The Aging Voice: an Acoustic, Electroglottographic and Perceptive Analysis of Male and Female Voices

R. Winkler, M. Brückl, W. Sendlmeier

Technical University Berlin, Germany

ABSTRACT

This study examines the differences between young and old adult voices. Acoustic cues in voices that enable listeners to recognize a speaker's vocal age are specified as well as acoustic cues that straightly indicate the speaker's chronological age. Electroglottographic data were used to directly examine glottal behaviour in aging voices.

We found a strong spectral attenuation of high frequencies in aging male voices assumed to result from rather sinusoidal glottal excitation. Increasing amplitude perturbation is an indicator of increasing age even on the basis of spontaneous speech. Reading rate decreases with increasing age, whereas there is no significant change in articulation rate of spontaneous speech in women's voices. Based on sustained vowels of female voices, the frequency tremor intensity index indicates age more accurately than F0 and amplitude perturbations. We also found evidence for the relevance of the vowel onset to recognize age more accurately.

1 INTRODUCTION

Our experiments investigate characteristics inherent in the speech signal that can be related to a concept of age. There has been a considerable amount of research concerned with acoustic phenomena of the aging voice as well as with the perception of aging voices. A comprehensive summary of the state of the art can be found in Linville [1].

The first studies investigating the aging voice were concerned with questions on the accuracy of age perception. According to Ptacek & Sander [2] listeners are able to assign single presented vowels correctly to two age groups (young vs. old) with a probability of 78%. Rating read speech, they even achieve 99%. Shipp & Hollien [3] found direct age estimations on read speech and chronological age correlating with r = 0.88. Factors influencing the accuracy of age estimations are the amount of information provided by the speech sample, the demanded precision of the age estimations, the dialect of the speakers and the age of the listeners. Linville [1] reports young and middle-aged adults to be the most accurate raters.

Acoustic examinations of the aging voice up to now included measurements of F0, intensity, perturbations (jitter, shimmer), spectral noise, speech velocity and formant frequencies. Relations of F0 and age are coherently reported: In female voices F0 is constantly decreasing with increasing age with a sharp decline around menopause. Men show a slightly decreasing curve followed by a strong increase starting around the 50th year. With increasing age fluent and controlled agility decreases. In speech production this affects vocal fold vibration, and therefore reflects in perturbation measurements. However, Ramig & Ringel [4] emphasize that frequency perturbations are likely to relate to physical fitness rather than to age. Ramig [5] also investigated vowel spectral noise as a function of age and physiological condition and found spectral noise to be related to fitness but not to age, at least in men's voices. Numerous studies demonstrated that older persons speak slower. Most of them concentrated on male speakers. If solely women were investigated, results are somewhat contradictory: Oyer & Deal [6] found a decline in articulation rate with increasing age in women. Hoit et al. [7] did not. Measurements of vocal tremor seem to be not yet analysed with respect to age. For the tremor measures applied in this study Xue & Deliyski report insufficient validity and did not use them in a comparable study [8].

First experiments dealing with extralinguistic information like speakers age and EGG data where done by Higgins & Saxman [9]. They found increasing contact duty cycle ratio values for men and decreasing values for women with increasing age. Glottal behaviour in the context of aging women was also interpreted indirectly by measurements based on oral airflow signal by Sapienza & Dutka [10]. Beside the weak relation between peak glottal airflow and chronological age they found no waveform features significantly correlated with age. We noticed strong correlation between spectral damping between 2 and 4 KHz and perceived age by listeners, and thus assume characteristic voice quality as the source for the characteristic spectral vowel shape. One factor of variation in the vowel spectra is the type of glottal excitation. The relation between glottal excitation and resulting spectral shape in the context of emotional speech was described by Klasmeyer & Sendlmeier [11]. They found steep glottal closure for happiness and anger with largely harmonic energy in all frequency bands, whereas sadness was related with nearly sinusoidal shape of glottal pulses and a marked spectral damping.

