2pSCb27. Advantage of talker differences and spatial separation for speech-on-speech listening in younger and older adults with good audiograms

Jana Besser*, M. K. Pichora-Fuller and Joost M. Festen

*Corresponding author's address: ENT/audiology, VU University Medical Center, De Boelelaan 1118, Amsterdam, 1081HZ, NH, Netherlands, besser.jana@gmail.com

Older adults have more difficulty than younger adults understanding speech when there is competing speech, even if they have good audiograms. Age-related differences in listening may be due to declines in auditory temporal processing and/or cognition. We administered the LiSN-S (Cameron et al., 2011) to measure speech reception thresholds (SRTs) in younger and older adults with good audiograms. There were four test conditions, in which the target and competing speech were presented with the same or different voices at the same or different locations. Compared to younger listeners, older listeners obtained worse SRTs in all test conditions and they realized less advantage from talker differences and spatial separation between the target and competing speech. For both groups, the results obtained in the four test conditions were strongly associated with each other. We also assessed cognitive abilities and auditory temporal processing in the older adults. LiSN-S results in this group were strongly associated with measures of cognition, as well as pure-tone averages (PTA) for 9 and 10 kHz, but not PTAs for frequencies in the standard audiometric range or measures of temporal processing (tapping the use of fine structure and gap cues).
INTRODUCTION

The Listening in Spatialized Noise-Sentences (LiSN-S) test (Cameron & Dillon, 2007) is a test of speech recognition in which the availability of spatial and/or voice cues is manipulated. In four test conditions, listeners are asked to repeat sentences spoken by a female target talker perceived to be located in front of the listener. An interfering female talker is speaking with the same or a different voice from the same or different perceived spatial locations (+90º and -90 º) relative to the target talker. The stimuli are presented binaurally under headphones and head-related transfer functions are used to simulate the spatial locations of the talkers. For each test condition, a speech reception threshold (SRT) is calculated. Furthermore, advantage measures index the benefit in SRT due to the perceived spatial separation of the talkers (spatial advantage), voice differences (talker advantage), or combined perceived spatial separation and voice differences (total advantage). The test was developed as a clinical tool for detecting deficits in central auditory processing in children and adults. Normative data for individuals with normal audiometric thresholds is available up to the age of 60 years for the Australian version of the test (Cameron et al., 2011) and up to the age of 30 years for the North-America version of the test (Brown et al., 2010). The advantage measures are not affected by age, but the SRTs become poorer in the sixth decade (Cameron et al., 2011). It is unclear how the LiSN-S measures change at older ages and whether they are associated with age as such or rather that the apparent age-related declines are a consequence of increased hearing impairment. In a recent study, Glyde et al. (2013) did not find an association between age and LiSN-S outcomes when controlling for pure-tone hearing acuity in a mixed group of children and adults up to 89 years of age with normal or impaired hearing. Average hearing loss over frequencies between .5 and 4 kHz accounted for more than half of the variance in three out of five reported LiSN-S measures. However, only approximately 20 out of the 80 participants had normal average hearing acuity in the range .5 – 4 kHz and only about 6 of these were above the age of 40 years. It is important to more extensively investigate if the ability to make use of spatial and voice cues during understanding speech when there is a competing talker varies with age for older adults who have normal hearing acuity but sub-clinical declines in supra-threshold auditory processing. For example, even for older listeners with normal or near-normal audiograms, the ability to make use of the temporal properties of the speech signal is likely to play a role in word recognition when there is an interfering voice that has the same or a similar spectrum as the target voice. Furthermore, general cognitive ability might be a determining factor for LiSN-S performance at older ages, as well as linguistic abilities that are relevant for speech comprehension independent of age (Besser et al., 2012). In the current study, we examined the relationships of cognitive, linguistic, and auditory temporal processing measures with LiSN-S performance in a group of older adults with clinically normal audiometric thresholds in the speech range, and we compared their LiSN-S performance to a group of young adults with clinically normal hearing, i.e., LiSN-S top performers (Cameron et al., 2011).

METHOD

Participants

The experiment included two groups of participants with pure-tone air-conduction audiometric thresholds of 25 dB HL or better at frequencies from 250 Hz to 3 kHz in both ears. Hearing acuity was tested up to 20 kHz, but thresholds above 3 kHz did not influence eligibility for the study. Each group consisted of 26 adults, either younger (18-27 years, M = 21.2 years, SD = 2.4 years) or older (66-82, M = 72.0 years, SD = 4.2 years). All participants were native English speakers or had learned English before the age of 5 years. Participants received monetary compensation of $10 CAD per hour and provided informed consent prior to testing.

