Women's Vocal Aging: a Longitudinal Approach

Markus Brückl

Institute of Communications Research, Technische Universität Berlin, Berlin, Germany

brueckl@kgw.tu-berlin.de

Abstract

A quasi-experimental longitudinal paired-samples study was carried out to explore, whether aging for 5 years can (1) audibly and (2) measurably change women's vocalisations, and if so, on which acoustic information (3) the listeners' performance possibly could relay on and (4) which parameters can contribute to detect the chronological difference.

Results indicate that (1) listeners can significantly correctly judge this difference based on sustained /i/ and /u/ vowels, but much better based on (spontaneous) speech samples. (2) Parameters depicting pitch, vowel resonance, voice perturbations, tremor and spectral energy distributions differ (significantly) between chronologically and perceptually younger and older samples. (3) Listeners tend to judge vowel samples as older, if increased (amplitude) perturbations can be measured, but in speech samples there seem to be overriding and objectively more reliable features. (4) The most reliable age-indicating measures in this study in speech samples are durations/ tempo measures – in vowels F_0 and tremor.

Index Terms: speaker age, human perception, acoustic-phonetic measures, longitudinal study

10.21437/Interspeech.2007-379

1. State of research

In the recent years speech sciences have considerably extended their focus on the investigation of extra- and paralinguistic information in speech. Primarily because non-linguistic speech information is on the one hand worth surveying for its own sake, on the other hand influencing the assessment of linguistic information, but further encouraged by improving technological possibilities as well as an growing social interest in how we age and how we can cope with it, research in the Acoustics of Aging has increased. A comprehensive summary of the state of research can be found in [1].

1.1. Human perception of speaker age

The modern investigation of acoustic-phonetic changes due to aging can be traced back to the 60s, when [2] found that listeners can correctly judge, if presented vowels were produced by either young or old speakers with a probability of 78%. Rating read speech, they even achieve 99%. [3] found judgments of age in years correlating to spontaneous speech samples at r=0.88 as well as a decreasing trend of information on age, starting with spontaneous speech over read speech to sustained /i/, /a/, /u/ vowels. [4] reports of speaker gender being the main factor of variance in a comparable study, using single words and spontaneous speech. Nevertheless, major factors influencing the accuracy of age estimations seem to be the overall amount (operationalised by concepts like duration, linguistic variance) of information encoded in speech and the difficulty of the listening task for / demanded precision of the age estimation. In a further perception study [4] "found that F_0 and duration are probably less important than non-prosodic cues" (everything else), but in the acoustical analysis "it was found that speech rate and intensity range seem to constitute the most important correlates of speaker age". In [5] it was shown that both, speech rate is highly relevant for the aural perception of age derived from synthesized single words. Glottal chink area was found more relevant than F_0.

Listeners report [cf.1] elderly speakers to speak slower, less precise, with longer pauses, less intense, at a lower pitch, and with a more harsh/rough, breathy and tremulous voice.

1.2. Acoustic measures of speaker age

The rather good human performance raised questions on how to specify and parameterise the acoustical information on the speakers' age that the listeners obviously can relay on, and further, if there are especially suited parameters to assert a speaker's chronological age and to which certain degree this can be achieved. Up to now acoustic assessments of the aging voice were mostly obtained by measurements that are related to the following concepts.

Speech tempo: Several studies demonstrated that older persons speak slower, but if solely women were investigated, results are contradictory: [6] found a decline in articulation rate with increasing age in women. [7] did not. **F_0**. Relations of F_0 and age are coherently reported: In female voices F_0 is constantly decreasing with increasing age; the relevance of measures of global F_0 dispersion (e.g. range, minimum, maximum) is strongly dependant on the speaking task. **Formants**: [8] examined in a longitudinal study formant frequency changes in 7 phoneme groups, produced by 4 men and 2 women over a time span of 13-15 years. They found that all "points of formant concentration" are lowering during 5 year steps continuously. [4] reports of a lowered F1, obtained in a study of 527 speakers. **Perturbations of vocal cord vibration**: [1] asserts that "firm conclusions as to the effect of aging on jitter and shimmer levels are not now possible". [3] found, that shimmer is a good indicator of age. [4] reports that perturbation measures did not yield much information on age. **Modulations of vocal cord vibration**: [3] found the intensity of the frequency tremor to be a better indicator of age than F_0 based on the investigation of sustained vowels. **Spectral energy distributions**: [1] states that "research is necessary to examine spectral noise as a correlate of perceived age estimates from women's voices". [3 and 4] found no obvious variation of parameters of spectral energy distribution as a function of age.