This paper summarises results of two experiments. We concentrate on changes in the acoustic and electroglottographic signal that come along with the chronological age of the speakers and with their perceived age.

 ISBN 1-876346-nn-n

2 EXPERIMENT A

The primary objective of this experiment is to investigate several aspects of the aging voice that according to Linville [1] have not been considered in detail, like articulation rate, amplitude SD and spectral noise as a function of age or of age estimates in female voices. Additional voice parameters supposed to vary with age are examined as well. The measured parameters are assumed to depict the voice qualities pitch, roughness, hoarseness, breathiness, speech velocity and, for the first time, vocal tremor. A perception test is accomplished to estimate the vocal age of the acoustically analysed voice samples. The perception test yields direct age estimations and does not distinguish age groups. A further aim of this experiment is to examine the relevance of the vowel onset on age-related information. Therefore two parts of equal length of each sustained vowel are analysed: the first containing the onset, the second without onset.

2.1 Data and Method

The voices of 56 women, mostly middle-aged (from 20 to 87 years; AM=49.77; SD=16.01) are recorded. All speakers judged themselves as healthy, especially as not suffering from hearing impairment or voice disease.
15 young adult listeners (6 female, 9 male, between 22 and 35 years of age (AM=28.67; SD=3.46), not hearing impaired) rated perceptual direct age of the speech samples. The reliability of these judgements was tested separately for each stimulus group via Kendall's coefficient of concordance (W). The lowest coefficient of concordance ($W=0.327$; $p<0.0005$) still denotes sufficient reliability.
Each speaker provided 8 different voice samples, assumed to differ in the amount and type of age-related information. They can be subdivided in (1-6) onset and quasi-stationary parts of the sustained vowels /a/ /i/ and /u/ (/a/-o, /a/-s, /i/-o, /i/-s, /u/-o, /u/-s), (7) read speech (r-sp) and (8) spontaneous speech (s-sp). For producing sustained vowels, speakers were instructed to keep pitch and loudness as constant as possible at their preferred levels. For read speech all speakers were reading the same visually presented path description applying their normal reading style. Spontaneous speech was created by describing a picture.
All stimuli were analysed with respect to their phonatory parameters via the Multi Dimensional Voice Program (MDVP) of KAY Elemetrics Corp.. This program was originally developed to obtain measures indicating degree and kind of voice diseases on the basis of sustained /a/-realisations. MDVP extracts 33 voice parameters. Only those parameters that hypothetically vary with age are examined in this experiment. They can be grouped according to their supposed perceptive equivalent. Pitch (average fundamental frequency (F0)) is supposed to decline in female voices with increasing age. Roughness or hoarseness (registered by perturbation measurements) is assumed to increase with age. The collected frequency and amplitude perturbation parameters measure either absolute variations or variations normalised by F0 or the mean amplitude, respectively. They also differ according to their time resolution, describing e.g. variations from one cycle to the next or variations from 55 cycles to the next 55 or variations from one cycle to all other cycles in the analysed sample. Breathiness (measured by the parameters of spectral energy distribution extracted from the long-term average spectrum) and vocal tremor (frequency tremor intensity index and amplitude tremor intensity index) are also expected to increase with age. The manually measured parameters of speech velocity (articulation rate , duration of read speech and accumulated duration of breaks) are expected to indicate slower speech in older voices.
All vocal parameters were correlated with the chronological age of the speaker as well as with the perceived vocal age of the stimuli.

