



# Listener estimation of speaker age based on whispered speech

Angelika Braun, Daniela Decker

Department of Phonetics, University of Trier, Germany

brauna@uni-trier.de; daniela.decker87@googlemail.com

## Abstract

The assessment of a given speaker's age has been shown to rest on laryngeal as well as vocal tract features. This study attempts to separate these two elements by comparing listener performance based on whispered and phonated speech. A total of 45 speakers belonging to three different age groups (20-35; 45-60; 70 and above) were assessed by 20 listeners. Results show that although listener performance decreases with whispered speech, it is still well above chance level. No significant difference was found between age groups, but female speakers were assessed correctly more often than males in both conditions. There are forensic implications to this finding, since in that context, age estimation sometimes has to be carried out on whispered speech, e.g. if whispering is used as a voice disguise.

**Index Terms:** age estimation, whisper, forensics

## 1. Introduction

Estimating a given speaker's age based on his or her speech is a task which forms part of the everyday abilities within a speech community. The reason why it seems reasonable to assume that it can be done is, of course, that the bodily structures used for speaking undergo changes with age: this concerns the degree of neuromuscular control, the respiratory process, phonation, vocal tract features, and last, but not least, hearing [9, 7]. It is only fair to assume that, whether consciously or not, these changes will not go undetected by listeners. What renders age estimation somewhat fuzzy is the fact that aging is neither a uniform nor a linear process. Speakers in relatively good health have been judged to be significantly younger than speakers whose physiological condition was poor [13]. Smokers were generally judged to be significantly older than non-smokers of the same calendar age [3]. Since the gap between biological age and calendar age tends to widen with an increase in the latter, older subjects can be expected to be more difficult to assess.

The accuracy of age estimation by listeners has been studied extensively. Accuracy was defined as either the ability to correctly place a speaker in an age group (cf. e.g. [12]) or to make an estimate of speaker age to the exact year (cf. e.g. [1, 15, 14]). Researchers agree that listeners are able to place speakers in their respective age group with a high degree of accuracy [9]. In the latter case, a significant correlation between calendar age and perceived age was considered to be a good indicator of listener accuracy (e.g. [15]). On second thought, however, this may not be such a good idea, because an offset of, say, 15 years for all speakers studied will result in a perfect correlation between calendar age and perceived age but very poor accuracy. Studies which aimed at a direct age estimation determined an average difference of about 8 years between perceived age and calendar age [1].

Even though results vary and several variables have been established [5], (e.g., speaker age, listener age, an interaction between the two) there can be no doubt that listener performance is well above chance. Among the factors which have been found to best signal speaker age and are utilized by listeners in their assessment are speaking tempo, e.g. [3, 4, 14, 16], breathing [16], and voice source characteristics like fundamental frequency and periodicity, cf. e.g. [4].

The bulk of previous research on age estimation has been devoted to phonated speech. This is in keeping with the fact that phonation seems to be a primary key to age assessment. In the forensic context, however, whispered samples may occur as a voice disguise or owing to circumstantial factors. That is why from a forensic perspective, it is of interest to establish just how exactly speaker age can be determined based on whispered speech.

The number of studies devoted to age estimation based on whispered speech, thus having an immediate bearing on the research presented here, is very small. Two experiments are reported by Linville and colleagues [10, 11]. Linville and Fisher (1985) studied the ability of young female listeners (N=23; aged 20-28 years) to determine the age group of 75 female speakers who were equally divided among three age groups: 25-35, 45-55, and 70-80 years. The speech material consisted of one-second chunks of the vowel /æ/ both in phonated and whispered condition. They found correct recognition rates of 51% for phonated speech and 43% for whispered speech, both being well above chance and the difference between the two significant. Linville and Korabic (1986) tested the same materials, this time with 23 elderly females (aged 65-90 years) as listeners who reported no significant health problems. These listeners correctly identified the age group of the speakers 45% of the time from phonated vowels and 38% of the time from whispered vowels. Both these rates are well above chance, but they were significantly worse than the recognition rates for young listeners.

Jacques and Rastatter (1990) recorded two groups of young adults (10 male, 10 female) aged 21 years on average and two groups of elderly speakers (10 male, 10 female), aged 74 years on average. (The age ranges are not provided, which makes it impossible to determine the age gap between the two groups.) The materials consisted of 10 sustained vowels in phonated and whispered condition. These materials were played to a total of 32 listeners (16 young and 16 old, both groups balanced for sex). Recognition rates were found to drop significantly from phonated to whispered speech in both young (75% vs. 58%) and old listeners (53% vs. 40%). No significant difference was established with respect to speaker sex.

In a pilot study carried out by the first author [2], a total of 45 speakers were recorded. They fell into three distinct age groups: young (20-35 years), middle-aged (45-60 years) and old (70 years+). Each group was balanced for sex. The

speakers were asked to produce about one minute of semi-spontaneous speech based on a drawing depicting a summer day in the countryside. Twenty-eight students served as listeners. They were presented with complete utterances of approximately 6 seconds' duration from the picture descriptions and were asked to place them into one of the three age groups. All listeners individually performed above chance level. Overall, 67% of the listener judgements were correct. Young speakers were more often assessed correctly than middle-aged and old ones: 79%, 65% and 57%, respectively. This is in keeping with the results of previous research which found that listeners are best at assessing the age of speakers who are about their own age [5]. In the young (87% vs. 71%) and old (69% vs. 42%) groups, female voices were assessed with a greater degree of correctness, whereas in the middle-aged group men were assessed correctly more often (60% vs. 71%).

In a forensic scenario, whispering will occur either as a means of deliberate voice disguise or circumstantially, i.e. if a victim or witness manages to call 911 with a perpetrator present in the immediate vicinity. In this context, the amount of speech which is available for examination will usually exceed one vowel. Therefore, it was decided to provide listeners with more material than the two studies cited above did.

