The Role of F0 and Duration in Perception of Female and Male Speaker Age

Schötz, Susanne

Published in:
Proceedings of Speech Prosody

2004

LUND UNIVERSITY

PO Box 117
221 00 Lund
+46 46-222 00 00

The Role of F_0 and Duration in Perception of Female and Male Speaker Age

Susanne Schötz

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

Abstract

Single word stimuli from twelve female and twelve male natural speakers of various ages and from two synthetic voices were acoustically analyzed for duration and F_0. Listening experiments were carried out to test if spectral features or F_0 and duration provide the more dominant age cues and to test if listeners are equally good at estimating the age of female and male speakers. Results of the listening tests indicate that listeners are equally good at estimating the age of female and male voices and that spectral information is more important than F_0 and duration in age perception of both male and female speakers. Strong correlations of duration with biological and perceived age were found for both female and male voices, indicating that duration is an important cue for agedness. When correlating F_0 (mean, range, SD in Hertz and semitones) with biological and perceived age, the result was significant only for the female speakers (F_0 range and F_0SD in Hz), but no indication was found that listeners use different strategies when judging the age of female and male speakers. The acoustics and perception of speaker age will be studied further using a larger and more varied material.

1. Introduction

Phonetic differences between female and male speakers and between old and young speakers can be observed in both laryngeal and supralaryngeal information, the causes being physical (size and condition of the vocal organs) as well as social (environmental etc.) [1, 5, 7, 8, 14, 15]. Listeners are usually able to judge speaker age to within ±5 years of biological age, but there seems to be a subdivision into typical and atypical speakers, the former displaying a stronger correlation between biological and perceived age [11, 12, 13].

Some studies have found F_0 and F_0SD to be dominant cues to age perception [5, 6, 9]. However, several recent studies have failed to find strong correlations between measurements of F_0 and agedness, implying that other factors, including speech rate, shimmer and spectral features are more important in perception of speaker age [10, 13, 19].

One reason for the different results could be the various types of material used in the studies. Another reason may be the effect of gender on age perception. Men and women age differently. Effects of the aging process on phonatory behaviors like F_0 and EGG duty cycles seem to differ in degree and kind for male and female voices [4].

In a previous study [13] it was found that spectral information was dominant over F_0 in age perception of single word stimuli produced by male speakers. The present study includes both female and male speakers, and aims at increasing the phonetic knowledge of the various cues and listener strategies to female and male speaker age. To what extent do listeners use spectral features, F_0 and duration as cues to age? Is female and male speaker age equally easy to estimate? Do listeners use different strategies when judging the age of female and male speakers? The study described here is an attempt to answer the above questions.

2. Material and Method

The speech material consisted of 24 natural, 4 synthesized and 96 resynthesized versions of the single Swedish word *rasa* [ˈʁɑːsa] (collapse). 6 older women (age: 60-82), 6 younger women (age: 18-31), 6 older men (age: 60-76) and 6 younger men (age: 20-29) had produced the natural versions by elicitation. They all spoke the same dialect (Småländska) and were non-pathological speakers taken from the SweDia 2000 dialect database [2]. In order to facilitate analysis of between-speaker variations the natural stimuli were named after each speaker, i.e. the stimulus produced by the first old woman from village *A* was named *aow1*, the stimulus from the third young man from village *S* was named *sym3* etc. Four synthesized versions with monotonous F_0 values were created using 2 young speaker MBROLA-based concatenative synthesizers. LUCAS [3] produced 2 synthetic male stimuli (age: 30; F_0 values 80 & 110 Hz), and OFELIA [16] produced 2 synthetic female stimuli (age: 24; F_0 values 160 & 220 Hz). The 4 synthesized and the 24 natural versions were then used to produce new PSOLA-resynthesized stimuli using a "prosody-switching" script called 'Mix-Prosody!' (developed by Johan Frid, Dept. of Linguistics and Phonetics, Lund University) for the speech analysis tool Praat (www.praat.org). This script produced 2 new output stimuli AB and BA out of 2 input stimuli A and B; one with the spectral features of A, but the F_0 and duration of B, and one with the spectral features of B but the F_0 and duration of A, as shown in Figure 1.

