Stimulus Duration and Type in Perception of Female and Male Speaker Age

Susanne Schötz

Department of Linguistics and Phonetics
Lund University, Sweden
susanne.schotz@ling.lu.se

Abstract

In a small perception study, our ability to estimate speaker age from speech samples was investigated with respect to stimulus duration, stimulus type and speaker gender. Four separate listening tests were carried out with four different sets of stimuli: 10 and 3 seconds of spontaneous speech, one isolated word, and 6 concatenated isolated words, all produced by the same 24 speakers. The results showed that the listeners' judgements were about twice as accurate compared to a baseline estimator, and that both stimulus duration and type affected the judgements. It was also found that stimulus duration influenced the listeners judgements of female speakers somewhat more, while stimulus type affected the age judgments of male speakers more, indicating that listeners may use different strategies when judging female and male speaker age.

1. Introduction

Most of us are able to make fairly accurate estimates of an unknown speaker's chronological age from hearing a speech sample, perhaps because of a constant confrontation with this task throughout our lives, for instance when answering the telephone, or when listening to the radio [1, 2, 3]. This paper adresses the question of how much and what kind of speech information we need to make as good estimates of speaker age as possible.

1.1. Background and previous studies

In an age estimation task, the accuracy depends, among other things, on the precision required and on the duration and type of the speech sample (prolonged vowels, read speech etc.). The less acoustic information present in a speech sample, the more difficult the task, but even with very little information, listeners are still not reduced to random guessing. Speaker and listener characteristics, including gender, age group, the speaker's physiological and psychological state, and the listener's experience or familiarity with similar speakers (dialect etc.) may also influence the accuracy [4, 3]. Consequently, some speakers may be more difficult to judge than others.

A considerable amount of research has been devoted to the issue of age recognition from speech [5, 1, 6, 7, 2, 8, 3, 9, 10]. Unfortunately, these studies are often difficult to compare due to differences in the stimuli as well as in the method. Differences concern (1) language, (2) stimulus duration, (3) speech type (prolonged vowels, whispered vowels, single words, read, backward or spontaneous speech etc.), (4) sound quality (HiFi, telephone-transmitted etc.), (5) speaker age and gender, (6) listener age and gender, (7) recognition task (classify into 2, 3 or 7 age groups, direct magnitude etc.), and (8) result measure (correlation, absolute mean error, % correct etc.).

Another question concerns whether listeners use different strategies when estimating female and male speaker age, since women and men age differently [11]. In a study of automatic estimation of elderly speakers, Müller et al. [12] successfully built gender-specific age classifiers. The author [13] found differences between female and male speakers in both human and machine age recognition from single word stimuli. While F0 was a better cue for estimation of female age, the formants seemed to constitute better cues when judging male age. One possible explanation is that the characteristics of female voices appear to be perceived as more complex than those of male speech [14], suggesting that listeners would need either a partly different set or a larger number of phonetic cues when judging female age.

1.2. Purpose and questions

The purpose of the present study was to determine to what extent stimulus duration and two different stimulus types (isolated words and spontaneous speech) influence estimation of female and male speaker age by answering the following questions:

1. In what way does stimulus duration and type affect the accuracy of listeners' perception of speaker age?
2. Is there a difference between perception of female and male speaker age with respect to stimulus duration and type?

2. Material and method

Figure 1: *The southern part of Sweden (Götaland = below the line) from where the speakers were selected.*

2.1. Material

Six speakers each from four different groups – older women (aged 63-82), older men (aged 60-75), younger women (aged 24-32) and younger men (aged 21-30), all from the southern part of Sweden (see Figure 1) – were selected semi-randomly from the SweDia 2000 database [15], which contains elicited isolated words and spontaneous narratives of non-pathological native speakers of Swedish, recorded in their homes. For each of the 24 speakers, four different speech samples were extracted, normalized for intensity, and used in the listening tests:

- Test 1: about 10 seconds of spontaneous speech
- Test 2: about 3 seconds of spontaneous speech
- Test 3: a concatenation of 6 isolated words: käke (jaw), saker (things), själen (the soul), sot (soot), typ (type) and tack (thanks) (dur.≈4 sec.)
- Test 4: 1 isolated word: rasa (collapse) (dur.≈0.65 sec.)