2.2 Results and Discussion

Age Perception: Relevance of Vowel Onset The amount of information on age, which is controlled with the different stimuli, as well as smoking habits and accent are investigated according to their impact on age estimations. A multivariate analysis of variance testing the influence of these factors on the mean of age estimations as well as on the inter-listener variance of age estimations showed only the intra-speaker factor amount of information to be significant. The influence of the onset can also be shown in the correlations of perceived age per stimulus group and chronological age (see table 1). The estimations approximate chronological age more accurately, if vowel onset is provided. The strongest relation can be found between chronological age and the spontaneous speech stimuli (s-sp), which are the items produced most naturally and which therefore should carry most information on age. The accuracy of age estimations on read speech stimuli (r-sp) corresponds with accuracies reported e.g. by Shipp & Hollien [3]. The accuracy of isolated vowel estimations is, as expected, lower than s-sp and r-sp correlations. If chronological age is to be estimated on the basis of these data, the best results will be obtained via linear regression of spontaneous speech stimuli, by means of the following equation:

	/a/-o	/a/-s	/i/-o	/i/-s	/u/-o	/u/-s	r-sp	s-sp
r	.559	.443	.738	.603	.460	.344	.862	.864
p	.000	.000	.000	.000	.000	.005	.000	.000

Table 1: Correlations of the chronological age and the perceived vocal age of the different stimulus groups.

$$age = 1.37 * perc.age(s - sp) - 9.63 \qquad (1)$$

Vowel onset also contributes to appointing vocal age more accurately. Listeners' estimations agreed best for continuous speech. At least for /a/ and /i/ concordance is stronger in the onset-stimuli. /u/-realisations are generally hard to rate, which points to a relative lack of information on age in /u/-vowels. Other interesting results of the perception test are first that all listeners estimate the speakers as a group significantly younger than they are. In addition, old speakers as a group are rated more divergently than young speakers.

Acoustic Correlates of Age An overall comparison of the data reveals generally higher correlations of the acoustic parameters with perceived age than with chronological age. Relatively weak relations of F0 in /i/ and /u/ vowels can be explained by the intrinsic pitch phenomenon supposed to be responsible for the generally high pitch in these vowels. Apart from that, decreasing F0 in women is a reasonable predictor of increasing age. /u/-realisations do not prove to be very useful in indicating age, although listeners do relate elevated /u/ perturbation measurements to ascending age. Amplitude perturbation compared to frequency perturbation correlates much better with age, confirming former findings which assign jitter to physical decrease rather than to age. But regarding time resolution, Linville's [1] suggestions that measures reflecting more general fluctuations (e.g. standard deviation) are better predictors of age can not be found in these data. In fact it seems that shimmer measures with a time window somewhere between 5 and 55 cycles would be the most accurate perturbation measurements of age. Surprisingly, because of the unconventional application of MDVP the shimmer measures provide best results in connection with spontaneous speech. Even more surprisingly, frequency tremor intensity index indicates chronological age on the basis of sustained vowels more accurately than F0 and amplitude perturbations and furthermore seems to be the most reliable measure of age independent of vowel type. Articulation rate only seems to be important in read speech, suggesting that the slowing of women's articulation with age is better described as decline in reading or cognitive performance. The applied measures of spectral noise show only little importance for age perception in sustained /a/-realisations and no correlation with chronological age in female voices.

3 EXPERIMENT B

Because of anatomical and physiological changes in the vocal folds with aging it is likely to find acoustical evidence in the speech signal resulting from such laryngeal changes. One possibility to alter spectral distribution in vowels is to alter the behaviour of the vocal folds during oscillation. The aim of this experiment is to evaluate electroglottographic data qualitatively to study vocal fold behaviour in men and women as a function of perceived age.

3.1 Data and Method

In this experiment our data consist of the sustained German vowel /a/ spoken with self-selected comfortable pitch and loudness by 26 men and 24 women differing in age. Speakers were instructed to sustain vowel /a/ as stable as possible. Audio and EGG signals were recorded synchronously. The EGG signal was linear high pass filtered (cutoff frequency 40 Hz) to minimize baseline shift and nonvocal EGG influences. Because of the possible relationship between age and health special care was taken to record aged speakers judged healthy by a physician.
18 listeners estimated the age of the speakers on the basis