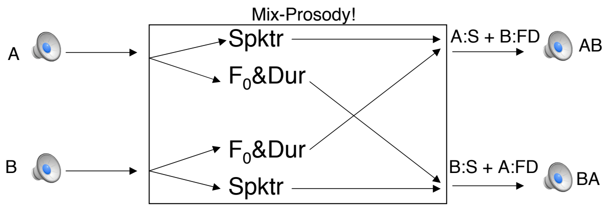

Figure 1: *Schematic diagram of the 'Mix-Prosody!' script.*

Two listening tests - one with female, one with male stimuli - were created. Each test was further divided into five parts based on the various stimuli types as described in Table 1. Part 1A and 3 comprised merely natural and synthesized versions, while the other parts contained only resynthesized stimuli. In part 1B the synthesized versions switched prosody with all older speakers, in part 2A one typical old speaker switched prosody with all younger speakers, and in part 2B one typical young speaker switched prosody with all older speakers.

Table 1: *The five parts and stimuli types of each listening test.*

Part	*Stimuli type*	*No of stimuli*
1A	6 older natural and 2 synthesized versions	12 pairs
1B	resynthesized versions (6 older natural and 2 synthesized versions)	12 pairs
2A	resynthesized versions (1 typical old speaker and 6 younger speakers)	6 pairs
2B	resynthesized versions (1 typical young speaker and 6 older speakers)	6 pairs
3	12 natural versions, 2 synthesized versions (all presented twice)	28

Word duration and F_0 (mean, range and SD) was measured for the natural and synthetic stimuli using Praat. All stimuli were normalized for intensity, and presented in a random order in all parts of the test using an Apple PowerBook G4 with Harman Kardon's SoundSticks speakers.

The subjects (students of phonetics at the Dept. of Linguistics and Phonetics, Lund University) were in part 1A-2B asked which stimulus sounded older out of each pair, and in part 3 to judge the exact age (in years) of each stimulus. 31 listeners (24 female, 7 male; age: 18-36, mean age: 21.6) took part in the test with female stimuli, while 29 listeners (24 female, 5 male; age: 18-28, mean age: 22,3) participated in the male speaker test.

3. Results

3.1. Part 1A, 1B, 2A and 2B of the listening test

In part 1A the older natural speakers were much more often judged to sound older than the synthetic versions, and in part 1B, 2A and 2B the resynthesized stimuli with *spectral features* of an *older speaker* in combination with *F_0 and duration* of a *younger speaker* (including the synthetic speakers) were significantly more often judged to sound older than the opposite combination with younger speaker spectral features and F_0 and duration of an older voice.

3.1.1. Female speakers

In part 1A the listeners judged the natural stimuli produced by older women to sound older than the two synthetic stimuli (Ofelia) in 88% of the stimuli pairs. Out of the 12% where Ofelia was judged to sound older, there was no large difference between the high F_0 version (5%) and the low F_0 version (7%). When the speaker responsible for the most "errors" (aow3, an atypical older speaker), was excluded from the results, this percentage rose to 92%. The result of part 1B was that 69% of the stimuli with older spectral features were judged older than the stimuli with older duration and F_0 by the listeners. In part 2A the result was somewhat poorer, with 66% older speaker spectral features judged older, and in part 2B results were even worse with only 57% of the stimuli judged older when they consisted of older spectral features.