2.2. Method

Four separate perception tests – one for each of the four sets of stimuli – were carried out. Two listener groups participated in one test each, while a third group took part in two of the tests. The gender and age distribution for the three groups is shown in Table 1, along with information on which test and set of stimuli each group was presented with. All subjects were students of phonetics at Lund University. The task was to make direct age estimations based on first impressions of the 24 stimuli, which were played only once in the same random order in all four tests using an Apple PowerBook G4 with Harman Kardon's SoundSticks loudspeakers. The listeners were also asked to name cues, which they believed had affected their judgements.

Table 1: *Test number, stimuli set, number of listeners, and gender and age distribution of the listener groups in the four tests.*

Test (stimuli)	N	F	M	Age range (mean/median)
1 (10 sec.)	31	18	11	19-65 (27/22)
2 (3 sec.)	33	22	11	19-57 (25/23)
3 (6 words)	37	33	4	19-55 (26/24)
4 (1 word)	37	33	4	19-55 (26/24)

3. Results

3.1. Accuracy

Figure 2 displays the mean absolute error, i.e. the average of the absolute difference between perceived age (PA) and chronological age (CA) in years, for female, male and all speakers in the four tests, while Figure 3 shows the corresponding correlations between CA and PA. Since the two graphs display similar results, and as it has been suggested that mean absolute error is a more realistic measure for listener accuracy [2], only this measure will be mentioned and discussed below.

The listeners' judgements were about twice as accurate as a baseline estimator, which judged all speakers to be 47.5 years old (the mean CA of all speakers) in the first three tests. In Test 4, the shortest (1 word) stimuli yielded results at levels approximately half-way between the baseline and the other tests. The sum, mean and median values of the errors for all speakers in the four tests as well as for the baseline are shown in Table 2.

In all of the four tests, the listeners' judgements of women were more accurate than those of men. The highest accuracy was obtained for the female speaker 10 second stimuli (6.5), while the male speaker 6 word stimuli received the lowest accuracy (15.3). Moreover, the listeners tended to overestimate the younger speakers, and to underestimate the older speakers.

Table 2: *Sum, mean and median values of the absolute errors for all speakers for the four tests as well as the baseline (BL).*

Test	1 (10 s.)	2 (3 s.)	3 (6 w.)	4 (1 w.)	BL
sum	196.5	256.1	277.6	348.7	497.0
mean	8.2	10.7	11.6	14.5	20.7
median	7.2	10.0	10.0	16.7	19.5

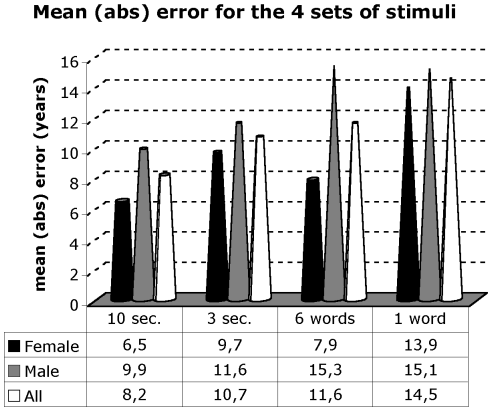

Figure 2: *Mean absolute error for the four sets of stimuli for female, male and all speakers.*

3.2. Stimulus and speaker gender effects

The listeners' mean absolute errors were subjected to two separate analyses of variance. In the first analysis, speaker gender and speaker age (old or young) were within-subject factors, and stimulus duration (short (1 word), medium (6 words and 3 sec.), long (10 sec.)) was the between-subjects factor. In the second analysis, the between-subjects factor was stimulus type (spontaneous or word stimuli) instead of stimulus duration.

3.2.1. Stimulus duration

Longer stimulus durations led to higher accuracy, an effect which was significant ($F(2,100)=71.059$, $p<.05$). A difference between the judgements for female and male speakers was also observed. Accuracy for increasingly longer stimuli improved more for the female than for the male speakers. For the female speakers, a lower mean absolute error was observed for the 10 second stimuli (6.5) compared to the 3 second stimuli (9.7), and the error for the 6 word stimuli (7.9) was lower than for the 1 word stimuli (13.9). The difference between longer and shorter stimulus durations was much smaller for the male speakers, with a mean absolute error of 9.9 for the longest (10 second) stimuli, higher errors for the medium long 3 second and 6 word stimuli (11.6 and 15.3), and a similar error for the 1 word stimuli (15.1). This interaction of speaker gender and stimulus duration was, however, not significant ($F(2,100)=2.171$, NS).

