interval in which the token had the deviating $F_0$. Feedback was given after each response by activating a light above the correct button. The $F_0$ difference between the tokens was 30 Hz for the first trial. Thereafter, a 3-up-1-down adaptive procedure was employed to find the 79.9\% point on the psychometric function (Levitt, 1971), with the difference in $F_0$ being halved after three correct responses in a row and doubled after one incorrect response. After the fifth reversal, the step size decreased to 1.25 times the previous $F_0$ for the increments and 0.8 times the previous $F_0$ for the decrements. The procedure stopped after a total of 16 reversals. The $F_0$DL was calculated.
as the average $F_0$ difference at the last 10 reversals. The participants completed at least three runs. If the thresholds for the participant were continuing to improve on the third run, then up to three more runs were completed until a plateau in performance was observed. The final $F_0$DL score (in Hz) was calculated as the mean of the two best test runs. A smaller outcome score indicates better performance.

4.2.3.4 TFS Sensitivity

We measured the ability of the participants to make use of the temporal fine-structure information with the TFS2 test (Hopkins & Moore, 2011; Moore & Sek, 2009). The test used a two-alternative forced-choice task. Each test trial consisted of two intervals of four successively presented 200 ms complex tones. The tones within the intervals were separated by 100 ms, and the time between the two intervals was 300 ms. One of the intervals contained four harmonic (H) tones (HHHH), whereas the other interval contained two identical harmonic and two identical inharmonic (I) tones (HIHI). The H and I tones were passed through a band-pass filter before presentation. The filter had a flat region with a width of $1F_0$ of the H tone and skirts with a slope of 30 dB/octave. The filter was centered at $9F_0$, that is, flanking harmonics were decreased in the level along the skirts of the filter. We performed TFS2 tests with an $F_0$ of 83.3 Hz and 333.3 Hz, respectively, such that the center frequency in the passband was either 749.7 Hz (TFS$_{750\text{Hz}}$) or 2999.7 Hz (TFS$_{3000\text{Hz}}$), approximating the test frequencies of the TiNG test. The tones were presented at 70 dB SPL in threshold equalizing noise (55 dB SPL), designed to give equal masked thresholds at 200 to 16,000 Hz (Moore, 2007).

The task of the participant was to identify the interval, in which the tones alternated from harmonic to inharmonic. The alternation can be perceived as a change in the pitch. The H and I tones had the same frequency difference between partials and the same envelope-repetition rate. However, relative to the frequencies of the harmonics of the H tone, the corresponding components of the I tones were shifted upwards by a given number of Hz which was manipulated using a 2-down-1-up adaptive procedure to determine the smallest detectable frequency shift for each participant, estimating the 70.7% correct point on the psychometric function (Levitt, 1971). The initial frequency shift was set to $0.5F_0$ of the H tone, which is the biggest possible shift. After two correct responses, the frequency shifts were reduced by a factor of 1.253 until the first reversal. Between the first and the second reversal, the frequency shifts were increased by a factor of 1.252. Thereafter, the frequency shifts were fixed to a factor of 1.25. The test terminated after eight reversals and the outcome score was calculated as the geometric mean of the frequency shifts at the last six reversals. Further details regarding the test can be found in Moore en Sek (2009) and Hopkins en Moore (2011). The participants
completed three test runs. The average of the two best outcome scores was used for the analyses. A smaller outcome score indicates better performance.

4.2.4 Cognitive and Linguistic Abilities

There were three measures of cognitive and linguistic ability, including a cognitive screening test, a measure of vocabulary, and a test for linguistic closure.

4.2.4.1 Vocabulary

Participants completed the Mill Hill vocabulary test (Raven, 1962). For each of the 20 target words, the participant scored a point by selecting the correct synonym from a set of six alternatives, for a possible total score of 20.

4.2.4.2 Cognitive Abilities

The participants were administered the Montreal cognitive assessment (MoCA) test (Nasreddine et al., 2005). The test includes tasks to evaluate visuospatial/executive abilities, naming abilities, attention and short-term memory, sentence repetition and language fluency, abstraction and delayed recall, and orientation in the time and space. A total score is calculated as the sum of all the test items with a maximum score of 30. When using the MoCA as a screening tool, those with a total score >=26 are classified as having normal cognition. In the present study, we used the MoCA total score as a continuous variable in our analyses with higher scores representing better performance.

4.2.4.3 Text Reception Threshold (TRT)

We assessed linguistic closure abilities with the TRT (Zekveld et al., 2007) using the TRT\textsubscript{500} test (Besser et al., 2012). During this test, the sentences in red font (26 point) were presented on a 17 inch computer screen of the type NEC MultiSync LCD 1760 NX placed at a distance of about 60cm in front of the participant. Each sentence was presented by adding one word at the time at a rate of 250 ms per word, until the sentence was complete. It then remained on the screen for another 500 ms. The text was partially masked with a pattern of black vertical bars. The participants were asked to read the sentences aloud filling in the missing information as accurately as possible. The initial percentage of masking was 60%. It was then adapted for every sentence in a 1-up-1-down test procedure. When the participant repeated a sentence entirely correctly, the next sentence was presented with 6% more masking; otherwise it was presented with 6% less masking. Every sentence was presented only once, except for the first sentence.
which was repeated with a decrease in the masking of 12% for every repetition until the response of the participant was correct. We conducted the TRT test with sentences from the speech perception in noise (SPIN) test corpus (Kalikow, Stevens & Elliott, 1977). We split the original SPIN sentence lists into two lists containing 25 low-context or 25 high-context sentences, respectively. Each participant was tested with three low-context lists and three high-context lists. The outcome score is the TRT or the percentage of the unmasked text at which the participants were able to read on an average 50% of the sentences correctly; that is, a lower test score indicates better performance because more was read with a smaller percentage of the un-masked text. Per test condition, we examined practice effects on the TRT tests for the three test runs with repeated measures analyses of variance, applying Bonferroni corrections for multiple comparisons. For the TRT HIGH, no practice effects were observed and the average of all the three test runs was included in the data analyses. TRT scores in the low-context condition improved significantly after the second test run in both the groups. Therefore, only the third run of TRT LOW was used in the analyses.

4.3 RESULTS

4.3.1 Effect of Age on LiSN-S Results

The older group had poorer LiSN-S performance than the younger group in all the four test conditions Table 4.1. As depicted in Figure 4.1, the pattern of results was the same for both groups: LiSN-S scores were best for the DV90 condition (both cues) and second best for the SV90 condition (only spatial cues), followed by the DV0 (only voice cues) condition. The scores were poorest for the SV0 condition (no cue).

An analysis of variance (ANOVA) with age as a between-subjects factor with two levels (younger, older) and LiSN-S condition as a within-subjects factor with four levels (DV90, SV90, DV0, SV0) confirmed the observed pattern of results. There were significant main effects of group, $F(1, 50) = 101.092, p < 0.001$, and LiSN-S test condition, $F(3, 150) = 940.217, p < 0.001$, as well as a significant two-way interaction between group and LiSN-S test condition, $F(3, 150) = 24.726, p < 0.001$. Pairwise comparisons using $t$ tests confirmed that the younger adults outperformed the older adults in all conditions and all conditions within the groups were significantly different from each other ($ps < 0.05$), with the size of the age effect differing across the conditions.

As shown in Figure 4.2, for the LiSN-S advantage measures, the older participants had a smaller talker advantage, spatial advantage, and total advantage compared with the younger participants. This
Table 4.1: Mean scores (M), SDs, minimum (Min) and maximum (Max) scores, and group differences for the LiSN-S outcome measures, organized by age group.

