INFLUENCES OF PITCH AND SPEECH RATE ON THE PERCEPTION OF AGE FROM VOICE

Ralf Winkler

Institute of Language and Communication, Technical University Berlin, Germany
ralf.winkler@tu-berlin.de

ABSTRACT

Listeners are able to rate a speaker's age with reasonable accuracy. Although several speech features are known to be characteristic for specific age groups, there is less knowledge about the perceptual relevance of those parameters. This paper describes the results of a perception study, where single word stimuli were synthesized and rated regarding the perceived age by 20 listeners. All combinations of pitch and speech rate were synthesized with male and female voices. Results show that (i) speech rate had the largest impact on listeners' judgement. Although pitch variations alone did not show a large impact on listeners' judgements, (ii) significant differences between selected pitch levels at slow and fast speech exist. Our results regarding (i) and (ii) contribute to the identification of the relevant features signaling a speaker's age. Results regarding (ii) further support the assumption that a set of parameters always interact in signaling a speaker's age.

Keywords: voice, perceived age, speech perception

1. INTRODUCTION

Previous research has shown that listeners can estimate a talker's age quite accurately based on listening to speech sounds alone ([8],[12]). Several features such as F_0, jitter, shimmer and spectral tilt as well as temporal features like segment durations and pauses have been identified as markers of chronological and perceived age (for details see [6]).

Amongst others, two features, voice pitch and speech rate, consistently appear in the literature to change with chronological age. But relationship between pitch and speech rate on the one hand and perceived age on the other are still controversial. While sometimes pitch does not influence age perception ([2]) sometimes pitch does have an influence ([4], [5], [7]). While speech rate sometimes influences the judgement ([2], [8], [13]), other studies do not find such a strong influence ([9]).

Synthesis experiments provide an insight into the contribution of selected features to the perception of a speaker's age. Shrivastav et al. [14] measured features related to F_0 and speech rate and resynthesized the young and old male voices by systematically manipulating pitch and speech rate to shift the perceived age of the groups towards each other. A significant shift was observed for the older, but not younger, voices. They successfully demonstrated that pitch and speech rate of older male voices can be manipulated to be perceived as significantly younger. But, because manipulated versions of natural stimuli were used, other possible features like roughness may interact with pitch and speech rate and might contribute to the perception of a speaker's age in an unpredictable way.

An approach to synthesize a speech sample with a specific perceived age was described in Schötz [10]. All parameters used for the formant synthesis of the target word were calculated by linearly interpolating between the parameters of the two speakers with the chronological age next to the target age. A total of four speakers differing in age were used to represent the voices of the desired age continuum ranging from 10 to 80 years. She successfully demonstrated that synthesizing voices differing in the mean perceived age is possible by means of formant synthesis. However, with her approach it would be impossible to determine the perceptual relevance of single parameters such as segment duration or pitch.

The aim of this study was to analyze the perceptual relevance of pitch and speech rate for listeners' judgement regarding a speakers' age. We used formant synthesis to ensure that no other cues vary in our stimulus set. Rather than interpolate between single speakers, in our approach the range of the variations has been defined based on measurements of a real speech database of young and old speakers.

2. METHODS & MATERIAL

2.1. Speech database

A database of 23 single words spoken by 30 female and 30 male subjects was recorded with speakers young and old one half each. The mean age of the female speakers was 26.27 (young) and 69.56 (elderly) years. For the group of male speakers the mean age was 25.5 (young) and 66.75 (elderly) years. Several acoustic and temporal features were measured in order to restrict the pitch and speech rate variations to a realistic range. Other features (e.g. group mean formant values) were used in the specification of the high-level synthesis parameters to avoid a possible judgement bias that could be regarded to e.g. low static formant values.

2.2. Synthesis

For synthesis the commercially available synthesizer HLsyn (Sensimetrics) was used. HLsyn is a high-level formant synthesizer that is based on a hybrid articulatory-acoustic model of speech production [3]. A Matlab environment was developed to calculate the HL-parameter trajectories. Every phone has been defined in terms of articulatory events, that influence at least one of the 13 variable parameters. In an initial step the single articulatory events were concatenated. The next step was to apply rules for adapting the formant transitions to the corresponding consonantal environment. Finally F_0 and subglottal pressure were manipulated to generate the appropriate prosody.