1.3. Methodological considerations

Except for the validated but moderate relevance concerning F_0 and duration/ tempo measures there are still many questions. Moreover the human precision of age perception is far from being explained and even further from being reached by automatic systems.

One major problem is that the variation in speech due to linguistic and "not primary interesting" non-linguistic modulations can up to now hardly be controlled and thus concepts of interest are hard to isolate successfully.

One approach to overcome this problem is to **use the human recognition ability**. Using the aurally perceived age provides the advantage of relying solely on the speech signal and leaving behind the variation that is introduced by differential biological aging of the speakers when their chronological age is used.

Another approach to eliminate a great part of the "irrelevant" variation is to conduct a **longitudinal study**: such a procedure yields the omission of inter-personal variation, hence highly comparable speech samples, and allows for a (quasi-) experimental design, combined with the advantage of investigating natural speech sounds. But although it is quite obvious that a longitudinal study seems to be predestined to investigate changes due to aging, this design has rarely [cf. 8] been used.

The study presented here combines these two approaches in order to (dis-) validate previously found correlates of vocal aging. Main questions to be surveyed are (1) if the voice and/or the speaking style changes perceivably during 5 years, (2) which parameter differences can be inferred to rely on this chronological difference and (3) which parameters possibly influence age perception.

2. Data and methods

The **design** of the presented study can be classified as quasi-experimental longitudinal paired-samples-investigation, based on natural utterances.

9 adult female **speakers**, 26, 27, 27, 30, 39, 40, 61, 67, and 87 (AM=44.89; SD= 21.79) years old when recorded the first time and correspondingly 5 years older when they were recorded the second time, participated in the presented study. They provided read and spontaneous German speech as well as the sustained vowels /a/, /i/ and /u/. The vowels were split up into three segments of 2.2 seconds each, containing either the vowel onset, the offset or a quasi-stationary middle part. The text, on which the **read speech** was based on, consists of 3 sentences of a road description. **Spontaneous speech** was induced by the task of a picture description. The speech samples were manually segmented and classified on phoneme, syllable, and word level.

2 **perception tests** were created, one for the speech and one for the vowel samples. During the test procedure the samples were randomised newly for each listener. Each sample pair was presented twice, in both orders one time. Listeners were instructed to listen to the samples as often as needed for the judgment, which of the two presented samples of the same speaker sounds older. 34 listeners, (age AM=31.35 SD=8.79; 14 male, 20 female) participated in the speech test, 31 of them (age AM=31.32 SD=9.11; sex 12 male, 19 female) in the vowel test.

The **acoustic-phonetic analyses** mainly parameterise the concepts duration, tempo, pitch, hoarseness/ roughness/ breathiness, vocal tremor, and vowel quality. They were computed by the Multi Dimensional Voice Program (MDVP) by KAY-Pentax and/or with PRAAT [9]. All parameters of the MDVP, covering measures of the fundamental frequency, frequency and amplitude perturbations, voicing, tremor and spectral energy distribution were observed, since they all are supposed to indicate vocal pathology and thus should also indicate age – according to the "wear and tear theory" of aging. The MDVP parameters were computed on all samples without using any segmentation, although this program originally is optimised for sustained /a/ vowels as input. Using PRAAT, the corresponding measures were computed as far as possible (voice report section) to get a second measure of the relatively unreliable voice parameters. Additionally the centre frequencies of the first 4 formants and the "point of formant concentration" [cf. 8] were determined.

Further, the PRAAT voice and resonance parameters were averaged segment-wise in the read speech samples. Tempo (speech rate and articulation rate) as well as durations, to-total-duration rates and quotients of durations of segments of these segment-groups were additionally appraised. Segment groups for these procedures included: vowels, diphtongs, laryngalisations, /a/, /i/, /u/ vowels, lexically stressed vowels, lexically unstressed vowels. Durations, rates and quotients were as well computed for segment groups like consonants, stops, fricatives, nasals, voiced consonants, unvoiced consonants, pauses, breathing-in pauses, breathing-out pauses.