<table>
<thead>
<tr>
<th>LiSN-S Measure</th>
<th>Younger</th>
<th></th>
<th>Older</th>
<th></th>
<th>Group Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
<td>Min</td>
<td>Max</td>
<td>M</td>
</tr>
<tr>
<td>DV90 (dB SNR)</td>
<td>−17.8</td>
<td>2.1</td>
<td>−21.5</td>
<td>−12.8</td>
<td>−12.5</td>
</tr>
<tr>
<td>SV90 (dB SNR)</td>
<td>−16.8</td>
<td>2.1</td>
<td>−21.1</td>
<td>−13.2</td>
<td>−10.4</td>
</tr>
<tr>
<td>DV0 (dB SNR)</td>
<td>−12.9</td>
<td>2.0</td>
<td>−19.5</td>
<td>−8.5</td>
<td>−8.9</td>
</tr>
<tr>
<td>SV0 (dB SNR)</td>
<td>−2.6</td>
<td>1.4</td>
<td>−6.0</td>
<td>−0.5</td>
<td>−0.7</td>
</tr>
<tr>
<td>Talker advantage (dB)</td>
<td>10.5</td>
<td>1.6</td>
<td>7.9</td>
<td>14.2</td>
<td>8.2</td>
</tr>
<tr>
<td>Spatial advantage (dB)</td>
<td>14.1</td>
<td>2.8</td>
<td>7.9</td>
<td>19.9</td>
<td>9.6</td>
</tr>
<tr>
<td>Total advantage (dB)</td>
<td>15.2</td>
<td>1.9</td>
<td>11.3</td>
<td>18.9</td>
<td>11.8</td>
</tr>
</tbody>
</table>

N = 26 per group.

DV0, SRT for different voices at ±0°; DV90, SRT for different voices at ±90°; LiSN-S, listening in spatialized noise–sentences; SNR, signal-to-noise ratio; SRT, speech reception threshold; SV0, SRT for same voice at ±0°; SV90, SRT for same voice at ±90°
Figure 4.1: Mean speech reception thresholds (SRTs) are shown for both the age groups in the four listening in spatialized noise–sentences test conditions. The thicker black horizontal bar indicates the mean group outcome for the respective measure. The two thinner black horizontal bars indicate group SDs for the respective measure. Y = younger group, O = older group, N = 26 per group. The diamonds represent individual outcome scores. SNR indicates signal-to-noise ratio.

description was confirmed by an ANOVA with age as a between-subjects factor with two levels (younger, older) and LiSN-S advantage condition as a within-subjects factor with three levels (talker advantage, spatial advantage, total advantage). There were significant main effects of group, $F(1, 50) = 51.293, p < 0.001$, and LiSN-S advantage condition, $F(2, 100) = 113.350, p < 0.001$, as well as a significant two-way interaction between the group and LiSN-S advantage condition, $F(2, 100) = 7.902, p < 0.005$. Pairwise comparisons using $t$ tests confirmed that the younger adults outperformed the older adults in all the conditions and all the conditions were significantly different from each other ($ps < 0.05$) with the size of the age effect differing across advantage measures.

As discussed in Section 4.1, there are many factors that might drive the observed differences in the LiSN-S outcomes between the two the age groups. We examined the relationship between LiSN-S outcomes and age, pure-tone thresholds, measures of temporal processing, and measures of cognitive and linguistic abilities.
Table 4.2: Audiometric thresholds in dB HL for both the ears, averaged per frequency for each age group

<table>
<thead>
<tr>
<th>Frequency (Hz)</th>
<th>Younger</th>
<th>Older</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Right</td>
<td>Left</td>
</tr>
<tr>
<td>250</td>
<td>1.9 (5.1)</td>
<td>2.1 (4.3)</td>
</tr>
<tr>
<td>500</td>
<td>0.5 (4.0)</td>
<td>-0.4 (5.1)</td>
</tr>
<tr>
<td>1000</td>
<td>-1.5 (4.6)</td>
<td>-2.2 (3.3)</td>
</tr>
<tr>
<td>1500</td>
<td>-0.2 (3.9)</td>
<td>-0.4 (3.8)</td>
</tr>
<tr>
<td>2000</td>
<td>-0.6 (4.5)</td>
<td>-2.3 (3.2)</td>
</tr>
<tr>
<td>3000</td>
<td>1.4 (4.4)</td>
<td>-1.2 (5.7)</td>
</tr>
<tr>
<td>4000</td>
<td>-0.2 (5.2)</td>
<td>-1.4 (5.0)</td>
</tr>
<tr>
<td>6000(1)</td>
<td>2.1 (4.8)</td>
<td>0.8 (4.9)</td>
</tr>
<tr>
<td>8000</td>
<td>3.5 (5.8)</td>
<td>1.7 (6.2)</td>
</tr>
<tr>
<td>9000</td>
<td>7.1 (6.7)</td>
<td>6.2 (6.7)</td>
</tr>
<tr>
<td>10,000</td>
<td>-0.2 (7.7)</td>
<td>2.7 (6.8)</td>
</tr>
<tr>
<td>11,200(2)</td>
<td>1.7 (6.9)</td>
<td>4.6 (9.4)</td>
</tr>
<tr>
<td>12,500(3)</td>
<td>-1.0 (6.0)</td>
<td>-0.4 (8.5)</td>
</tr>
</tbody>
</table>

Numbers in parentheses indicate SDs. N = 26 per group, if not stated otherwise.

(1) Audiometric thresholds at 6000 Hz were only assessed for younger participants if the difference for the thresholds at 4000 and 8000 Hz was at least 20 dB. This was the case for six participants only. For the remaining younger participants, thresholds at 6000 Hz were obtained by averaging thresholds at 4000 and 8000 Hz.

(2) Thresholds at 11,200 Hz with N = 25 for the older group, because the threshold could not be measured for one of the older adults.

(3) Thresholds at 12,500 Hz with N = 20 for the older group, because the threshold could not be measured for six of the older adults.
4.3.2 Associations of LiSN-S With Age and Audiometric Measures

While age is a critical factor when comparing the test performance of older and younger people, it might not be as crucial within the age groups. We examined the associations of LiSN-S outcomes with age and two measures of pure-tone hearing acuity (4FAHL, PTA_{HIGH}). Table 4.2 provides details about the hearing acuity of the participants for frequencies up to 12,500 Hz (at frequencies higher than 12,500 Hz, the thresholds for more than half of the older individuals could not be measured). Results of the correlational analyses are given in Table 4.3. In Figure 4.3, the relationship between the pure-tone measures and the LiSN-S SRTs is plotted by age group for 4FAHL and for PTA_{HIGH} in the same style as Figure 4 of the article by Glyde, Cameron et al. (2013).

For the younger group, neither age nor PTA_{HIGH} were significantly associated with LiSN-S outcome scores. Surprisingly, worse 4FAHL was statistically significantly associated with better outcomes on LiSN-S conditions without spatial separation (SV0 and DV0); however, this association is unlikely to be of practical significance given that the mean 4FAHL value in the younger group was below 0 dB HL (range −6.3 to 6.3). For the present sample of older adults whose pure-tone thresholds were clinically normal at frequencies up to and including...
Table 4.3: Pearson correlations between LiSN-S outcomes and age and audiological measures.