Values for the articulatory events (e.g. formant targets) were taken from the natural stimuli (see sec. 2.1.) if possible. Speed values for the articulators were adopted from [11]. To produce stimuli with a female voice the pitch values were adapted based on the natural speech database. We further multiplied all male formant targets by a factor of 1.2 to account for different vocal tract lengths of men and women. Three German words were synthesized: /libaneːzə/ (Lebanese), /laviːnə/ (avalanche) and /masiːf/ (solid). The three words were selected because of their different length (2 to 4 syllables) and the different consonants they are comprised of.

2.3. Parameter manipulation

The stimulus set was produced by varying systematically pitch, speech rate, lengthening and an introduction of a glottal chink while keeping all other parameters constant. We will here focus on the results for pitch and speech rate variations.

Pitch and speech rate dimensions were sampled at three points each. The precise values used with the synthesizer are given in Table 1.

2.4. Perception test

Perception tests were done using the software Praat [1]. The listener's task was to listen to a single word and immediately rate the age of the simulated speaker. While there was the ability to repeat listening, the number of repetitions was limited to twice. Listeners were asked to rate by mouse-clicking on one of 16 boxes labeled with age value of a time span of 5 years, beginning with 15 years and ending with an age value of 90 years. All boxes were visually arranged on a horizontal line without gaps to emulate an age continuum. Stimuli were presented in random order, starting with all male voices first and then going on with the female voices after a short break. All participants used earphones.

Table 1: Range of feature space spanned by speech rate (SR) [Phonem/s] and pitch (first value in Hertz [Hz]/ second value in semitones [sm]).

	low/ slow	middle	high/ fast
		female voice	
F_0	170.0 (5.0)	214.3 (9.0)	270.0 (13.0)
SR	4.5	7.0	9.5
		male voice	
F_0	100.0 (-4.0)	126.1 (0.0)	159.0 (4.0)
SR	4.5	7.0	9.5

2.5. Listeners

A total of 20, ten female and ten male listeners participated in our perception experiment. The mean age (and SD) of the group of female listeners was 26.7 years (2.54). Mean age of the male listeners was 30.6 years (6.47). All participants declared to have normal hearing ability.

2.6. Statistical analysis

In order to investigate the influence of the single factors a two way ANOVA (pitch×speech rate) with the listeners as repetitions was conducted on the full data set. To explore the factors in more detail other repeated measure ANOVAs were performed on subsets of the data. Detailed analysis of the relationship between pitch and listeners' judgements was done by conducting further ANOVAs on each level of speech rate level separately. The key point in using repeated measure ANOVAs was to analyze the judgement deviations attributed to the manipulated speech parameters while controlling for every listener's characteristic judgement bias.

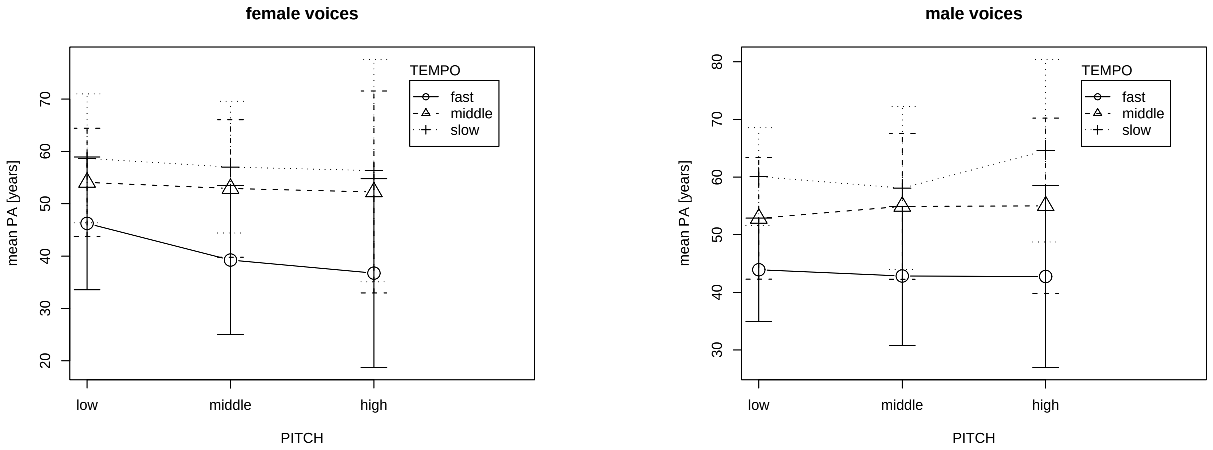
To account for possible differences in judging female and male voices judgements were analyzed for female and male voices separately. For statistical analysis the judgements of the three single words for every combination of pitch and speech rate condition were aggregated. Statistical analyses were done using R [15].

