A perceptual study of speaker age

Susanne Schötz
Dept. of Linguistics and Phonetics, Lund University
Susanne.Schotz@ling.lu.se

Abstract
The present paper examines listeners’ ability to judge speaker age from stimuli consisting of phonated isolated words, and also suggests subdivision (categorisation) of voices into typical and atypical groups, depending on the listeners’ age estimating abilities.

1 Introduction
1.1 Purpose and hypotheses
The present paper, based upon my B.A. essay in phonetics (Schötz 2001) examines listeners’ ability to judge speaker age from stimuli consisting of isolated words pronounced by elderly and young male and female speakers. Two hypotheses were tested:
   1. Listeners are able to judge speaker age within ±10 years of chronological age.
   2. Some voices are more easily age-estimated than others, perhaps due to individual levels of instability and variability in the age-related cues of the voice, hence suggesting a subdivision into typical and atypical voices, depending on the degree of correspondence between perceived and chronological age.

This perceptual study serves as a preparatory investigation to an acoustical study of the same speech material, which is currently (April-May 2001) in progress.

1.2 The aging of the voice
Researchers have tried to relate certain acoustic and perceptual parameters of voice to the aging process. Although data are still sparse, literature often considers the following to be the main correlates of the aging voice: (a) fundamental frequency – pitch, (b) vocal intensity – loudness, (c) jitter and shimmer – roughness, (d) formant frequencies and spectral slope – voice quality, and (e) duration and pausation – speech rhythm and timing.

1.3 Previous studies
Most researchers agree that age-related changes in laryngeal and supralaryngeal structures affect the perceived acoustic properties of speech, that listeners’ age judgements depend on both laryngeal and supralaryngeal cues (Jacques & Rastatter 1990), and that untrained listeners are able to judge speaker age within ±5 years of chronological age at levels considerably better than chance (Hollien 1987). In most previous studies the listeners were asked only to place speakers in two (Jacques & Rastatter 1990, Ptacek & Sander 1966), three (Linville 1987) or four (Traunmüller & van Bezooijen 1994) age categories, rather than estimate age more precisely. Also, in most cases, the investigations focused on age perception of phonated (sometimes also filtered and whispered) vowels and/or whole sentences, and not isolated words. The phonated stimuli usually produced the best results, indicating that fundamental frequency is the most dominant cue for age estimation. Results from several studies have led researchers to suggest that there is a connection between speaker age

2 Method

2.1 Speech material

The examined speech material consisted of the three Swedish words ‘tack’ ‘thanks’, ‘rasa’ ‘fall down’ and ‘tusendollarsedlar’ ‘thousand-dollar bills’, pronounced in the same dialect (småländska) by eight non-pathological speakers, giving a total of 24 words. All speakers (four female, age 73, 61, 30 and 28 years, and four male, age 71, 66, 27 and 21 years) were taken from the research project SweDia-2000’s database. The speech material underwent a preliminary auditive analysis, which subdivided the speakers into four typical and four atypical voices in accordance with the second hypothesis of the study.

2.2 Procedure

In a perception test presented as a web page, 38 listeners (19 male, 19 female, age 14-60 years) were asked to age-estimate the 24 words, presented in a random order, by selecting one of 18 age alternatives, ranging from 10 to 95 years, from a menu. The results for each of the 24 words, for each of the eight speakers, and for speaker and listener groups, were analysed using calculations of the mean values, standard deviations (SD) and ranges, as well as the number and percentages of correct age estimates.

![A typical young woman, 28 years](image)