<table>
<thead>
<tr>
<th>LiSN-S Measure</th>
<th>Younger</th>
<th></th>
<th></th>
<th>Older</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Age</td>
<td>4FAHL</td>
<td>PTA_{HIGH}</td>
<td>Age</td>
<td>4FAHL</td>
<td>PTA_{HIGH}</td>
</tr>
<tr>
<td>DV90</td>
<td>0.07</td>
<td>-0.31</td>
<td>-0.14</td>
<td>0.08</td>
<td>0.30</td>
<td>0.78**</td>
</tr>
<tr>
<td>SV90</td>
<td>-0.07</td>
<td>-0.29</td>
<td>0.36</td>
<td>0.31</td>
<td>0.42*</td>
<td>0.81**</td>
</tr>
<tr>
<td>DV0</td>
<td>0.07</td>
<td>-0.53**</td>
<td>0.09</td>
<td>0.10</td>
<td>0.39*</td>
<td>0.72**</td>
</tr>
<tr>
<td>SV0</td>
<td>0.18</td>
<td>-0.42*</td>
<td>-0.16</td>
<td>0.54**</td>
<td>0.32</td>
<td>0.19</td>
</tr>
<tr>
<td>Spatial advantage</td>
<td>0.14</td>
<td>0.01</td>
<td>-0.34</td>
<td>-0.08</td>
<td>-0.31</td>
<td>-0.8**</td>
</tr>
<tr>
<td>Talker advantage</td>
<td>0.16</td>
<td>-0.02</td>
<td>0.10</td>
<td>0.17</td>
<td>-0.23</td>
<td>-0.62**</td>
</tr>
<tr>
<td>Total advantage</td>
<td>0.06</td>
<td>0.03</td>
<td>-0.27</td>
<td>0.18</td>
<td>-0.16</td>
<td>-0.70**</td>
</tr>
</tbody>
</table>

N = 26 per group. *Two-tailed significant at p < 0.05 level. **Two-tailed significant at p < 0.01 level.
DV0, SRT for different voices at ±0°; DV90, SRT for different voices at ±90°; 4FAHL, 4-frequency average hearing loss for 500, 1000, 2000, and 4000 Hz; LiSN-S, listening in spatialized noise–sentences; PTA_{HIGH}, mean pure-tone hearing threshold for 6000, 8000, 9000, and 10,000 Hz; SNR, signal-to-noise ratio; SRT, speech reception threshold; SV0, SRT for same voice at ±0°; SV90, SRT for same voice at ±90°.

3000 Hz, the correlations between LiSN-S outcomes and 4FAHL were relatively weak and they only reached significance for SV90 and DV0, the conditions where only one type of cue, either spatial separation or voice difference cue, was available. However, all LiSN-S outcomes for the older group were highly correlated with the average pure-tone thresholds at high frequencies (6000, 8000, 9000, and 10,000 Hz), with the exception of the SRT in the SV0 condition, in which neither cue was available. Interestingly, the SV0 condition was the only LiSN-S outcome associated with age.
Figure 4.3: Listening in spatialized noise–sentences speech reception thresholds (SRTs) for younger (A and C) and older (B and D) participants for all the four test conditions plotted with 4-frequency average hearing loss (4FAHL, for 500, 1000, 2000, and 4000 Hz; A and B) and mean pure-tone hearing threshold (PTA_{HIGH} for 6000, 8000, 9000, 10,000 Hz; C and D).
<table>
<thead>
<tr>
<th>Measure</th>
<th>Younger</th>
<th></th>
<th></th>
<th>Older</th>
<th></th>
<th></th>
<th>Group Diff.</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years)</td>
<td>21.7</td>
<td>2.6</td>
<td>18</td>
<td>27</td>
<td>72.0</td>
<td>4.3</td>
<td>66</td>
<td>82</td>
<td>50.3</td>
</tr>
<tr>
<td>4FAHL (dB HL)</td>
<td>-0.4</td>
<td>3.1</td>
<td>-6.3</td>
<td>6.3</td>
<td>13.0</td>
<td>5.1</td>
<td>1.3</td>
<td>22.5</td>
<td>13.4</td>
</tr>
<tr>
<td>PTAHIGH (dB HL)</td>
<td>3.5</td>
<td>5.2</td>
<td>-5.0</td>
<td>15.0</td>
<td>51.8</td>
<td>12.8</td>
<td>31.3</td>
<td>80.0</td>
<td>48.3</td>
</tr>
<tr>
<td>WIN (dB SNR)</td>
<td>dnt</td>
<td>dnt</td>
<td>dnt</td>
<td>dnt</td>
<td>7.6</td>
<td>2.3</td>
<td>3.6</td>
<td>12.4</td>
<td>na</td>
</tr>
<tr>
<td>TFS750Hz (Hz)(1)</td>
<td>6.3</td>
<td>2.9</td>
<td>2.0</td>
<td>11.2</td>
<td>13.1</td>
<td>4.1</td>
<td>7.1</td>
<td>19.9</td>
<td>6.8</td>
</tr>
<tr>
<td>TFS3000Hz (Hz)(2)</td>
<td>20.5</td>
<td>9.4</td>
<td>8.1</td>
<td>44.7</td>
<td>38.7</td>
<td>13.4</td>
<td>14.4</td>
<td>59.8</td>
<td>18.2</td>
</tr>
<tr>
<td>F0DL (Hz)</td>
<td>0.9</td>
<td>1.9</td>
<td>0.4</td>
<td>9.0</td>
<td>2.7</td>
<td>6.4</td>
<td>1.1</td>
<td>21.8</td>
<td>0.5</td>
</tr>
<tr>
<td>TiNG750Hz (dB)(3)</td>
<td>18.7</td>
<td>4.7</td>
<td>8.3</td>
<td>26.7</td>
<td>16.4</td>
<td>4.2</td>
<td>10.9</td>
<td>27.0</td>
<td>2.3</td>
</tr>
<tr>
<td>TiNG3000Hz (dB)(4)</td>
<td>19.2</td>
<td>4.2</td>
<td>9.7</td>
<td>25.4</td>
<td>12.5</td>
<td>3.8</td>
<td>5.2</td>
<td>20.3</td>
<td>6.7</td>
</tr>
<tr>
<td>GapSU (ms)(5)</td>
<td>8.3</td>
<td>9.5</td>
<td>4.6</td>
<td>39.1</td>
<td>21.9</td>
<td>23.2</td>
<td>6.3</td>
<td>100.5</td>
<td>0.4</td>
</tr>
<tr>
<td>Vocabulary (/20)</td>
<td>13.4</td>
<td>2.4</td>
<td>7</td>
<td>20</td>
<td>15.5</td>
<td>2.0</td>
<td>12</td>
<td>19</td>
<td>2.1</td>
</tr>
<tr>
<td>MoCA (/30)</td>
<td>28.1</td>
<td>1.2</td>
<td>26</td>
<td>30</td>
<td>28.0</td>
<td>3.1</td>
<td>18</td>
<td>30</td>
<td>0.1</td>
</tr>
<tr>
<td>TRTLOW (%)</td>
<td>67.3</td>
<td>3.3</td>
<td>61.0</td>
<td>74.1</td>
<td>74.1</td>
<td>6.9</td>
<td>63.7</td>
<td>91.6</td>
<td>6.8</td>
</tr>
<tr>
<td>TRTHIGH (%)</td>
<td>67.7</td>
<td>3.3</td>
<td>62.8</td>
<td>75.2</td>
<td>73.8</td>
<td>5.5</td>
<td>65.6</td>
<td>83.4</td>
<td>6.1</td>
</tr>
</tbody>
</table>

*Table continues on next page*
Table 4.4: Continued from previous page

Absolute group differences (Group Diff.) are provided along with \( p \) values and \( t \) values for \( t \) tests of the group differences. \( N = 26 \) per group, if not stated otherwise. Results printed in italics are based on log-transformed variables.

1. Results for TFS\(_{750\text{Hz}}\) with \( N = 24 \) in the younger group as a result of the exclusion of one outlier and noninterpretable data for one participant. Results with \( N = 17 \) for the older group as a result of the noninterpretable data for nine participants.

2. Results for TFS\(_{3000\text{Hz}}\) with \( N = 18 \) for the older group as a result of the noninterpretable data for seven participants and the exclusion of one outlier.

3. Results for TiNG\(_{750\text{Hz}}\) with \( N = 25 \) in the younger group as a result of the exclusion of one outlier.