The **statistical analyses** include χ^2 frequency tests (H_0: equal frequencies) to evaluate the chronological correctness of the listeners' judgments and the intra-listener consistency as well as mean comparisons (Wilcoxon Test) to evaluate the parameter differences for each age grouping – chronological and perceptive. The perceptive age grouping was determined by χ^2-testing comparable (same speaker and sample type, differing speaker age) samples for unequal (p=0.1) rating in either direction. Only sample-pairs being rated unequal remain in the analysis and with their perceived age label.

The status of most of the acoustical parameter's results of this study must be considered explorative.

3. Results

3.1. Human perception

Speech Perception: Table 1 summarises results of two different χ^2-tests for each sample subset: The first test's alternative hypothesis (H_1) is: "There are more chronologically correct listeners' judgments." The second test's H_1 is: "There are more listener-internally equal judgments of same sample-pairs." It can be observed that listeners are able to judge significantly correctly – with a probability of approximately 2/3 if the presented speech samples were older or younger. The spontaneous speech can be judged slightly easier chronologically correctly, indicating that there is more information on age in these samples than in read speech. This observation is also supported by the evaluation of listener-internal consistency.

Table 1. *Results of the statistical analyses of the listeners' judgments of the speech test.*

sample type	listeners' judgements are	frequencies		χ^2	p(sig.)
		yes	no		
all speech	correct	817	407	137.337	0.000
	equal	417	195	80.529	0.000
read speech	correct	396	216	52.941	0.000
	equal	200	106	20.876	0.000
spon. speech	correct	421	191	86.438	0.000
	equal	217	89	53.542	0.000

Vowel Perception: The analysis of the judgments of the vowel samples (cf. Table 2) reveals a more complex situation: Taken together, the vowels seem to provide no information on

the chronological age difference. But observed in detail it becomes obvious that the /a/ vowels are rated significantly falsely, while the /i/ and the /u/ vowels both are rated correctly. This indicates, that there is misleading information especially in the presented /a/ vowels.

Table 2. *Results of the statistical analyses of the listeners' judgments of the vowel test.*

sample type	listeners' judgements are	frequencies		χ^2	p(sig.)
		Yes	no		
all vowel parts	correct	2555	2467	1.542	0.214
	equal	1650	861	247.918	0.000
all /a/ parts	correct	700	974	44.848	0.000
	equal	571	266	111.141	0.000
all /i/ parts	correct	931	743	21.114	0.000
	equal	542	295	72.890	0.000
all /u/ parts	correct	924	750	18.086	0.000
	equal	537	300	67.108	0.000
all onset parts	correct	878	778	5.974	0.015
	equal	544	293	75.270	0.000
all stat. parts	correct	821	853	0.612	0.434
	equal	540	297	70.548	0.000
all offset parts	correct	847	827	0.239	0.625
	equal	566	271	103.973	0.000

A similar effect probably emerges if different vowel segments (onset, quasi-stationary, offset) are compared – although moderated by the stronger variation due to vowel quality: the quasi-stationary middle parts are judged false, although not significantly. This indicates that there is more information on age in the outer vowel segments.

A look at the listener-internal consistency evaluation shows that it is just slightly worse than the one of the speech test, indicating that the listeners have a concept of speaker age to base their vowel judgment on, but a partly misleading one.

3.2. Acoustic differences in chronologically grouped samples

Changes in speech samples: The in-total-analyses of all speech samples show in the older (lately recorded) samples lowered amplitude perturbation measures, a lowered first Formant (F_1), a raised F_3 and a relative increase of lower harmonic energy compared to higher harmonic as well as turbulent energy (SPI, HNR). The same analyses, performed solely on spontaneous speech samples still demonstrate a strong trend for increased amplitude perturbation measures in the younger samples and a higher breathiness (higher NHR, lower HNR) but also a lower intensity of frequency tremor (FTRI). Read speech reveals in this analysis longer overall durations in the older samples. The speech tempo measures syllables per second (p=0.033; df=8) and phonemes per second (p=0.014) belong to the most reliable acoustic indicators of chronological age for speech samples in this study. Even better performs the quotient of the durations of canonically voiced vs. those of unvoiced phonemes, indicating more voiced duration in older speech. But this measure just as other nearly highly significant parameters as e.g. the minor duration of canonically unstressed vowels (p=0.008; two-tailed) in older has not been assessed for measuring age before and thus awaits further validation.