Measures

Both participant groups performed the LiSN-S test described above in a sound-attenuating booth using a commercial version of the North-American version of the test (Phonak, Version: V2.003). The order in which the test conditions were administered followed the recommendations of Cameron and Dillon (2007); i.e., 1) different voices at different locations, 2) different voices at the same location, 3) same voice at different locations, 4) same voice at the same location.

The older group also completed the Montreal Cognitive Assessment (MoCA), a screening tool for the detection of mild cognitive impairment (Nasreddine et al., 2005), and the total score was used as a continuous variable indicating general cognitive ability. We used the text reception threshold (TRT) test (Zekveld et al., 2007) to assess modality-independent processing abilities required to make sense of incomplete linguistic
information. The TRT is a visual analogue of the speech reception threshold that has been found to be strongly associated with ability to understand masked speech (Besser et al., 2012). The participants’ task during this test was to read as accurately as possible high- and low-context sentences from the Speech Perception in Noise (SPIN) test (Kalikow et al., 1977) corpus that were presented partially masked by a pattern of black vertical bars.

Auditory temporal processing was assessed by means of four tests, two using non-speech stimuli and two using speech stimuli, with one measure for each of the stimulus types focused primarily on the use of fine structure cues and the other focused primarily on the use of envelop cues. The two non-speech tests used targets of 750 Hz and at 3 kHz. Auditory sensitivity to temporal fine-structure (TFS) information was measured with the TFS2 test (Hopkins & Moore, 2011) in a two-interval two-alternative forced choice procedure to evaluate the listener’s ability to detect changes in the waveform fine structure of a harmonic tone compared to a complex inharmonic signal resulting from concurrent frequency shifts of the harmonic tone’s partials. Another task measured difference in dB SNR between the threshold for detecting a tone in a noise gap (TiNG) and the threshold for detecting it in continuous noise, i.e., the TiNG benefit due to ability to take advantage of temporal modulations in the noise masker. Using speech stimuli, we used a two-interval, two-alternative forced choice procedure to determine the smallest difference in voice F0 (voice F0-DL) that a participant was able to detect between synthesized tokens of the vowel /a/. The F0-DL threshold is assumed to represent abilities in periodicity coding of fine structure cues arising from the F0 and harmonic structure of speech that likely play a role in speaker segregation (Vongpaisal & Pichora-Fuller, 2007) and release from informational masking of a speech target by interfering speech (Ezzatian et al., 2012). Finally, a test for auditory temporal resolution (gapSU) measured detection thresholds for gaps in speech markers, i.e., a gap introduced in the speech envelope between the initial consonant and the final vowel of the syllable [su] to render [s_u] which is perceived as /spu/ when the gap is sufficiently large (Pichora-Fuller et al., 2006).

**Statistical Analyses**

We used general linear models to assess the effects of age group (younger or older) and test condition on LiSN-S SRTs and advantage measures. Correlation coefficients were calculated for all 7 LiSN-S outcome measures with hearing acuity in different frequency regions, age, MoCA, TRT scores, and the four tests of auditory temporal processing. We calculated Pearson correlation coefficients, except for not normally distributed variables (F0-DL and gapSU), for which Spearman correlation coefficients were calculated. For three of the outcome measures, a data point strongly deviating from the remaining data was observed. Excluding these outliers from the analyses did not change significance levels of the observed associations. We therefore decided to retain all available data in the analyses. Some participants were not able to perform the test for TFS sensitivity; therefore, analyses of these data included fewer data points (17 at 750 Hz, 18 at 3 kHz).

**RESULTS**

We calculated participants’ pure-tone hearing acuity in the worse ear in different frequency regions, i.e., the average threshold of .5 and 1 kHz (PTA\textsubscript{low}), the average threshold of 2 and 4 kHz (PTA\textsubscript{mid}), the 4 frequency average hearing loss (4FAHL) covering all frequencies of PTA\textsubscript{low} and PTA\textsubscript{mid}, and the average threshold of 9 and 10 kHz (PTA\textsubscript{high}). Mean values and standard deviations are reported in Table 1. Group audiograms are also displayed in Figure 1. The older adults had higher thresholds for all four measures. All group differences were found to be statistically significant ($p < .001$), tested with independent-samples t-tests.