4. Results for TiNG\(_{3000\text{Hz}}\) with \( N = 25 \) in the older group as a result of the exclusion of one outlier.

5. All auditory test results are reported for the right ear, except Gap\(_{SU}\), which was tested binaurally.

6. Results for TRT measures are with \( N = 24 \) in the younger group as a result of the exclusion of two outliers per condition.

dnt, did not test; F\(_0\)DL, F\(_0\) difference limen; 4FAHL, 4-frequency average hearing loss for 500, 1000, 2000, and 4000 Hz; Gap\(_{SU}\), gap detection in speech markers; MoCA, Montreal Cognitive Assessment; na, not applicable; PTA\(_{HIGH}\), mean pure-tone hearing threshold for 6,000, 8,000, 9,000, and 10,000 Hz; SNR, signal-to-noise ratio; TFS, temporal fine-structure sensitivity; TiNG, detection of tones in noise gaps; TRT, text-reception threshold; WIN, words in noise test.
4.3.3 Multiple Linear Regression Analyses

We performed linear regression analyses to examine how much of the variance in LiSN-S outcomes could be explained by models with the combinations of the included auditory, cognitive, and linguistic measures as well as age and hearing acuity. Table 4.4 shows the test outcomes (excluding outliers\(^2\)) for the younger and the older groups on the set of predictor variables. Some variables (\(F_0\)DL and \(Gap_{SU}\), in both groups and MoCA in the older group) were log10-transformed\(^3\), because they had non-normal distributions; descriptive statistics before transformation are reported in Table 4.4, but the transformed, normally distributed variables were used for the group comparisons reported in the last three columns and all the subsequent analyses. The age groups differed significantly on all the measures except temporal resolution at 750 Hz (\(TiNG_{750Hz}\)) and on the MoCA test.

To examine potential cases of collinearity between the predictor variables in each participant group, we performed correlation analyses including all the predictors, the detailed results of which are not reported here. Variables whose correlation coefficient was \(\geq 0.7\) were classified as collinear. This was the case for two correlations in the older group: \(F_0\)DL with TFS\(_{750Hz}\) (\(r = 0.71\)) and TRT\(_{LOW}\) with TRT\(_{HIGH}\) (\(r = 0.85\)). We transformed each of these variables to z scores, which were then used to obtain the combined z-score variables Z-\(F_0\)DL–TFS\(_{750Hz}\) and Z-TRT, respectively. Z-\(F_0\)DL–TFS\(_{750Hz}\) replaced the original variables \(F_0\)DL and TFS\(_{750Hz}\) as predictors for the regression analyses in the older group. Equally, TRT\(_{LOW}\) and TRT\(_{HIGH}\) were replaced by the Z-TRT variable in the older group. No collinearities were observed in the younger group.

For each group, we created one prognostic regression model for each of the LiSN-S SRTs (DV90, SV90, DV0, SV0) and each of the LiSN-S advantage measures (spatial advantage, talker advantage, total advantage). For each prognostic model, we made an individual selection of potential predictors to be included in the initial model, based on the preparatory regression analyses using association mod-

\(^2\) Outliers were removed from the data set if their value differed by at least 3 SD from the group mean. This was the case for the TFS\(_{750Hz}\) score of one participant and the \(TiNG_{750Hz}\) score of another participant and the TRT\(_{HIGH}\) and TRT\(_{LOW}\) scores of two participants in the younger group. In the older group, the TFS\(_{3000Hz}\) score of one participant and the \(TiNG_{3000Hz}\) score of another participant were classified as outliers.

\(^3\) The distributions of the variables \(F_0\)DL and \(Gap\) were positively skewed in both groups and had outliers in the young group according to the definition given earlier. Rather than excluding the outliers, we applied log10-transformations to account for the skewness and preserve all the data points. The distribution of the MoCA was negatively skewed in the old group. Therefore, we first reversed and then log10-transformed the variable. The transformations yielded normal variable distributions. The transformed variables \(F_0\)DL\(_{lg10}\) and \(Gap_{SU}_{lg10}\) were used for the analyses in both the groups and MoCA\(_{R_{lg10}}\) was used for the analyses in the older group.
els to examine each variable’s association with each of the outcome measures separately. Variables whose association with the respective outcome measure had a p value > 0.2 were eliminated because they clearly had no linear relationship with the LiSN-S outcome measures. Additional variables were excluded if the direction of the associations seemed counter-intuitive. For the younger group, all the initial models for the LiSN-S SRTs would have included 4FAHL as a predictor according to the criteria stated above, but it was eliminated from the regression analyses because it was negatively associated with the LiSN-S SRTs such that better LiSN-S SRTs were related to worse hearing acuity, and because the mean 4FAHL value of the younger group was extremely good (< 0 dB HL, range −6.3 to 6.3 dB HL) and such small differences in this range are not clinically meaningful. Similarly, other counter-intuitive associations resulted in the elimination of variables in analyses for the younger group; poorer TFS\textsubscript{750Hz} scores were related to better SV90 thresholds and larger talker advantage and spatial advantage in the younger group. Poorer TFS\textsubscript{3000Hz} scores were related to better DV0 and SV90 thresholds and better talker advantage and spatial advantage; poorer Gap\textsubscript{SU-lg10} scores were related to better DV0 thresholds and a larger talker advantage. Likewise, for the older group, poorer TiNG\textsubscript{750Hz} scores were related to better SV90 thresholds and larger spatial advantage. It is unlikely that these negative associations reflect meaningful relationships between the measures. We therefore decided to exclude the 4FAHL, TFS\textsubscript{750Hz}, and TFS\textsubscript{3000Hz} variables from the regression analyses in the younger group and TiNG\textsubscript{750Hz} from the analyses in the older group. Table 4.5 lists the initial models for each outcome measure for each age group.

Using a backward selection procedure, the selected variables were excluded stepwise; at each step removing the variable with the highest p value, until only variables remained that had a p value < 0.157, following the recommendations by Royston, Moons, Altman en Vergouwe (2009). The final models are reported in Table 4.6. Next to stating the variance accounted for by each model’s ultimate set of predictors (R\textsuperscript{2} total), the table also shows the variance in LiSN-S SRT that each predictor would account for on its own in a model including no other predictors (R\textsuperscript{2} per predictor).

For the younger group, TRT\textsubscript{LOW} predicted outcomes for all LiSN-S SRTs and alone it accounted for approximately 15% of the variance for each SRT. Additionally, DV90 thresholds were predicted by TiNG\textsubscript{750Hz} scores, and DV0 thresholds were predicted by the vocabulary score, but TiNG\textsubscript{750Hz} and vocabulary added to the explained variance in LiSN-S DV90 and DV0 thresholds only in combination with TRT\textsubscript{LOW}. Vocabulary alone accounted for 23.5% of the variance in the talker advantage. The spatial advantage was predicted by PTA\textsubscript{HIGH}, which accounted for 34.4% of the variance.

For the older group, more variance was explained for the SRTs
in conditions with spatial separation cues (68.1% for the DV90 and 74.0% for the SV90), while less was explained for the SRT conditions without spatial separation cues (52.1% for the DV0 and 59.6% for the SV0). Similarly, the models accounted for more variance in the spatial advantage measure (63.9%) than in the talker advantage measure (39.0%). Age was a significant predictor of LiSN-S thresholds in the SV0 condition whereas PTA_{HIGH} was the best predictor of LiSN-S SRTs in all the conditions except SV0. Furthermore, PTA_{HIGH} was the only predictor of all the advantage measures and DV0 SRTs (only voice cues available). Performance on the MoCA test explained variance in all the conditions except DV0\(^4\). No other variables were included in the final models.