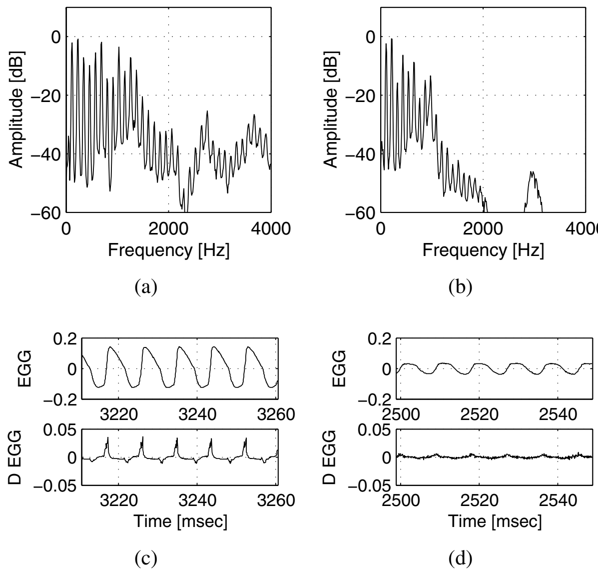


Figure 1: LTAS, EGG and differentiated EGG signals for men perceived as young (a, c) and old (b, d).

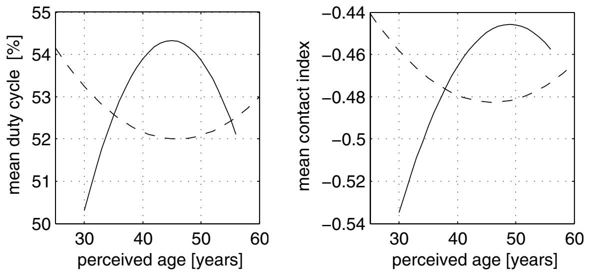
of the vowel segments. From the vowel segments of the 4 men/women perceived as eldest and youngest long-term average spectra (LTAS) where calculated and visually evaluated.
The glottal behaviour was interpreted by inspecting the electroglottograms. We further calculated contact duty cycle, contact index, and the EGG relative rise time along the lines of Orlikoff [12] over 5 glottal cycles per speaker centered around the midpoint of the vowel.

3.2 Results and Discussion

We found a characteristic spectral distribution in the /a/-vowel of male voices. LTAS of the stimuli of male voices perceived as old had remarkable little harmonic energy between 2 and 4 KHz (compare figure 1(a) and 1(b)). We did not find spectral damping in the LTAS of females perceived as young vs. old. The difference lies more in the amount of harmonic energy rather than spectral damping between 2 and 4 KHz.

Male Speakers To describe the effect of aging qualitatively, a 30 year and a 63 year old man were chosen.
The young male shows a prototypical EGG waveform (see figure 1(c)) for modal voice. The waveform is characterized by a steep increase of vocal fold contact in the closing phase (increasing signal), where closing phase is short and peak-to-peak amplitude is high. The opening phase (falling signal) lasts somewhat longer than the closing phase. The instant of glottal opening is clearly visible in the EEG-signal as is typical only in modal voice. That instant is seen as a knee in the upper panel and as a small minimum in the lower panel of figure 1(c)).
The EGG waveform of the old male (see figure 1(d)) is rather tense than modal. The peak-to-peak amplitude is

ISBN 1-876346-nn-n © 2003 UAB



(a) mean contact duty cycle (b) mean contact index

Figure 2: Quadratic polynomial least square fit between listeners age ratings and two parameters from EGG measurement. (solid lines = male, dashed lines = females)

very small. The shape of the EGG-signal is nearly sinusoidal. The steepness of the increasing signal slope is reduced and the maximum contact phase starts early and lasts relatively long. While the instant of glottal closure indicated by a maximum in the differentiated EGG signal is hardly detectable, the instant of glottal opening is not visible at all.