3. RESULTS

3.1. Main effects: Speech rate & Pitch

Mean perceived age values broken according to pitch and speech rate for female and male voices are

Figure 1: Means (± 1 SD) of the age judgements for the three levels of pitch. Lines represent the levels of speech rate. Left figure shows judgements for female, right picture for male voices.



(a) female

(b) male

given in Table 2. These numbers are visualized in Fig. 1 for female (left) and male (right) voices. Mean judgements of female (50.4 years) and male voices (52.8 years) are almost equal.

Table 2: Mean perceived age for the judgements broken according to pitch and speech rate for female (upper panel) and male voices (lower panel).

	$Fast$	$Middle$	$Slow$	$\bar{x}_{row}$
Low	46.3	54.1	58.7	53.0
$Middle$	39.3	52.9	57.0	49.7
$High$	36.8	52.3	56.3	48.4
$Mean$	40.8	53.1	57.3	50.4
Low	43.9	52.8	60.1	52.3
$Middle$	42.8	54.9	58.1	51.9
$High$	42.8	55.0	64.6	54.1
$Mean$	43.2	54.3	60.9	52.8
$\bar{x}_{column}$	42.0	53.7	59.1	51.6

Regarding the variations of speech rate a comparison of the columns of Table 2 shows a strong impact of the speech rate variations on listeners' judgements. Mean age judgements increased with decreasing speech rate for both, male and female synthesized voices. For female voices the mean judgements rise by 16.6 years from fast to slow speech rate. A rise by 17.8 years has been found for male voices.

In order to assess the impact of pitch variations on listeners' judgements we compared the row means. Regardless of the gender of the synthesized voices mean listeners' judgements do not show a strong impact. Mean listeners' judgements fall by an amount of 4.6 years for female and show a rise by 1.8 years for male voices between low and high pitch.

Statistical analysis revealed a significant main effect for speech rate in case of female [F(2,38)= 27.63, p=0.00] as well as male [F(2,38)=40.28 ,p=0.00] synthesized voices, but not for pitch. There was no significant interaction between pitch and speech rate.

3.2. Differences between levels of speech rate

The mean perceived age consistently increase with decreasing speech rate (cp. separate lines in Fig. 1). The pairwise analysis between two levels of speech rate always revealed significant differences. For female voices the difference between slow and middle speech rate [F(1,19)=5.16, p=0.03] is significant. The difference between middle and fast [F(1,19)=39.42, p=0.00] as well as slow and fast speech rate [F(1,19)=31.40, p=0.00] is even more highly significant. For the male voices separate ANOVAs revealed statistically significant differences in the comparison of slow vs. normal [F(1,19)=19.74, p=0.00], normal vs. fast [F(1,19)=26.07, p=0.00] and slow vs. fast speech rate [F(1,19)=62.88, p=0.00]. For female as well as for male voices no interaction reached statistical significance.

3.3. Differences between levels of pitch

The effect of different pitch levels on listeners' judgement differs between different levels of speech rate.These dependencies are depicted in Fig. 1.

Each figure shows two lines corresponding to two levels of speech rate without strong differences between pitch levels. While for the female voices judgements regarding different pitch levels do not differ for slow and normal speech rate, a characteristic pattern can be observed for the fast speech rate (see Fig. 1 left). Results for mean listeners' judgement show a rise by 9.5 years from high to low pitch level. For the male voices, listeners' judgements seem to be less influenced by the pitch level in fast and middle, but more in slow speech rate. If stimulus words were spoken slow (see dotted line in Fig. 1 right), the high pitch level was associated with a remarkable increase in the mean listeners' judgement of a talker's age.