\(^4\) For the interpretation of the contribution of the MoCA should be kept in mind that MoCA scores were reversed and log10 transformed for the analyses such that a positive \(B\) indicates that higher (worse) reversed MoCA scores were associated with higher (worse) LiSN-S thresholds. As a result of the log transformation, it makes sense to look at the MoCA variable in terms of percentage changes. Decreasing the original MoCA score by 10%, for example from 30 to 27, will give an increase in LiSN-S SNR of \(\log_{10}(1.1) \times B = 0.41 \times B\) dB
Table 4.5: Selection of variables included in the initial regression models of the backward selection procedure for each outcome measure, organized by the age group

<table>
<thead>
<tr>
<th>Outcome Measure</th>
<th>Younger</th>
<th>Older</th>
</tr>
</thead>
<tbody>
<tr>
<td>DV90</td>
<td>TRT_LOW, TiNG_750Hz</td>
<td>4FAHL, PTA_HIGH, MoCA_Rlg10, Z-TRT</td>
</tr>
<tr>
<td>SV90</td>
<td>PTA_HIGH, Vocabulary, TRT_LOW</td>
<td>Age, 4FAHL, PTA_HIGH, WIN, MoCA_Rlg10, Z-TRT</td>
</tr>
<tr>
<td>DV0</td>
<td>Vocabulary, TRT_LOW</td>
<td>4FAHL, PTA_HIGH, MoCA_Rlg10, Z-TRT</td>
</tr>
<tr>
<td>SV0</td>
<td>TRT_LOW</td>
<td>Age, 4FAHL, WIN, MoCA_Rlg10</td>
</tr>
<tr>
<td>Spatial advantage</td>
<td>PTA_HIGH, Vocabulary</td>
<td>4FAHL, PTA_HIGH, Z-TRT</td>
</tr>
<tr>
<td>Talker advantage</td>
<td>Vocabulary</td>
<td>PTA_HIGH</td>
</tr>
<tr>
<td>Total advantage</td>
<td>PTA_HIGH, TiNG_750Hz</td>
<td>PTA_HIGH, Z-TRT</td>
</tr>
</tbody>
</table>

DV0, SRT for different voices at ±0°; DV90, SRT for different voices at ±90°; 4FAHL, 4-frequency average hearing loss for 500, 1000, 2000, and 4000 Hz; MoCA\_Rlg10, Montreal Cognitive Assessment, reversed and log10-transformed; PTA\_HIGH, mean pure-tone hearing threshold for 6000, 8000, 9000, and 10,000 Hz; SRT, speech reception threshold; SV0, SRT for same voice at ±0°; SV90, SRT for same voice at ±90°; TiNG, detection of tones in noise gaps; TRT, text reception threshold; WIN, words in noise test; Z-TRT, z-transformed TRT scores.
Table 4.6: Predictive regression models for the LiSN-S measures by age group. Presentation of regression coefficients ($B$), standardized regression coefficients ($\beta$), 95% confidence intervals (95% CI), levels of significance ($p$), and $R^2$ values per predictor ($R^2$ per pred.) and for the whole model ($R^2$ total)

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Younger</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th>Older</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$B$</td>
<td>$\beta$</td>
<td>95% CI</td>
<td>$p$</td>
<td>$R^2$ p.</td>
<td>$R^2$ total</td>
<td>$B$</td>
<td>$\beta$</td>
<td>95% CI</td>
<td>$p$</td>
</tr>
<tr>
<td>TiNG$_{750\text{Hz}}$ (1)</td>
<td>- .190</td>
<td>- .385</td>
<td>[- .385, .005]</td>
<td>.056</td>
<td>(.265)</td>
<td>.286</td>
<td>PTAHIGH</td>
<td>.113</td>
<td>.702</td>
<td>[.072, .153]</td>
</tr>
<tr>
<td>TRT$_{LOW}^{(2)}$</td>
<td>.223</td>
<td>.340</td>
<td>[- .036, .482]</td>
<td>.088</td>
<td>.139</td>
<td></td>
<td>MoCA$_{Rigto}$</td>
<td>1.917</td>
<td>.283</td>
<td>[.203, 3.631]</td>
</tr>
<tr>
<td><strong>Outcome: SV90</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TRT$_{LOW}^{(2)}$</td>
<td>.245</td>
<td>.381</td>
<td>[- .018, .508]</td>
<td>.067</td>
<td>.145</td>
<td>.145</td>
<td>PTAHIGH</td>
<td>.127</td>
<td>.732</td>
<td>[.087, .166]</td>
</tr>
<tr>
<td>MoCA$_{Rigto}$</td>
<td>2.159</td>
<td>.296</td>
<td>[.491, 3.826]</td>
<td>.013</td>
<td>.244</td>
<td></td>
<td>MoCA$_{Rigto}$</td>
<td>1.797</td>
<td>.557</td>
<td>[.908, 2.686]</td>
</tr>
<tr>
<td><strong>Outcome: DV0</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vocabulary</td>
<td>-.451</td>
<td>-.463</td>
<td>[- .796, -.107]</td>
<td>.013</td>
<td>.100</td>
<td>.403</td>
<td>PTAHIGH</td>
<td>.107</td>
<td>.722</td>
<td>[.063, .150]</td>
</tr>
<tr>
<td>TRT$_{LOW}^{(2)}$</td>
<td>.241</td>
<td>.381</td>
<td>[.018, .464]</td>
<td>.036</td>
<td>.192</td>
<td></td>
<td>MoCA$_{Rigto}$</td>
<td>1.797</td>
<td>.557</td>
<td>[.908, 2.686]</td>
</tr>
<tr>
<td><strong>Outcome: SV0</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TRT$_{LOW}^{(2)}$</td>
<td>.158</td>
<td>.404</td>
<td>[.000, .317]</td>
<td>.050</td>
<td>.163</td>
<td>.163</td>
<td>MoCA$_{Rigto}$</td>
<td>1.797</td>
<td>.557</td>
<td>[.908, 2.686]</td>
</tr>
<tr>
<td>Age</td>
<td>.109</td>
<td>.478</td>
<td>[.046, .172]</td>
<td>.002</td>
<td>.289</td>
<td></td>
<td>MoCA$_{Rigto}$</td>
<td>1.797</td>
<td>.557</td>
<td>[.908, 2.686]</td>
</tr>
</tbody>
</table>

Table continues on next page
Table 4.6: Continued from previous page

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Younger</th>
<th></th>
<th>Older</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$B$</td>
<td>$\beta$</td>
<td>95% CI</td>
<td>$p$</td>
</tr>
<tr>
<td><strong>Outcome: spatial advantage</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PTA$_{HIGH}$</td>
<td>$-.214$</td>
<td>$-.344$</td>
<td>$[-.460, .032]$</td>
<td>$.085$</td>
</tr>
<tr>
<td><strong>Outcome: talker advantage</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vocabulary</td>
<td>$.311$</td>
<td>$.485$</td>
<td>$[.075, .548]$</td>
<td>$.012$</td>
</tr>
<tr>
<td><strong>Outcome: total advantage</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TiNG$_{750Hz}$</td>
<td>$.132$</td>
<td>$.318$</td>
<td>$[-.038, .303]$</td>
<td>$.122$</td>
</tr>
</tbody>
</table>

$N = 26$ per group.

(1) $p$ value $< .157$ only in combination with TRT$_{LOW}$. Therefore, the $R^2$ per pred. value is given in parentheses.

(2) Models with TRT$_{LOW}$ are with $N = 24$ in the younger group as a result of the exclusion of two outliers.