Considering all data, in stimuli perceived as very young or very old smaller contact duty cycle values are found (figure 2(a)). In the range of 40-50 years of perceived age the highest values were measured. A similar pattern was found for contact index and mean rise time. Values increase nearly linearly with increased perceived age. There is a tendency from a sharp glottal closure with a relative long contact phase to a less abrupt closure together with an overall shorter contact phase with increasing perceived age in male voices. While mean rise time increases with increasing perceived age the shape of the EGG signal tends to get more sinusoidal. The increasing values for contact index (from -0.54 to -0.44) means a trend of the contact phase from left-asymmetric to symmetric shape, which is in relation with longer rise times.

Female Speakers The shape of the relation between perceived age and contact duty cycle ratio is parabolic with a minimum at approximately 45 years (see dashed line in figure 2(a)). Stimuli perceived as very young correlate strongly and stimuli perceived as old correlate well with high values. Only weak relation between perceived age and relative rise time was found. Values tend to be greater for voices perceived as old. High values of contact index correlate with stimuli perceived as very young, while smaller values were measured for old voices. There is some evidence, however, that values increase for voices perceived as old.

4 CONCLUSION

The first experiment shows that the parameters correlating with chronological and perceived age are the amplitude perturbation, at best normalised by the mean amplitude and with a time resolution between 20 and 250 ms, the frequency tremor intensity index and F0. Conditionally correlated are measures of frequency perturbation and speech velocity. Thus, adult female speakers speak more quiveringly, lower, rougher and read aloud more slowly as they grow older. We also found evidence for the relevance of the vowel onset to recognize age more accurately.

The second experiment reveals that there is a tendency from a sharp glottal closure to a less abrupt closure with increasing perceived age in male voices. The EGG measurements do not support hypotheses for a tendency of glottal excitation to become more sinusoidal in women.

ACKNOWLEDGMENTS

This work was supported by the Deutsche Forschungsgemeinschaft (German Research Fund, SE 462/5-1).

REFERENCES

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

[2] P. Ptacek and E. Sander, "Age recognition from voice," *JSHR*, vol. 9, pp. 273–277, 1966.

[3] T. Shipp and H. Hollien, "Perception of the aging male voice," *JSHR*, vol. 12, pp. 703–71, 1969.

[4] L. Ramig and R. Ringel, "Effects of physiological aging on selected acoustic characteristics of voice," *JSHR*, vol. 26, pp. 22–30, 1983.

[5] L. Ramig, "Effects of physiological aging on vowel spectral noise," *Journal of Gerontology*, vol. 38, pp. 223–225, 1983.

[6] E. Oyer and L. Deal, "Temporal aspects of speech and the aging process," *Folia Phoniatrica*, vol. 37, pp. 109–112, 1985.

[7] J. Hoit, K. Hixon, M. Altman, and W. Morgan, "Speech breathing in women," *JSHR*, vol. 32, pp. 353–365, 1989.

[8] S. A. Xue and D. Deliyski, "Effects of aging on selected acoustic voice parameters: preliminary normative data and educational implications," *Educational Gerontology*, vol. 21, pp. 159–168, 2001.

[9] M. Higgens and J. Saxman, "A comparison of selected phonatory behaviors of healthy aged and young adults," *JSHR*, vol. 34, pp. 1000–1010, 1991.

[10] C. M. Sapienza and J. Dutka, "Glottal airflow characteristics of women's voice production along an aging continuum," *JSHR*, vol. 39, pp. 322–328, April 1996.

[11] G. Klasmeyer and W. Sendlmeier, "Voice and emotional states," in *Voice Quality Measurement*, R. Kent and M. Ball, Eds., pp. 339–357. Singular Thomson Learning, San Diego, 2000.

[12] R. Orlikoff, R.J. Baken, and D. Kraus, "Acoustic and physiologic characteristics of inspiratory phonation," *JASA*, vol. 102, no. 3, pp. 1838–1845, September 1997.

ISBN 1-876346-nn-n © 2003 UAB