<table>
<thead>
<tr>
<th>Age group</th>
<th>PTA\textsubscript{low}</th>
<th>PTA\textsubscript{mid}</th>
<th>4FAHL</th>
<th>PTA\textsubscript{high}</th>
</tr>
</thead>
<tbody>
<tr>
<td>older group</td>
<td>8.3 (SD = 4.6)</td>
<td>20.5 (SD = 7.7)</td>
<td>13.8 (SD = 4.7)</td>
<td>65.2 (SD = 13.9)</td>
</tr>
<tr>
<td>young group</td>
<td>0.1 (SD = 3.8)</td>
<td>0.4 (SD = 3.8)</td>
<td>-.1 (SD = 3.3)</td>
<td>7.3 (SD = 7.9)</td>
</tr>
</tbody>
</table>
As seen in Figure 2, the older group had poorer LiSN-S performance in all four conditions. LiSN-S scores were best for different voices at different spatial locations, followed by scores for the same voice at different locations, and poorest for the same voice at the same location. This description was confirmed by 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 (different voices at different locations, same voice at different locations, different voices at the same location, same voice at the same location). There were significant main effects of group, $F(1,50) = 101.092, p < .001$, and LiSN-S condition, $F(3,150) = 944.215, p < .001$, as well as a significant two-way interaction between group and LiSN-S condition, $F(3,150) = 18.273, p < .001$. Pairwise comparisons using t-tests confirmed that younger adults out-performed older adults in all conditions and all conditions were significantly different from each other ($p$s < .05).

As shown in Figure 3, for the LiSN-S advantage measures, older participants had smaller talker advantage, spatial advantage, and total advantage compared to the younger participants. This description was confirmed by another 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, $F(1,50) = 28.679, p < .001$, and LiSN-S advantage condition, $F(2,100) = 150.193, p < .001$, as well as a significant two-way interaction between group and LiSN-S advantage condition, $F(2,100) = 15.352, p < .001$. Pairwise comparisons using t-tests confirmed that younger adults out-performed older adults in all conditions and all conditions were significantly different from each other ($p$s < .05).
FIGURE 3. Mean SNR advantage is shown for each age group for the three LiSN-S advantage measures. Error bars indicate the Std. Error of the Mean.

For the older group, we calculated correlations of the LiSN-S outcomes with measures of hearing acuity and with age, cognitive ability, linguistic processing skills, and auditory temporal processing (see Table 2). PTA_{mid} and 4FAHL were associated with the LiSN-S SRT with the same voice at different perceived locations, with different voices at the same perceived location, and with the spatial advantage. PTA_{high}, was associated with all LiSN-S outcome measures, except the condition in which talkers with the same voice were perceived at the same location. We also performed correlation analyses for the LiSN-S measures and hearing acuity in the younger group. The only significant associations were found for PTA_{high} with the LiSN-S SRT for the same voice at spatially separated locations ($r = .45$, $p < .05$) and the spatial advantage ($r = -.39$, $p < .05$) and for PTA_{low} with the LiSN-S threshold for different voices at spatially separated locations ($r = -.39$, $p < .05$); however, note that the direction of the latter correlation means that poorer hearing acuity at low frequencies was associated with better SRTs.

TABLE 2. Two-tailed Pearson correlations (*$p < .05$, **$p < .01$) between LiSN-S outcome scores and measures that were significantly correlated with at least one of the LiSN-S outcomes, for the older listeners (N = 25). Hearing acuity measures represent thresholds in the participant’s worse ear.

<table>
<thead>
<tr>
<th>LiSN-S outcome</th>
<th>PTA_{mid}</th>
<th>4FAHL</th>
<th>PTA_{high}</th>
<th>Age</th>
<th>MoCA</th>
<th>TRT low context</th>
<th>TRT high context</th>
</tr>
</thead>
<tbody>
<tr>
<td>different voices ±90°</td>
<td>.38</td>
<td>.36</td>
<td>.71**</td>
<td>.07</td>
<td>-.48*</td>
<td>.48*</td>
<td>.52**</td>
</tr>
<tr>
<td>same voice ±90°</td>
<td>.51**</td>
<td>.47*</td>
<td>.71**</td>
<td>.32</td>
<td>-.51**</td>
<td>.38</td>
<td>.37</td>
</tr>
<tr>
<td>different voices ±0°</td>
<td>.41*</td>
<td>.42*</td>
<td>.68**</td>
<td>.12</td>
<td>-.32</td>
<td>.42*</td>
<td>.32</td>
</tr>
<tr>
<td>same voice ±0°</td>
<td>.24</td>
<td>.14</td>
<td>-.06</td>
<td>.46*</td>
<td>-.36</td>
<td>.14</td>
<td>.09</td>
</tr>
<tr>
<td>talker advantage</td>
<td>-.28</td>
<td>-.34</td>
<td>-.69***</td>
<td>.11</td>
<td>.13</td>
<td>-.35</td>
<td>-.27</td>
</tr>
<tr>
<td>spatial advantage</td>
<td>-.43*</td>
<td>-.44*</td>
<td>-.77***</td>
<td>-.13</td>
<td>.37</td>
<td>-.36</td>
<td>-.36</td>
</tr>
<tr>
<td>total advantage</td>
<td>-.27</td>
<td>-.29</td>
<td>-.73***</td>
<td>.14</td>
<td>.28</td>
<td>-.42*</td>
<td>-.49*</td>
</tr>
</tbody>
</table>