The voice and formant measures in the segment(-group-)wise analyses of the read speech samples also vary as a function of chronological age, but not as consistently as duration measures: In the old raised SD(F_0) as well as higher F2, F3 and again lower amplitude perturbation can be observed for the group of vowel segments. The diphtongs finally provide a in older samples highly significantly lowered F_1 ($p<0.008$, one-tailed) as well as statistically even more extraordinarily raised F_2, F_3 (both p=0.011; two-tailed). But due to their more sporadic appearance these measures are considered less reliable to indicate chronological age than the duration/ tempo measures.

Even the pauses seem to carry information on age: The number, the mean duration, the to-total-duration-rate and the to-pauses'-total-duration-rate of breathing-out-pauses perform as good as the standard tempo-measures.

Changes in sustained vowel samples: A broad variety (32 parameters) of the used voice quality and formant measures shows significant differences in these samples. All sustained vowels analysed in the same statistical test reveal in the younger samples raised F_0, min. F_0, max. F_0, more lower harmonic spectral energy, lower F_3 and F_4, but again strikingly often increased values of most of the applied amplitude and pitch perturbation measures.

These higher perturbations in the set of all vowels can be deduced to mostly the same effect in the subset of /a/ vowels and weakened in /i/ vowels. However /u/ vowels do show significantly lower values of (long-term-) perturbations in the younger samples. Also the sum/integral of the frequency tremor intensity contour (a derivate measure of the MDVP tremor section) is increased in the older /u/ samples, indicating more intense frequency tremor.

3.3. Acoustic differences in perceptually grouped samples

Changes in speech samples: The in-total-analyses of the speech samples demonstrate again in the perceptually older samples – just like in chronological grouping – decreased amplitude perturbation, increased HNR and SPI. The differences in amplitude perturbation and spectral measures is found both in spontaneous and in read speech subsets. Spontaneous speech additionally shows an increased pitch range due to a lowering of the min. Pitch as well as again a higher F2 in the older samples.

The segment(-group-)wise analyses of the read speech samples – again like those of chronological differences – show a decrease of the syllable per second measure but only a nearly significant difference in phonemes per second. But again the quotient of the durations of voiced vs. unvoiced phonemes, indicating more voiced duration in older speech performs better than both established (tempo) measures. The pauses' total duration as well as it's to-total-duration-rate are higher in the perceptually older samples. The voice quality and formant measures again vary more arbitrarily in segment-wise extraction, also between perceptually differing samples: In older samples SD(F_1) and F_0-perturbations are higher. Duration and HNR are increased, amplitude perturbation is decreased. A lowered F_1 for all vowels can be explained as a result of more extreme lowered F_1 in the subset of the diphtongs again.

Changes in sustained vowel samples: The selected voice measures (also massively the MDVP perturbation measures, which were scarcely significant in the other analyses) provide even more (44 in number) significant differences than in the respective analysis of chronological correlates. The direction of the parameter differences remains also the same as for the chronological analysis, but with two contradictions:

perturbation parameters are raised in the perceptually older samples as well as F_2 is decreased. This can be explained as a consequence of the stronger increased perturbation in the subset of sustained /a/ vowels in perceptually older samples.

4. Discussion

Human perception: Although it was reported by the listeners that the judging task was very hard, they could even rate the vowel samples correctly to a satisfying extent. That indicates that already the difference of five years yields audible changes in women's voices, assuming that a sustained vowel contains only information on the construct voice including vocal tract resonance features. Perceived together with the temporal (and highly probable as well prosodic) changes that additionally occur in speech samples, there is enough information to rate this minimal age difference correctly with a probability of approximately 2/3. The findings of [3] with respect to the degree of information on age in vocal utterances could be replicated, and partly the findings of [4] concerning the stimulus-relative perceptive relevance of different parameters.