**Figure 1.** 38 listeners’ age estimates of a typical young female speaker.

<table>
<thead>
<tr>
<th>Young woman (28 years) (38 listeners)</th>
<th>'tack'</th>
<th></th>
<th>'rasa'</th>
<th></th>
<th>'tusendollarsedlar'</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>number</td>
<td>percent</td>
<td>number</td>
<td>percent</td>
<td>number</td>
<td>percent</td>
<td></td>
</tr>
<tr>
<td>Correct</td>
<td>9</td>
<td>24%</td>
<td>11</td>
<td>29%</td>
<td>6</td>
<td>16%</td>
</tr>
<tr>
<td>Correct ±5 years</td>
<td>23</td>
<td>61%</td>
<td>25</td>
<td>66%</td>
<td>27</td>
<td>71%</td>
</tr>
<tr>
<td>Correct ±10 years</td>
<td>28</td>
<td>74%</td>
<td>35</td>
<td>92%</td>
<td>30</td>
<td>79%</td>
</tr>
<tr>
<td>Mean</td>
<td>33.68</td>
<td>28.55</td>
<td>27.24</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>9.42</td>
<td>7.96</td>
<td>9.13</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Max</td>
<td>55</td>
<td>50</td>
<td>60</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Min</td>
<td>15</td>
<td>15</td>
<td>10</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Range</td>
<td>40</td>
<td>35</td>
<td>50</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Table 1.** 38 listeners’ age estimates of a typical young female speaker.
3 Results and analysis

3.1 Typical and atypical voices

One of the main purposes of this study was to find out whether typical and atypical voices exist. The results indicate that such a subdivision actually may exist, and that the preliminary subcategorization of the speakers turned out to be accurate enough to support the hypothesis. The four typical speakers did get a greater amount of correct age estimates as well as lower SD and range values than the four atypical speakers. Figures 1-2 and Tables 1-2 illustrate age estimates for two of the speakers of the study: a typical young woman and an atypical elderly man.

![An atypical elderly man (71 years)](image)

Figure 2. 38 listeners’ age estimates of an atypical elderly male speaker.

<table>
<thead>
<tr>
<th>Elderly man (71 years)</th>
<th>'tack'</th>
<th>'rasa'</th>
<th>'tusendollarsedlar'</th>
</tr>
</thead>
<tbody>
<tr>
<td>(38 listeners)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Correct</td>
<td>4</td>
<td>11%</td>
<td>1</td>
</tr>
<tr>
<td>Correct ±5 years</td>
<td>10</td>
<td>26%</td>
<td>3</td>
</tr>
<tr>
<td>Correct ±10 years</td>
<td>18</td>
<td>47%</td>
<td>7</td>
</tr>
<tr>
<td>Mean</td>
<td>56.97</td>
<td>45.79</td>
<td>55.66</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>17.61</td>
<td>13.18</td>
<td>9.53</td>
</tr>
<tr>
<td>Max</td>
<td>85</td>
<td>75</td>
<td>80</td>
</tr>
<tr>
<td>Min</td>
<td>15</td>
<td>25</td>
<td>30</td>
</tr>
<tr>
<td>Range</td>
<td>70</td>
<td>50</td>
<td>50</td>
</tr>
</tbody>
</table>

Table 2. 38 listeners’ age estimates of an atypical elderly male speaker.

3.2 Different sets of comparisons between speaker and listener groups

Differences between results from different speaker groups (in terms of the relationship between chronological and estimated age) and listener groups (regarding their age estimation abilities), such as male vs. female or young vs. older, were found in several of the previous studies. Other researchers, however, found no such indications (Kukol et al. 1980). This investigation did only find minor differences, which is probably due to the small size of the study. Also, a comparison of groups of trained vs. untrained listeners was made, but no notable differences were found.

4 Conclusions and discussion

Listeners were not always able to judge speaker age within ±10 years of chronological age, but the results show that for some speakers (i.e. the typical speakers) the estimates were considerably better than chance, which in this case could be considered as $3/18 \approx 16.7\%$ as
there were 18 alternatives in the perception test. Between 19 (50%) and 35 (92%) of the 38 listeners made correct age estimates within ±10 years for the four typical speakers, whereas the result obtained for the atypical voices was poorer; only between 7 (18%) and 22 (58%) correct age estimates within ±10 years.

The better results for the typical speakers suggest that a subdivision of voices into typical and atypical categories could help in further understanding the cues of age perception, for instance when integrating age-related parameters in speech synthesis.

The monosyllabic word ‘tack’ got the poorest results for all speakers. It did not seem to hold enough information about the speaker to enable several of the listeners to make reasonably good age estimates. This indicates that more than a monosyllabic word is probably needed in order to judge age correctly.

The fact that the results of different studies do not always agree, indicates that more research in this field is needed to fully understand the acoustic and perceptual age-related properties of speech.

References
Kukol R, Hutchinson, D. & Duane, M. 'Listener group comparisons of perceived speaker age'. Paper presented at the ASHA convention in Detroit, Mi.