DV0, SRT for different voices at $\pm 0^\circ$; DV90, SRT for different voices at $\pm 90^\circ$; MoCA$_{Rlg10}$, Montreal Cognitive Assessment, reversed and log$_{10}$-transformed; PTA$_{HIGH}$, mean pure-tone hearing threshold for 6000, 8000, 9000, and 10,000 Hz; SRT, speech reception threshold; SV0, SRT for same voice at $\pm 0^\circ$; SV90, SRT for same voice at $\pm 90^\circ$; TiNG, detection of tones in noise gaps; TRT, text reception threshold.
In the present study, we examined the effect of age on the ability to make use of voice difference cues and spatial separation cues in speech-in-speech listening on the LiSN-S test. We compared the performance of the two groups with good audiograms, one group of participants aged 65 years and older and the other a younger group (18–27 years). Within the age groups, contributions of audiometric thresholds and several suprathreshold factors to LiSN-S outcomes were also examined. The discussion of the results is structured in the following way: First, the age differences in LiSN-S outcomes are discussed. Thereafter, contributions of audiometric thresholds to LiSN-S outcomes are discussed, followed by contributions of auditory temporal processing, and last, the contributions of cognitive and linguistic factors. Finally, a summary and general conclusions are given.

4.4.1 Age-Related Differences on LiSN-S Measures

The older adults had poorer SRTs than younger adults in all LiSN-S conditions and they obtained less benefit than younger adults from the spatial separation of the competing talkers, voice differences, and the combination of these factors (i.e., they had poorer advantage scores than the younger participants). These results confirm previous observations that LiSN-S thresholds become significantly poorer for those above the age of 50 years (Cameron et al., 2011) compared with younger adults. Importantly, compared with younger adults, and also compared with people aged 50 to 60 years, we found that LiSN-S thresholds declined at ages above 65 years. Following the approach of Cameron et al. (2011), we divided the current older group into two subgroups, roughly splitting the sample by life decade with a group of 10 people aged 65 to 69 years and a group of 16 people aged 70 to 82 years. In Table 4.7, data of the present study are compared with the normative data of Cameron et al. (2011)§ for the Australian version of the LiSN-S test and to the normative data of Brown et al. (2010) for the North-American version of the LiSN-S test.

Cameron et al. (2011) reported that, unlike the SRTs, the advantage measures for people between 50 and 60 years of age were not worse than those of their younger groups. In contrast, in the present study we found that all the advantage measures were poorer for the older group than for the younger group. Additionally, as shown in Table 4.7, we found that the values for the spatial and total advantage of our older group were also lower than the same advantages in the 50 to 60 year olds in the study by Cameron et al., suggesting that not only the SRTs, but also two of the advantage measures are poorer at ages above 65 years. Surprisingly, the talker advantage of the present

§ No normative data is available for the LiSN-S SRTs in the SV90 or DVO conditions.
Table 4.7: LiSN-S outcome scores of the present study compared with the normative data for the Australian LiSN-S test reported by Cameron et al. (2011) and the normative data for the North-American LiSN-S test reported by Brown et al. (2010), organized by age group.

<table>
<thead>
<tr>
<th>Age Group (yrs)</th>
<th>Present Study</th>
<th>Cameron et al.</th>
<th>Brown et al.</th>
</tr>
</thead>
<tbody>
<tr>
<td>LiSN-S SRT for DV90</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18 – 29</td>
<td>−17.8</td>
<td>−17.5</td>
<td>−15.8</td>
</tr>
<tr>
<td>30 – 39</td>
<td>−18.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>40 – 49</td>
<td>−16.9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>50 – 60</td>
<td>−15.6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>65 – 69</td>
<td>−12.7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>70 – 82</td>
<td>−12.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LiSN-S SRT for SV0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18 – 29</td>
<td>−2.6</td>
<td>−2.9</td>
<td>−1.8</td>
</tr>
<tr>
<td>30 – 39</td>
<td>−2.8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>40 – 49</td>
<td>−2.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>50 – 60</td>
<td>−1.6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>65 – 69</td>
<td>−1.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>70 – 82</td>
<td>−0.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LiSN-S spatial advantage</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18 – 29</td>
<td>14.1</td>
<td>13.1</td>
<td>12.0</td>
</tr>
<tr>
<td>30 – 39</td>
<td></td>
<td>14.1</td>
<td></td>
</tr>
<tr>
<td>40 – 49</td>
<td></td>
<td>13.5</td>
<td></td>
</tr>
<tr>
<td>50 – 60</td>
<td></td>
<td>13.3</td>
<td></td>
</tr>
<tr>
<td>65 – 69</td>
<td>10.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>70 – 82</td>
<td>9.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LiSN-S talker advantage</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18 – 29</td>
<td>10.5</td>
<td>6.6</td>
<td>9.0</td>
</tr>
<tr>
<td>30 – 39</td>
<td></td>
<td>7.9</td>
<td></td>
</tr>
<tr>
<td>40 – 49</td>
<td></td>
<td>6.4</td>
<td></td>
</tr>
<tr>
<td>50 – 60</td>
<td></td>
<td>6.3</td>
<td></td>
</tr>
<tr>
<td>65 – 69</td>
<td>7.8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>70 – 82</td>
<td>8.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LiSN-S spatial advantage</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18 – 29</td>
<td>15.2</td>
<td>14.5</td>
<td>14.0</td>
</tr>
<tr>
<td>30 – 39</td>
<td></td>
<td>15.2</td>
<td></td>
</tr>
<tr>
<td>40 – 49</td>
<td></td>
<td>14.9</td>
<td></td>
</tr>
<tr>
<td>50 – 60</td>
<td></td>
<td>14.0</td>
<td></td>
</tr>
<tr>
<td>65 – 69</td>
<td>11.6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>70 – 82</td>
<td>11.8</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Mean scores for the data of Cameron et al. and Brown et al. are estimations, because no exact outcomes per age group were reported for these studies. The numbers represent values in dB and dB SNR for the advantage measures and SRTs, respectively. LiSN-S, listening in spatialized noise–sentences, SNR, signal-to-noise ratio, SRT, speech reception threshold.
study’s older group (8.2 dB) was higher than that of any group in the study of Cameron et al. This apparent discrepancy in the results may stem from the test differences between the North-American and the Australian version of the LiSN-S test in terms of dialect and/or talker variability. Using the North-American LISN-S test, Brown et al. (2010) also observed a higher talker advantage for younger adults than had been observed by Cameron et al. using the Australian LiSN-S test.

Age-related differences on LiSN-S measures could be a consequence of the differences in the susceptibility to semantic interference or of the differences in the use of spatial separation and voice difference cues, which will be discussed in the following.

### 4.4.1 Susceptibility to Semantic Interference

Age-related differences on LiSN-S measures could be a consequence of greater susceptibility to semantic interference in the older group, that is, poorer ability to ignore the meaningful content of the interfering talkers’ speech. In a previous study, older adults had poorer speech recognition scores than the younger adults in conditions with competing single-talker, two-talker, and multitalker maskers; however, there was no group difference by masker interaction, which was interpreted as the evidence that the amount of semantic interference did not affect performance (Tun & Wingfield, 1999). Note that individual streams of meaningful speech are usually not distinguishable from multitalker maskers such that these introduce less semantic interference than single- or two-talker maskers. Furthermore, performance of the younger and older adults was affected similarly by semantic interference, when energetic masking was accounted for by comparing accuracy for a noise-only masker with the accuracy for a noise-plus-speech masker (Agus, Akeroyd, Gatehouse & Warden, 2009). Younger and older listeners also achieved similar word recognition accuracy when they were instructed to focus on one of the two simultaneous speech streams (Meister et al., 2012), although older adults may need more time than the younger adults to segregate target speech from a speech masker (Ben-David, Tse & Schneider, 2012; Ezzatian, Li, Pichora-Fuller & Schneider, z. j.).