Table 2: *The number and percentage of female spectral quality and F_0 & duration being judged older in part 1B, 2A and 2B.*

stimuli pairs judged older:	*no. of results*	*spectral quality*		*F_0 and duration*	
		no. of	*%*	*no. of*	*%*
1B. older speakers + OFELIA (mixtures)	372	258	69%	114	31%
2A. one typical older speaker + younger speakers (mixtures)	183	121	66%	62	34%
2B. older speakers + one typical younger speaker (mixtures)	181	104	57%	77	43%

3.1.2. Male speakers

The listeners judged the natural stimuli produced by older men to sound older than the two synthetic stimuli (Lucas) in 87% of the stimuli pairs in part 1A. Out of the 13% where Lucas was judged to sound older, the low F_0 version accounted for 9% and the high F_0 version for 4%. Two rather atypical male speakers (aom2 and aom3) were found. These speakers were judged younger than Lucas' low F_0 version, but older than the high F_0 version. In part 1B 93% of the older spectral features were judged to sound older than the older duration and F_0 features by the listeners. Only one speaker (aom3) failed to produce good results in this part. One of the atypical male speakers (aom2) was unfortunately chosen to represent a typical older speaker in part 2A. In 3 of the stimuli pairs (aom2 + aym1, sym1 & sym2) the results were in fact close to 50%. Therefore the complete results for this part is down to 69%. Part 2B displays somewhat better results, as 80% of the stimuli were judged older when they consisted of older spectral features in combination with younger duration and F_0 as compared to stimuli with the opposite combination.

Table 3: *The number and percentage of male spectral quality and F_0 & duration being judged older in part 1B, 2A and 2B.*

stimuli pairs judged older:	*no. of results*	*spectral quality*		*F_0 and duration*	
		no. of	*%*	*no. of*	*%*
1B. older speakers + LUCAS (mixtures)	347	322	93%	25	7%
2A. 1 typical older speaker + younger speakers (mixtures)	174	120	69%	54	31%
2B. 1 typical younger speaker + older speakers (mixtures)	174	139	80%	35	20%

3.1.3. Comparing the results of female and male speakers

The results were significant for both female and male speakers, though slightly better for the male speakers, except for part 1A.

Table 4: *Female and male χ^2-results of part 1A to 2B.*

part	*1A*		*1B*		*2A*		*2B*	
gender	*$\chi^2(1)$*	*p <*	*$\chi^2(1)$*	*p <*	*$\chi^2(1)$*	*p <*	*$\chi^2(1)$*	*p <*
female	219.882	.001	55.742	.001	19.022	.001	4.028	.045
male	194.253	.001	254.205	.001	25.034	.001	62.161	.001

3.2. Part 3 of the listening test

In part 3 the results were fairly good even though many older speakers were underestimated and several younger (as well as all synthetic) stimuli were overestimated. Mean standard deviation of the judgements for the female speakers was 13.4 years (13.0 for older, 12.3 for younger and 16.0 for Ofelia), and 12.7 years (11.0 for older, 15.1 for younger and 10.0 for Lucas) for the male speakers.

3.2.1. Female speakers

For the older women (to the left in Figure 2) the results range from nearly perfect (aow1, sow1) to fairly good (aow2, sow2, sow3). As expected from the previous parts, one atypical speaker (aow3) was judged much younger than her biological age. The two synthetic voices (in the middle of Figure 2) were judged much older than the "biological" age, as were two of the younger voices (to the right in Figure 2). However, the other young voices showed results very close to the biological age of the speakers.

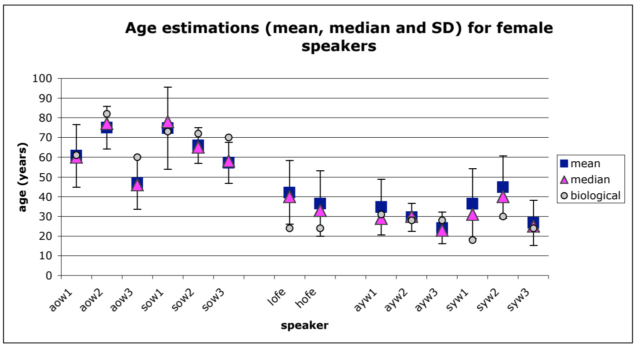

Figure 2: *Perceived (mean, median, SD) and biological age of the female speakers.*

When correlating the mean perceived age with the biological age, the result was significant for older women (r=.825; p<.05), but not for the younger (r=.097) voices.