Acoustic measures: Again the tempo measures were validated as chronologically relatively reliable and potentially perceptive relevant indicators of age, as well in women's speech, confirming the findings of [6]. But it must be noted, that (pure) duration measures again performed better, as previously found [cf. 3]. F0 is also validated in it's function as indicator of age – although as a less reliable one. The formants did behave as expected by projecting the results of [4]: Only F1 can serve as an indicator of age, interestingly at best assessed in diphtongs. The relevance of the by [8] introduced parameter "point of formant concentration" could not be validated. This possibly could be explained as an effect of automation of formant measurement. Tremor measures could as well be seen differing between age-groups, but the by [3] objected relevance could not fully be confirmed.

Combining the results: The counter-hypothetical changes in amplitude perturbation measures are at first sight very conspicuous, especially compared to the findings of [3]. But seen together with the results of the perception test, it may be assumed that the selected speakers show increased amplitude perturbations in their younger recordings either by accident or due to intra-personal short-term (emotion, arousal, …) differences, which have not been raised. So increased amplitude perturbations still seem to be a good predictor of advanced age. But they as well seem more easily influenced by short term variations than other measures like durations or tempo, and thus are the reason for false judgments on hardly to distinguish samples.

But regardless if the raised amplitude perturbation in (chronologically) younger samples was pure accident, it was a revealing accident. It made possible to test very reliably the relative perceptive relevance of perturbations against the relevance of additional information on age supplied in speech. Perceptive concepts that are depicted by perturbation measures are focused by the listener when there is only little information but their impact can easily be overruled by other, probably more reliable, features – at least in this study.

5. Conclusions

Humans are able to judge, based aural perception, the rather small difference of 5 years in women's vocalisations. And there are a plenty of acoustic parameters that show significant differences during this rather small time span.

These multiple differences in both chronologically and perceptually grouped samples suggest both an immense complexity of the (perceptive) construct speaker age as well as an, with respect to the affected parameter groups, wide-spread impact of (chronological) aging on speech production. This suggestion is further supported by including the results of other studies of speaker age: the assessment of appropriate parameters to indicate chronological age or to influence perceptive age and moreover the assessment of their relative relevance seems to be strongly depending on a multitude of factors. It should further be noted that at the moment the better age-indicating parameters are those that can be appraised more reliable – this does not mean that they parameterise the more valid concepts with respect to age.

Since there is obviously most information on age in spontaneous speech, and since particularly this type is especially hard to handle, there is still plenty of work, as well in the field of the acoustics of aging. Above all, linguistically normalised measures must be developed. For this purpose the robust and reliably age-indicating duration/ tempo measures seem to provide the best first basis. Probably there is as well much, yet unrevealed age information in (more) prosodic cues (than pure time derivates). Thus it seems unrealistic to capture a speakers age reliably and in general manner in the near future.

This leads to the conclusion that future research in speech acoustics (of aging) should try to capture and combine much more factors that influence speech production and reliable parameters to depict its output in one study.

6. Acknowledgements

The reported study was made possible by the voluntary participation of the speakers and listeners, partly by grants of the German Research Foundation DFG for the project "Junge und Alte Stimmen" and by the support of Prof. Walter Sendlmeier. I thank the reviewers and Ralf Winkler for their help to improve this article.

7. References

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

[2] Ptacek, P. and Sander, E., "Age recognition from voice", Journal of Speech and Hearing Research, 9:273-277, 1966.

[3] Brückl, M. and Sendlmeier, W., "Aging Female Voices: an Acoustic and Perceptive Analysis", Proc. of the ISCA Workshop Voqual'03, Geneva, 163-168, 2003.

[4] Schötz, S., Perception, analysis and synthesis of speaker age, Media-Tryck, Lund, 2006.

[5] Winkler, R., "Influences of pitch and speech rate on the perception of age from voice", Proc. of the 16th Int. Congress of Phonetic Sciences, Saarbrücken, 2007.

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

[7] Hoit, J., Hixon, K., Altman, M. and Morgan, W., "Speech breathing in women", Journal of Speech and Hearing research, 32:353-365, 1989.

[8] Endres, G., Bambach, A. and Flösser, M., "Voice Spectrograms as a Function of Age, Voice Disguise and Voice Imitation", J. Acoust. Soc. Amer., 49:1842-1847, 1971.

[9] Boersma, P. and Weenink, D., "Praat: doing phonetics by computer (Version 4.5.02)", [Computer Program], URL: http://www.praat.org, 2006.