### 4.4.1.2 Use of Spatial Separation Cues and Voice Difference Cues by Age Group

The spatial advantage was bigger than the talker advantage in both the groups Figure 4.1 and removing spatial separation cues had a stronger detrimental effect on the SRTs than removing voice difference cues. However, we want to point out that the spatial and talker advantages cannot be compared in absolute terms in the LiSN-S test, because the conditions were not designed to yield equivalent advantages from each cue type. The conditions used in the test favor spatial advantage over talker advantage, presumably because the spatial sep-
aration of 90° between the talkers provides richer cues than the voice differences between the talkers of the same sex.

For both the age groups of the present study, adding one cue type (spatial separation or voice differences) resulted in significantly better LiSN-S SRTs compared with the condition in which no cues were available. The effect was larger in the younger group than in the older group for both the cue types (change from SV0 to SV90 or from SV0 to DV0). In contrast, compared with the condition in which both cue types were available, removing one of the cue types (change from DV90 to SV90 or from DV90 to DV0) had less effect on LiSN-S SRTs and changes in SRTs were approximately equivalent for both the age groups. Taken together, these results suggest that older adults are less able than younger listeners to make use of each one of the cue types in separation, whereas both the age groups benefit equally from the provision of a second cue. Thus, listeners in both the age groups seem to achieve similar benefit from the combination of both the cues, but older adults are less able to benefit from the availability of a single cue. This is in line with earlier research, in which younger adults outperformed older adults in recognizing noise-vocoded words, but both the age groups benefited equally from the addition of different prosodic cues (Wingfield, Lindfield & Goodglass, 2000). In other words, the ability to sum information from several cues seems to be preserved at higher ages.

Age-group differences were bigger for the use of spatial separation cues than for the use of voice differences. For example, the biggest age difference in SRTs (6.4 dB) was observed for the SV90 condition. This observation is in line with the findings from a recent study, which found clear effects of age on spatial release from masking while controlling for audibility (Gallun et al., 2013). According to Gallun et al. (2013), the reduced ability to make use of spatial separation cues reflects age-related changes in the temporal nuclei, probably along with cortical changes, that influence auditory scene analysis and thus stream segregation.

### 4.4.2 Contributions of Audiometric Thresholds to LiSN-S Outcomes

In the younger group, $\text{PTA}_{\text{HIGH}}$ accounted for 34.4% of the variance in the spatial advantage. Given that the pure-tone thresholds of our younger group were $\leq 20$ dB HL at all the measured frequencies up to and including 10,000 Hz, this result is not obviously the consequence of significant reductions in audibility. Dobreva, O’Neill en Paige (2011) found that 9000 and 10,000 Hz play a role in sound localization, which would explain why $\text{PTA}_{\text{HIGH}}$ may be related to the spatial advantage in the younger listeners, but not to any other LiSN-S measure.

The older group in the present study had statistically significantly
poorer pure-tone hearing acuity than the younger participants, tested for frequencies up to and including 10,000 Hz. Nevertheless, for our older adults who had clinically normal audiometric thresholds up to and including 3000 Hz, 4FAHL was associated with LiSN-S SRTs to some degree (DV0, SV90; Table 4.3), but did not predict the outcomes in the regression analyses. Instead, hearing acuity at higher frequencies above and including 6000 Hz was a strong predictor of LiSN-S outcomes. Importantly, the finding that typical audiometric data do not predict LiSN-S performance for older adults with relatively good audiograms contrasts to the earlier findings of Glyde, Cameron et al. (2013) who found that pure-tone hearing acuity rather than age determined LiSN-S outcomes for individuals up to 89 years of age in their Australian sample, in which hearing loss was confounded with age.

There are several potential explanations for the contributions of very high-frequency pure-tone thresholds to LiSN-S outcomes. As for the younger adults, these contributions may be attributable to the role of 9000 and 10,000 Hz in sound localization (Dobreva et al., 2011). However, in the older group PTA\textsubscript{HIGH} also influenced outcomes in conditions in which no spatial cues were available (DV0 and talker advantage). Another possibility is that the strong prediction of LiSN-S SRTs and advantage measures by very high-frequency audiometric thresholds in our older group could be related to audibility issues, if the ear-specific gain amplifier based on pure-tone audiometric thresholds up to and including 8000 Hz did not compensate sufficiently for elevated thresholds at the higher frequencies. Indeed, Glyde, Cameron et al. (2013), state that the gain amplifier built into the LiSN-S test provides less high-frequency gain than recommended by other amplification methods, which could lead to reduced access to some interaural intensity difference cues. However, it is unlikely that the audibility issues alone account for the relationship between PTA\textsubscript{HIGH} and LiSN-S outcomes because there is relatively little speech information present at such high frequencies. A final possible explanation is that increased high-frequency thresholds could be an indicator of cochlear damage that has altered aspects of auditory processing that play a role in LiSN-S outcomes even though the damage has not yet resulted in threshold shifts at lower frequencies. However, the monaural suprathreshold auditory temporal processing abilities tested in the present study did not predict LiSN-S outcomes in the older group.

The influence of high-frequency hearing acuity on LiSN-S SRTs might be explained by recent findings in neuro-imaging studies. Peelle, Troiani, Grossman en Wingfield (2011) observed that people with decreased hearing acuity had reduced neural responses to speech stimuli. The same listeners also had reduced gray matter volume in primary auditory regions. These findings were confirmed in a later study by Eckert, Cute, Vaden Jr, Kuchinsky en Dubno (2012), who...
were also able to attribute the reductions of gray matter in the auditory cortex more specifically to high-frequency hearing loss (investigated for frequencies ≤ 8000 Hz). In both the studies, the effects of hearing loss on gray matter volume were distinct from general age-related reductions in gray matter. The findings from the studies by Peelle et al. and Eckert et al. indicate that there is a solid link between sensory acuity and cortical integrity. Thus, the behavioral differences in speech understanding in noise between the younger and older adults with good audiograms and the dependence of the LiSN-S outcomes on high-frequency hearing acuity in the older group might actually stem from cortical changes rather than being a direct consequence of peripheral aspects of auditory processing.

4.4.3 Contributions of Auditory Temporal Processing to LiSN-S Outcomes

Overall, the results of the present study for both the age groups agree with those of the previous studies for the TiNG (Larsby & Arlinger, 1998; van Esch & Dreschler, 2011), F0 DL (Vongpaisal & Pichora-Fuller, 2007), Gap SU (Pichora-Fuller et al., 2006), and TFS6 (Hopkins & Moore, 2011) tests.

The effect of interaural time differences on speech-in-speech listening with spatial separation cues has been examined before using speech material from the LiSN-S test (Glyde, Buchholz et al., 2013). However, to our knowledge, the present study is the first to examine relationships between LiSN-S outcomes and specific measures of auditory temporal processing. We expected that LiSN-S outcomes would be influenced by the ability of the listeners to use both temporal fine-structure and temporal envelope cues, because of the role of these cues in voice-stream segregation and benefit from masker fluctuations (see Section 4.1.2.1). As anticipated, the group comparisons showed that our older group performed more poorly on all the tests of temporal processing, except temporal resolution at 750 Hz (TiNG750Hz). However, surprisingly, auditory temporal processing abilities did not predict the outcomes on the LiSN-S test in the older group and played only a minimal role in predicting LiSN-S outcomes in the younger group. When accounting for linguistic closure abilities in low-context sentences (TRT LOW) in the younger group, listeners with better temporal resolution at low frequencies (TiNG750Hz) had better DV90 thresholds. Temporal resolution has often been related to the ability to make use of masker fluctuations (George et al., 2006).