3.2.2. *Male speakers*

Results for the older men (to the left in Figure 3) range from nearly perfect (som3) to underestimations at about 10-15 years (aom2, aom3, som1). Many of the younger speakers (to the right in Figure 3) were perceived as much older than the biological age, but some estimations came very close (aym2, sym2). The two synthetic versions (in the middle of Figure 3) were both overestimated, and the low F_0 version was perceived as older than the high F_0 version.

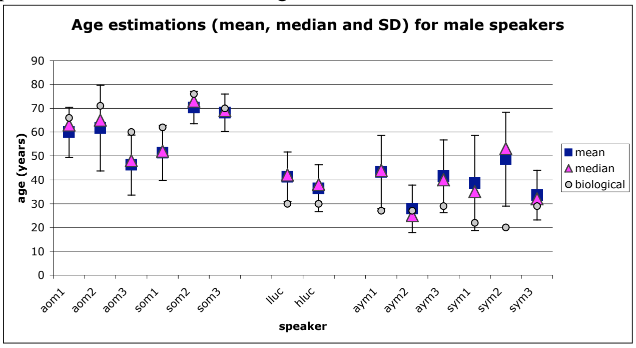

Figure 3: *Perceived (mean, median, SD) and biological age for male speakers.*

Correlations of mean perceived age and biological age was significant for the older speakers (r=.944; p<.05), but not for the younger speakers (r=.522).

3.2.3. *Comparing the results of female and male speakers*

The correlation between perceived and biological age was somewhat better for the male voices, both older and younger. Significant results were found only for the older speakers of both sexes, but the younger speaker results were slightly better for the male speakers.

Table 5: *Female and male r-values in part 3 of the listening test (significant values in boldface).*

	older		*younger*	
	r	*p <*	*r*	*p <*
female	**.825**	0.05	0.097	ns
male	**.944**	0.05	0.522	ns

The listeners did make fairly good age judgements. Still, the younger speakers were frequently overestimated, while the older speakers often were underestimated. No significant difference between the listeners' ability to judge the age of female and male speakers was found (t(30)=0.3; p>.05).

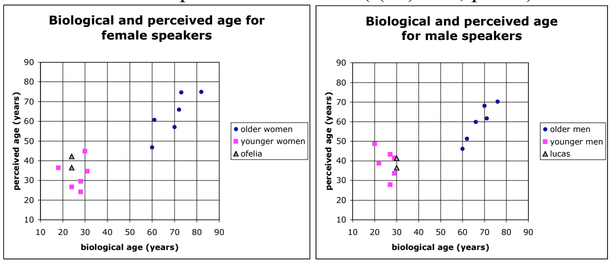

Figure 4: *Correlations of perceived and biological age for female (left) and male (right) speakers.*

3.3. Duration and F_0

Correlations of word duration with biological and mean perceived age were found for both male and female speakers, but for SD of mean perceived age only for female speakers.

Table 6: *Correlations of duration with biological age, mean perceived age and its SD for female and male speakers.*

Duration and	*biol. age*		*mean perc. age*		*mean perc. SD*	
	r	*p <*	*r*	*p <*	*r*	*p <*
female	**0.620**	.001	**0.756**	.001	**0.567**	.005
male	**0.598**	.005	**0.751**	.001	-0.220	ns

The acoustic values for mean F_0, F_0 range and F_0SD were also correlated with biological age, mean perceived age and its SD. Significant results were obtained only for female speakers. Though mean F_0 and F_0SD in semitones (ST) did neither correlate with biological age nor with mean perceived age for the female speakers, both F_0 range and F_0SD in Hz did. The strongest correlation was found between F_0SD in Hz and biological age.