Age was only associated with LiSN-S performance when both talkers were speaking in the same voice at the same location. Better MoCA scores were associated with better LiSN-S SRTs in the two conditions where the talkers were speaking from different perceived locations, but MoCA scores were not associated with any of the advantage measures. Higher (poorer) TRTs for high- and low-context sentences were associated with higher (poorer) LiSN-S thresholds in the condition with different voices at different perceived locations and with the combined advantage of spatial and voice cues. Low-context TRTs were also associated with the other condition where target and interfering voice differed. LiSN-S outcome scores were not significantly correlated with any other of the included tests. Correlation analyses also revealed that MoCA outcomes were not associated with age, and age was not associated with any of the measures of hearing acuity either, as expected given that only those with good audiograms were included.

As stated earlier, there were some people who were not able to perform the test for TFS sensitivity, i.e., were not able to detect fine-structure differences at the easiest test level. Accordingly, there were fewer data points for the TFS measures. We therefore performed further analyses coding the TFS outcome as a dichotomous variable,
indicating whether or not the participant was able to perform the test. We performed t-tests to check for differences in LiSN-S outcomes and hearing acuity between the people who were and those who were not able to do the test. There were no statistically significant differences between these groups.

DISCUSSION

Older adults performed significantly worse in all LiSN-S conditions and had lower advantage scores than the younger adult participants, confirming the decrease in LiSN-S performance above the age of 50 years observed previously (Cameron et al., 2011). Older adults also had statistically significantly poorer pure-tone hearing acuity than the young participants, but group audiograms in the speech range did not differ from a clinical perspective and the audiograms of the older adults would be considered to be normal for their age (ISO 7029-2000). Importantly, although older adults performed worse than younger adults in all LiSN-S conditions, associations between hearing acuity at up to 4 kHz and LiSN-S SRTs with differing voice and spatial cues were not evident in the current study sample of older adults with good audiograms. The lack of further significant associations could be an effect of the relatively small spread in pure-tone thresholds in this frequency range in the current participant group. Support for this interpretation comes from Glyde et al. (2013), who found LiSN-S outcomes to be strongly influenced by hearing acuity at frequencies between .5 and 4 kHz. On the other hand, the ability to use spatial cues in the absence of voice cues was related to PTA_{mid} for older adults and to PTA_{high} in the young adults, who had mostly normal thresholds even at 9 and 10 kHz. In the older group, hearing acuity at high frequencies beyond the usual audiogram range (PTA_{high}) affected speech understanding in all tested listening situations, except when the target and interfering speakers were speaking at the same location and with the same voice. The connection between high-frequency hearing acuity and ability to make use of spatial cues may be attributable to the role of 9-10 kHz in localization (Dobreva et al., 2011) and the connection to ability to use voice cues may involve the use of information provided by the higher harmonics (Russo et al., 2012). A remarkable finding is that none of the measures of auditory temporal processing was associated with any of the LiSN-S outcomes. This suggests that the ability of older adults with normal hearing acuity in the speech range to make use of voice cues and spatial cues in speech recognition may be related to different auditory functions or non-auditory factors. The results of the correlational analyses confirm that linguistic processing ability (TRT) and general cognitive ability (MoCA) play a role in listening situations requiring spatial or voice segregation, although not in other LiSN-S conditions. This is in contrast with Glyde et al. (2013) who did not find associations between a test of cognitive ability and any of the LiSN-S outcomes. Also in the current study, TRT and MoCA are not associated with all LiSN-S measures. However, this could partially be a consequence of the relatively small size of our study sample. It is interesting that the LiSN-S threshold with the same voice at the same location is the only LiSN-S measure that is associated with age, but it is clearly not connected to any of the hearing acuity measures, the auditory processing measures or the TRT measure. In summary, older adults perform worse than young adults on the LiSN-S test. Although the older group had statistically poorer audiometric thresholds than the young group, their audiometric thresholds were normal in the speech range. For the older group, performance on the LiSN-S test was largely unrelated to audiometric thresholds up to 4 kHz, and measures of auditory temporal processing were not associated with any of the LiSN-S measures. Age determined performance only in situations where no spatial separation and no voice difference cues were available. When either of these cues were available, modality-independent linguistic processing, cognitive ability, and high-frequency (9-10 kHz) hearing acuity came into play.

ACKNOWLEDGMENTS

We would like to thank all people who participated in the study and Dario Coletta for his efforts in conducting part of the data collection. The study was funded by a grant from the Natural Sciences and Engineering Research Council of Canada to M.K. Pichora-Fuller and a travel grant from the EMGO+ Institute of Health and Care Research to J. Besser.

REFERENCES