6 Regarding the TFS test it should be noted that to be able to include the test results of those who were and those who were not able to complete the adaptive test procedure, Hopkins en Moore (2011) converted the outcome scores to d’ values. The participants who could not complete the adaptive test received 40 items at the easiest test level and a percentage correct score was calculated. Instead, we excluded the data of such participants, because they usually performed at chance, suggesting that they were unable to recognize frequency shifts at the test frequency.
The extent of masker fluctuations was comparable in all LiSN-S conditions. The finding that TiNG\textsubscript{750Hz} contributed to DV90 thresholds, but not to the other outcomes may indicate that only in the conditions where the target and the masker can be separated based on rich cues (in this case a combination of voice and spatial cues), does temporal resolution play an additional role and influence performance.

The general lack of significant influence of temporal processing ability on LiSN-S outcomes could be a consequence of several factors. In general, there have been inconsistent findings in the studies examining the relationship between the measures of temporal processing and speech understanding in noise (e.g., Strouse, Ashmead, Ohde & Grantham, 1998; Snell, Mapes, Hickman & Frisina, 2002; Vestergaard, Fyson & Patterson, 2011). These inconsistencies may be attributable to the properties of the target speech; for example, temporal processing may play a greater role as the speech rate increases or when accurate word identification relies on phonemic contrasts involving temporal distinctions (e.g., Gordon-Salant et al., 2006; Pichora-Fuller et al., 2006). Another possibility is that the effect of temporal processing abilities was dominated by the influence of semantic interference and/or by predominant influences of the binaural spatial processing abilities.

4.4.4 Contributions of Cognitive and Linguistic Abilities to LiSN-S Outcomes

In the younger group, the linguistic measures (TRT\textsubscript{LOW} and vocabulary) dominated contributions to LiSN-S outcomes that we examined. The TRT has previously been found to be a good predictor of SRTs when both speech and modulated noise maskers are used to test middle-aged or older participants with normal hearing (Besser et al., 2013; Chapter 3, but many of the studies reviewed by Besser et al. (2013) failed to find associations between TRT and speech understanding in younger adults with normal hearing. In contrast, in the present study, TRT scores did not predict LiSN-S outcomes in the older group, whereas TRT\textsubscript{LOW} was the best overall predictor of LiSN-S SRTs in the younger group. The involvement of explicit higher-order linguistic processing is assumed to be more pronounced in challenging listening situations, in which implicit lower-order processing is insufficient to enable comprehension (see Rönnberg et al., 2013, for a model). None of the studies reviewed by Besser et al. (2013) examined the associations between the TRT and the understanding of speech in speech or in spatial listening situations in younger adults with normal hearing. The findings of the present study suggest that the conditions of the LiSN-S test were challenging enough to bring forth influences of higher-order linguistic abilities on speech understanding in younger adults. Similarly, the finding that linguistic abilities did
not play a role in the performance of our older group may be because of the relatively challenging conditions used in the LiSN-S test conditions. Younger adults were probably performing extremely well and relatively homogeneously at many levels of the processing hierarchy (hearing acuity, monaural auditory temporal processing, and general cognitive abilities) such that individual differences in linguistic abilities played a greater role in predicting LiSN-S outcomes. In the older group on the other hand, the effect of linguistic abilities was likely overshadowed by the abilities at lower levels of the processing hierarchy. Overall, this sort of explanation implies a general hierarchy of processing, in which linguistic influences are subordinate to other cognitive mechanisms and acoustic processing.

Of note, semantic contextual cues in the sentence materials used for the TRT\textsubscript{LOW} test in the present study were reduced to such a degree that the linguistic closure on the TRT\textsubscript{LOW} test likely depended less on semantic integration and more on syntactic and lexical knowledge. The minimal availability of the semantic context may also explain why the TRT results in the present study were poorer than the previously reported TRT results. Nevertheless, the pattern of older adults needing a somewhat higher percentage of unmasked text than younger adults agrees with previous results (Besser et al., 2013).

For the older group, MoCA scores predicted most LiSN-S SRTs. The influence of general cognitive ability on the performance of our older listeners fits well with a number of previous research studies that underscore the importance of different cognitive functions in speech understanding by older adults (e.g., Wingfield & Tun, 2007; Schneider et al., 2010) and with electrophysiological evidence that an auditory-cognitive neural system is involved in speech understanding in noise, independent of a listener’s hearing acuity (Anderson et al., 2013). The fact that Glyde, Cameron et al. (2013) did not observe an association between cognition and LiSN-S outcomes might be a consequence of their choice of cognitive test (COGNITSTAT Mueller et al., 2001), the predominant role of hearing thresholds in their sample, and/or the smaller sample of very old adults with good audiograms in their study.

Interestingly, the MoCA score was the strongest predictor of the LiSN-S SRT in the SV0 condition, which was also the only condition predicted by age. Together, age and MoCA scores accounted for more than half of the variance observed in the SV0 condition. Acoustic cues for stream segregation in the SV0 condition are extremely limited because neither spatial separation nor voice difference cues could be exploited. Presumably, acoustic cues in the other conditions are rich enough to provide for stream segregation independent of attention, consistent with the hierarchical model of stream segregation proposed by Cusack, Deeks, Aikman en Carlyon (2004). It has also been found that attention only modulates the buildup of acoustic
streams in the absence of robust segregation cues (Snyder et al., 2012). Thus, the additional influence of age on performance in this condition may reflect attentional mechanisms not assessed by the tests used in the present study.

4.4.5 Summary and General Conclusions

The listeners aged 65 years and older who participated in the present study had poorer LiSN-S outcomes than the younger adults in the present study and also than the middle-aged (50–60 years) listeners from a previous study (Cameron et al., 2011), indicating that the ability to make use of spatial separation and voice difference cues during speech-in-speech listening becomes poorer by increasing age. However, the data also suggest that while older adults benefit less than younger adults from the provision of a single cue type, both the age groups obtain similar SRT benefit from combined spatial and separation cues.

The results of the present study show that higher-order processing plays a role in speech-in-speech listening, regardless of the availability of voice differences and spatial separation cues. The contributions of specific aspects of higher-order processing depend on age. In the younger group, linguistic abilities were the most decisive factor for LiSN-S outcomes. In the older group, general cognitive ability predicted most LiSN-S thresholds. Different from our hypotheses, the influence of auditory temporal processing was very limited for listeners of both the age groups. Of note, general cognitive ability did not differ for the two age groups, whereas there were age-related differences for most measures of auditory temporal processing. This pattern of results could indicate either that the older listeners relied on their strongest suprathreshold processing domain, i.e., cognitive ability, or that the monaural temporal processing abilities assessed were not decisive for performance on the LiSN-S test. Furthermore, pure-tone thresholds at high frequencies (mean of 6000, 8000, 9000, 10,000 Hz) dominated contributions to LiSN-S outcomes in the older group. The strong influence of PTA<sub>HIGH</sub> might reflect the reductions in the peripheral auditory status and/or the functioning of the auditory cortex caused by reduced high-frequency acuity before the threshold elevations at lower frequencies are manifested. In the absence of voice differences and spatial separation cues, age predicted LiSN-S thresholds in the older group, possibly related to attentional mechanisms that influence streaming in the absence of strong segregation cues. Overall, our findings are in support of a highly integrated auditory-cognitive processing system where changes in one domain are associated with changes in the other domain.
The authors thank all the people who participated in the study and Dario Coletta and Jacob Maracle for their assistance with the data collection. The study was funded by a grant from the Natural Sciences and Engineering Research Council of Canada (to M.K.P.), a research grant from Foundation Het Heinsius Houbolt Fonds, and a travel grant from the EMGO+ Institute of Health and Care Research (to J.B.).

Preliminary results of this study have been presented in poster presentations at the 21st International Congress on Acoustics (Montreal, June 2013), at the 2nd International Conference on Cognitive Hearing Science for Communication (Linköping, June 2013), and at the 5th International and Interdisciplinary Research Conference on Aging and Speech Communication (Indiana, October 2013).