Table 7: *Correlations of mean F_0, F_0 range and F_0SD with biological age, mean perceived age and its SD for female (A) and male (B) speakers (significant r-values in boldface).*

A	*biol. age*		*mean perc. age*		*mean perc. SD*	
female	*r*	*p <*	*r*	*p <*	*r*	*p <*
mean F_0	0.258	ns	0.248	ns	0.202	ns
F_0 range	**0.636**	.005	**0.544**	.005	-0.052	ns
F_0SD (Hz)	**0.670**	.001	**0.593**	.005	-0.027	ns
F_0SD (ST)	0.409	ns	0.201	ns	-0.277	ns

B	*biol. age*		*mean perc. age*		*mean perc. SD*	
male	*r*	*p <*	*r*	*p <*	*r*	*p <*
mean F_0	0.020	ns	0.185	ns	0.204	ns
F_0 range	-0.113	ns	-0.239	ns	0.032	ns
F_0SD (Hz)	0.239	ns	0.284	ns	0.295	ns
F_0SD (ST)	0.400	ns	0.447	ns	-0.218	ns

4. Discussion

The results from part 1B, 2A and 2B of the listening tests indicate that duration and F_0 are not as important as the spectral cues in the age judgments by listeners. This is in agreement with a previous study [13] of male speakers, where similar results were obtained with an older listening group (mainly staff at the Dept. of Linguistics and Phonetics, Lund University). A question that probably should be considered is whether university students and staff really represent a normal population. One way to avoid such problems would be to include naive listeners from various social and regional groups in future studies.

Atypical speakers sometimes affected the results of part 1A, 1B, 2A and 2B. It is possible that very different results would be achieved if a prior subdivision of speakers into age-typical and age-atypical speakers were made. In part 3 of the listening test listeners often over- or underestimated the atypical speakers. A possible explanation for this may be the influence of speaker-specific social factors, such as dialect, wisdom or naivety etc. that may be present in the voice. The effect of such factors on perceived age will have to be further examined.

The range of biological age is wider for the older speakers than for the younger ones, so the direct comparison above of correlation between the groups is likely to be a bit misleading.

Several studies have reported that female F_0 contains more variability in than male F_0 [7]. However, since F_0 in ST seems to be closer to perception than F_0 in Hz [17], it is possible that the importance of F_0 to age perception is equally small for female speakers as it is for male speakers. The fact that in this study, female F_0SD in Hz - but not F_0SD in ST - correlated with age, indicates that the distinction between acoustics and perception also involves differences between female and male voices. Not every acoustic variation is perceived. Furthermore, the different types of speech material used in different studies of agedness are likely to have influenced the results. This should probably be taken more into consideration in future analysis and synthesis of speaker age as well as in perceptual studies. Moreover, prosodic patterns, choice of words, grammar, sentence structure etc. in all probability also contribute to the perception of speaker age, especially in longer sequences of speech.

Comments concerning the acoustic part of the study include both duration and F_0:

Correlations of word duration with biological and mean perceived age were found for both male and female speakers, but not for the SD of mean perceived age. This may indicate that duration is an age cue, but that between-listener variation cannot be explained in terms of word duration.

When analyzing the F_0 range of the stimuli, segments containing creaky voice were not excluded. This may be one of the reasons why the male results were not significant, as speech with abrupt frequency halvings more often is perceived as having a coarser voice quality than being lower in pitch [18].

The findings of this study can only be thought of as tentative. Studies without creaky segments and with more varied speech material (prolonged vowels, CVC-structures, single words, short phrases, read speech and spontaneous speech etc.) will be carried out in further pursuit of phonetic knowledge on speaker age.