Consequently selective tests regarding statistically significant differences between two pitch levels were done for female voices in the fast and for male voices in the slow speech rate condition. We found a significant difference between low and middle [F(1,19)=6.19, p=0.02] and between low and high pitch [F(1,19)=6.27,p=0.02] for the female synthesized voices. For the male voices subset analysis revealed a significant difference between middle and high pitch [F(1,19)=9.87, p=0.01].

4. DISCUSSION & CONCLUSION

Our results are in line with results from Schötz [10] in a sense, that formant synthesis is applicable and capable of producing speech with an intended perceived age. Results are in line with Shrivastav et.al. [14] as well, where speech rate and pitch were manipulated to shift the perception of a speaker's age. For the first time in our work synthetic words varying in nothing but speech rate and pitch have been judged by listeners regarding the perceived age. Our results show, that variations of speech rate and pitch exclusively already produce a stimulus set with a reasonably broad continuum of perceived age. With our approach it is possible to estimate the contribution of single features and combinations of them to the perception of age.

However, listeners were forced to use exactly those cues to rate the talker's age. Other cues (e.g. roughness), if present in the stimulus, probably contribute to the perception of age from voice as well. Furthermore, up to now no judgements of the naturalness of our stimuli has been collected. Hence, our cues could be weighted more heavily than in natural speech due to the reduced naturalness of the synthesized stimuli.

In our experiment listeners' judgement on a speaker's age has mainly been influenced by the speech rate of the stimulus words. Dependent on the gender of the synthesized voice and the speech rate, pitch also contributed to a characteristic perception of a speaker's age.

ACKNOWLEDGMENTS

Parts of this work were supported by the Deutsche Forschungsgemeinschaft (German Research Fund, SE 462/5-1 and SE 462/5-2).The author would like to thank the reviewers for many helpful comments and suggestions on an earlier version of this paper.

5. REFERENCES

[1] P. Boersma and D. Weenink. Praat: doing phonetics by computer (version 4.5.01). [Computer program], 2006.

[2] A. Braun and T. Rietveld. The influence of smoking habits on perceived age. In *Proceedings of The XIIIth International Congress of Phonetic Sciences*, volume 2, pages 294–297, Stockholm, Sweden, 1995.

[3] H. M. Hanson and K. N. Stevens. A quasiarticulatory approach to controlling acoustic source parameters in a Klatt-type formant synthesizer using HLsyn. *Journal of the Acoustical Society of America*, 112(3):1158–1182, 2002.

[4] D. E. Hartmann and J. L. Danhauer. Perceptual features of speech for males in four perceived age decades. *Journal of the Acoustical Society of America*, 59:713–715, 1976.

[5] Y. Horii and W. Ryan. Fundamental frequency characteristics and perceived age of adult male speakers. *Folia Phoniatrica*, 33:227–233, 1981.

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

[7] S. Linville and E. Korabic. Elderly listener's estimates of vocal age in adult females. *Journal of the Acoustical Society of America*, 80:692–694, 1986.

[8] P. Ptacek and E. Sander. Age recognition from voice. *Journal of Speech and Hearing Research*, 9:273–277, 1966.

[9] W. Ryan and K. Burk. Perceptual and acoustic correlates in the speech of males. *Journal of Communication Disorders*, 7:181–192, 1974.

[10] S. Schötz. Data-driven formant synthesis of speaker age. In *Proceedings of Fonetik*, Lund, 2006.

[11] Sensimetrics. HLsyn (version 2.2). [Turorial, part of Manual], 2002.

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

[13] T. Shipp, Y. Qi, R. Huntley, and H. Hollien. Acoustic and temporal correlates of perceived age. *Journal of Voice*, 6:211–216, 1992.

[14] R. Shrivastav, H. Hollien, W. Brown, H. B. Rothman, and J. D. Harnsberger. Shifting perceptions of age in voice. *Journal of the Acoustical Society of America*, 114(2):2336–2337, 2003.

[15] R. D. C. Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2006. ISBN 3-900051-07-0.