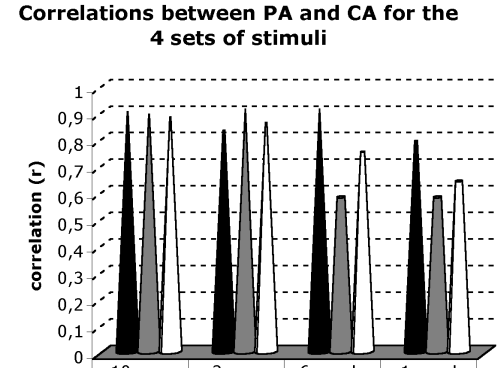

Figure 3: *Correlations between perceived age (PA) and chronological age (CA) for the four sets of stimuli for female, male and all speakers.*

3.2.2. Stimulus type

Stimulus type also influenced the age estimations significantly ($F(1,68)=61.143$, $p<.05$). The listeners' judgments of the male speakers were more accurate for the spontaneous stimuli than for the word stimuli. Lower mean absolute errors were obtained for the two sets of spontaneous stimuli (9.9 and 11.6) compared to the two sets of word stimuli (15.3 and 15.1). This effect was not observed for the female speakers. Here, the mean absolute error for the 6 word stimuli (7.9) was lower than for the 3 second spontaneous stimuli (9.7), but sightly higher than the longer spontaneous stimuli (6.5). The interaction of speaker gender and stimulus type was significant ($F(1,68)=39.296$, $p<.05$).

3.3. Listener cues

Most of the listeners named several cues, which they believed had influenced their age judgements. The cues mentioned by most speakers are presented in Table 3. Dialect, pitch and voice quality affected the listeners' estimates in all four tests, while semantic content influenced the judgements in the tests with spontaneous stimuli. A common listener remark in the tests with spontaneous stimuli concerned speakers talking about the past. They were often judged as being old, regardless of other cues. Additional listener cues included speech rate, choice of words or phrases and experience or familiarity with similar speakers (age group, dialect etc.).

Table 3: *Cues named by the listeners to affect their judgements.*

Test	Dialect	Pitch	Voice quality	Sem. content
10 sec.	24	23	25	22
3 sec.	25	26	32	17
6 & 1w.	33	31	34	0

4. Discussion and future work

Although only a limited number of stimulus durations and types were investigated in this study, a few interesting results were found. These are discussed below, along with a few suggestions for future work.

4.1. Accuracy

The listeners performed significantly better than the baseline estimator (about twice as good) in three of the tests, which is in line with previous work. However, it remains unclear what accuracy levels can be expected from listeners' judgements of age. When dealing with accuracy of perceived age, differences in speakers' chronological age have to be taken into account as well. A mean absolute error of 10 years could be considered less accurate for a 30 year old speaker (a PA of 20 could be regarded as 20/30 = 66.7% correct), compared to an 80 year old speaker (a PA of 70 could be regarded as 70/80 = 87.5% correct). Obviously, there is a need for a better measure of accuracy for age estimation tasks. The fact that three different listener groups participated in the tests may also have influenced the accuracy, as the between-listener group consensus was not checked.

In all of the four tests, the listeners' estimations of women were more accurate than those of men, perhaps because the listeners were mainly women themselves. However, the influence of listener gender on performance in age estimation tasks is still unclear. Although most researchers have not reported any difference in performance between male and female listeners, some studies have found females to perform better than males, while others still have found male listeners to perform somewhat better [8]. Another explanation could be that the male speaker group contained a larger number of atypical speakers, who consequently would be more difficult to judge, than the female speakers. Shipp and Hollien [1] found that speakers who were difficult to age estimate had standard deviations of nine years and over. Perhaps such a measure can be used to decide whether speakers are typical representatives of their CAs or not.

4.2. Stimulus effects

In this study, longer durations for the most part yielded higher accuracy for the listeners' age estimates. This raises the question of optimal durations for age estimation tasks. When does a further increase in duration for a specific speech or stimulus type no longer result in a higher accuracy? Further studies with a larger and more systematic variation of stimulus duration for each stimulus type are needed to answer this question.

Significant effects for both accuracy and speaker gender differences were found for the two stimulus types compared in this study. However, elicited isolated words and spontaneous speech can be difficult to compare in a study of speaker age. Several listeners mentioned that they were highly influenced in their judgements by the semantic content of the spontaneous stimuli, which may explain why the male speaker spontaneous stimuli yielded higher accuracy compared to the word stimuli. Besides providing more information about the speaker (dialect, choice of words etc.), spontaneous speech is also likely to contain more prosodic and spectral variation than isolated words. However, for the female speakers, the lower accuracy obtained for the 3 second spontaneous stimuli compared to the only slightly longer 6 word stimuli cannot be explained by stimulus type effects alone. It would be interesting to compare a larger number of speech types in order to find the types best suited for both female and male speaker age estimation tasks. Future work should include studies where several different speech types are compared and varied more systematically with respect to phonetic content and quality as well as variations and dynamics.