5. Conclusions

From the study the following conclusions may be drawn:

1. Listeners are able to make fairly good age estimations from a single two-syllable word.
2. It is likely that listeners are equally good at judging the age of female and male speakers.
3. Spectral features seem more dominant to age perception than F_0 and duration. Duration may still be important, but the role of F_0 remains unclear, although it may probably be minor.
4. The results from this study found no strong indications that listeners use different strategies when judging the age of female and male speakers.

6. References

[1] Biemans, M.; Van Bezooijen, R., 1999. Biological gender and social gender in relation to voice quality. *Proceedings of the XIVth ICPhS.* San Francisco. 1249-1252.

[2] Bruce, G.; Elert, C.-C.; Engstrand, O.; Eriksson, A., 1999. Phonetics and phonology of the Swedish dialects - a project presentation and a database demonstrator. *Proceedings of the XIVth ICPhS*. San Francisco. 321–324.

[3] Filipsson, M.; Bruce, G., 1997. LUKAS – a preliminary report on a new Swedish speech synthesis. *Working Papers 46*. Dept. of Linguistics and Phonetics. Lund University.

[4] Higgins, M.B.; Saxman, J.H., 1991. A Comparison of Selected Phonatory Behaviors of Healthy Aged and Young Adults. *JSHR*. vol. 34. 1000-1010.

[5] Hollien, H., 1987. "Old Voices": What Do We Really Know About Them? *Journal of Voice.* 1(1), 2-13.

[6] Jacques, R.D.; Rastatter, M.P., 1990. Recognition of Speaker Age from Selected Acoustic Features as Perceived by Normal Young and Older Listeners. *Folia Phoniatrica*. vol. 42, 118-124.

[7] Klatt D.; Klatt L., 1990. Analysis, Synthesis, and Perception of Voice Quality Variations Among Female and Male Listeners. *JASA*. 87(2): 820-857.

[8] Lindblad, P., 1992. *Rösten*. Lund: Studentlitteratur.

[9] Linville, S.E., 1987. Acoustic-Perceptual Studies of Aging Voice in Women. *Journal of Voice.* 1(1). 44-48.

[10] Minematsu, N.; Sekiguchi, M.; Hirose, K., 2002. Performance Improvement in Estimating Subjective Agedness with Prosodic Features. *Proceedings of Speech Prosody 2002.* Aix en Provence.

[11] Morris, R.J.; Brown, Jr., W.S., 1987. Age-related Voice Measures Among Adult Women. *Journal of Voice*. 1(1). 38-43.

[12] Ringel, R.L.; Chodzko-Zajko, W.J., 1987. Vocal Indices of Biological Age. *Journal of Voice*. 1(1). 31-37.

[13] Schötz, S., 2003. Towards synthesis of speaker age: a perceptual study with natural, synthesized and resynthesized stimuli. *PHONUM 9.* 153-156. (Umeå University, Dept. of Philosophy and Linguistics).

[14] Schötz, S., 2003. Speaker Age: A First Step From Analysis To Synthesis. *Proceedings of the XVth ICPhS.* Barcelona. 2585-2588.

[15] Stölten, K.; Engstrand, O., 2003. Effects of perceived age on perceived dialect strength: A listening test using manipulations of speaking rate and F_0. *PHONUM 9.* 29-32.

[16] Svensson, A., 2001. *Ofelia – en ny syntesröst*. M.A. thesis in Computational Linguistics, Dept. of Linguistics and Phonetics, Lund University.

[17] Traunmüller, H.; Eriksson, A., 1993. The frequency range of the voice fundamental in the speech of male and female adults. www.ling.su.se/staff/hartmut/f0_m&f.pdf (manuscript)

[18] Traunmüller, H.; Eriksson, A., 2000. Acoustic effects of variation in vocal effort by men, women, and children. *JASA*. 107: 3438-3451.

[19] Winkler, R.; Brückl, M.; Sendlmeier, W., 2003. The Aging Voice: an Acoustic, Electroglottographic and Perceptive Analysis of Male and Female Voices. *Proceedings of the XVth ICPhS.* Barcelona. 2869-2872.