4.3. Speaker gender effects

As already mentioned in the previous paragraph, there were differences between female and male speakers with respect to which stimulus type and durations yielded higher age estimation accuracy. One explanation for the differences between female and male speakers may be that listeners use different strategies when judging female and male speaker age. As suggested in [13], it is possible that listeners use more prosodic cues (mainly F0) when judging female speaker age, but that spectral cues (i.e. formants, spectral balance etc.) are preferred when judging male speaker age. Consequently, the results from this study may indicate that for male speakers, spontaneous stimuli provide the listeners with more spectral information, while longer stimuli contain more prosodic information needed to estimate female speaker age more accurately. The differences in perception of female and male speaker age has to be studied further, and speaker gender has to be taken into consideration in future research, when investigating acoustic as well as perceptual correlates to speaker age.

5. Conclusions

Although more research is needed to verify and further explore the results of this study, two tentative conclusions are drawn:

1. Longer stimulus durations and spontaneous speech samples (which contain more speaker-specific information) seem to improve the accuracy of listeners' perception of speaker age.
2. There are differences between female and male age perception, possibly explained by differences in listeners' strategies when judging female and male age.

6. References

[1] T. Shipp and H. Hollien, "Perception of the aging male voice," *Journal of Speech and Hearing Research*, vol. 12, pp. 703–710, 1969.

[2] L. Cerrato, M. Falcone, and A. Paoloni, "Age estimation of telephonic voices," in *Proceedings of the RLA2C conference*. Avignon, 1998, pp. 20–24.

[3] S. E. Linville, *Vocal Aging*. San Diego: Singular Thomson Learning, 2001.

[4] L. A. Ramig and R. L. Ringel, "Effects of physiological aging on selected acoustic features," *Journal of Speech and Hearing Research*, vol. 26, pp. 22–30, 1983.

[5] P. H. Ptacek and E. K. Sander, "Age recognition from voice," *Journal of Speech and Hearing Research*, vol. 9, pp. 273–277, 1966.

[6] R. Huntley, H. Hollien, and T. Shipp, "Influences of listener characteristics on perceived age estimations," *Journal of Voice*, vol. 1, pp. 49–52, 1987.

[7] A. Braun, "Age estimation by different listener groups," *Forensic Linguistics*, vol. 3, pp. 65–73, 1996.

[8] A. Braun and L. Cerrato, "Estimating speaker age across languages," in *Proceedings of ICPhS 99*. San Francisco, CA, USA, 1999, pp. 1369–1372.

[9] S. Schötz, "Speaker age: A first step from analysis to synthesis," in *Proceedings of ICPhS 03*. Barcelona, Spain, 2003, pp. 2528–2588.

[10] M. Brückl and W. Sendlmeier, "Aging female voices: An acoustic and perceptive analysis," in *Proceedings of VOQUAL'03*. Geneva, Switzerland, 2003, pp. 163–168.

[11] M. B. Higgins and J. H. Saxman, "A comparison of selected phonatory behaviours of healthy aged and young adults," *Journal of Speech and Hearing Research*, vol. 13, pp. 1000–1010, 1991.

[12] C. Müller, F. Wittig, and J. Baus, "Exploiting speech for recognizing elderly users to respond to their special needs," in *Proceedings of Eurospeech 2003*, Geneva, Switzerland, 2003, pp. 1305–1308.

[13] S. Schötz, "Prosodic cues in human and machine estimation of female and male speaker age," in *Nordic Prosody. Proceedings of the IXth Conference, Lund, 2004*, G. Bruce and M. Horne, Eds. Frankfurt am Main: P. Lang, 2005, pp. 215–223.

[14] T. Murry and S. Singh, "Multidimensional analysis of male and female voices," *Journal of the Acoustical Society of America*, vol. 68 (5), pp. 1294–1300, 1980.

[15] G. Bruce, C.-C. Elert, O. Engstrand, and A. Eriksson, "Phonetics and phonology of the Swedish dialects - a project presentation and a database demonstrator," in *Proceedings of ICPhS 99*. San Francisco, CA, USA, 1999, pp. 321–324.

[16] D. Hartman, "The perceptual identity and characteristics of aging in normal male adult speakers," *Journal of Communication Disorders*, vol. 12, pp. 53–61, 1979.