The research questions addressed in this paper are as follows:

- To what extent are listeners able to estimate speaker age based on whispered utterances (as opposed to sustained vowels)
- To what extent does this estimation differ from one which is based on phonated speech?
- Do recognition rates vary with speaker age (group)?
- Do recognition rates vary with speaker sex?
- Do recognition rates vary with listener age?

## 2. Materials and methods

### 2.1. Speakers

A total of 45 speakers, who were not identical with those reported on in our previous study, took part in the present experiment [2]. They fell into three distinct age groups: 20 to 35 years, 45 to 60 years, and 70 years and above. Each age group was represented by 15 speakers. Special care was taken to ensure that the old speakers were in good physical condition, thus making the task potentially more difficult for the listeners since they could not rely on cues like heavy breathing or voice breaks which may be indicative of a bad state of health (and thus old age). We aimed to balance these groups for speaker sex. This worked well for the young and middle-aged speakers (7 males and 8 females each), but turned out to be more difficult for the old speakers owing to the health requirements (4 males and 11 females).

### 2.2. Materials

The speakers were asked to read a set of 13 sentences first in phonated mode and subsequently in whispered mode. We decided to use read speech because any (semi-) spontaneous speech involves the risk of cues to speaker age being provided by choice of words or grammatical form. The utterances were

selected from the comprehensive recordings based on fluency and acoustic quality. Thus, a test tape of 26 min. duration was assembled which contained 90 utterances.

### 2.3. Listeners

Twenty listeners (8 males and 12 females) between 19 and 65 years of age with a mean of 31 for the men and 38 for the women participated in the experiment. They reported no known hearing impairments. In the listening experiment, they were provided with a total of 90 utterances, 45 of which were phonated and 45 whispered. Their task was to place each utterance into an age group. Furthermore, listeners were asked to name criteria for their decision, with a standard set of criteria provided to them: speaking tempo, respiratory noises, articulatory precision, loudness, and voice quality.

## 3. Results

Figure 1 shows the overall results:

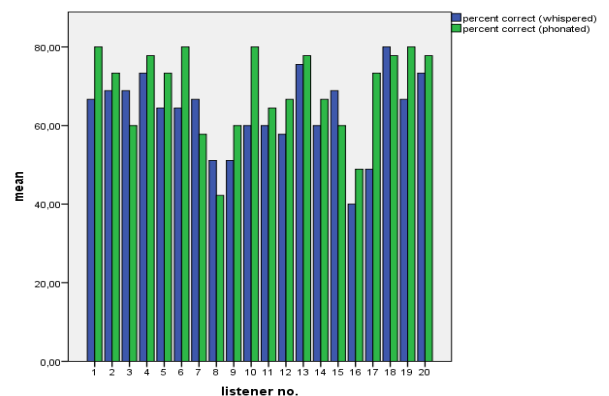


Figure 1: Percentage of correct assignments to the respective age group for whispered (left/blue column) and phonated (right/green column) utterances.

It is evident from the results that 15 out of 20 listeners made more correct assignments to speaker age group for phonated than for whispered speech. This underlines the importance of laryngeal information for making age estimations.

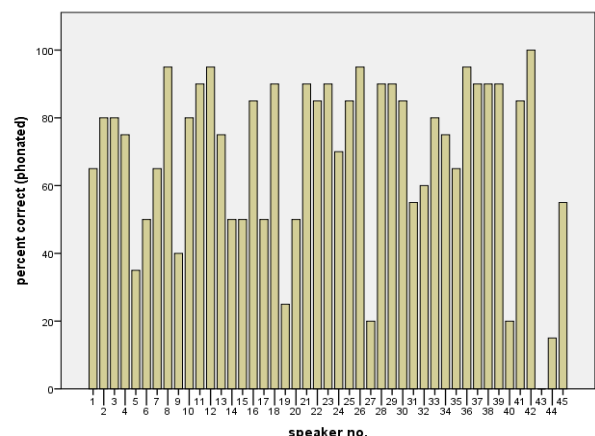


Figure 2: Percentage of assignments of phonated stimuli to the correct age group – individual speakers.

While there is considerable between-listener variation (compare listeners no. 8 and 18), all listeners performed well above chance.

Figure 2 shows the estimates for all speakers in terms of correct assignment to the respective age group. Listeners correctly assigned the speakers in 69% of the trials; however, only one speaker (no. 42) was always placed correctly. On the other hand, one speaker (no. 43) was never assessed correctly. This demonstrates that some speakers are more difficult to assess that others, which can probably be attributed to the discrepancy between calendar age and biological age.

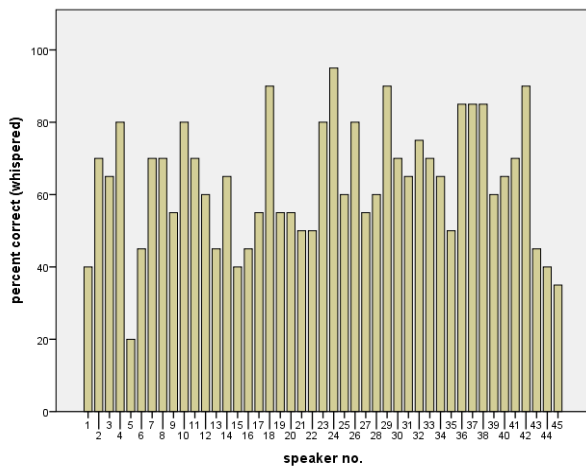


Figure 3: Percentage of assignments of whispered stimuli to the correct age group – individual speakers.

The percentage of correct age group assignments for the whispered stimuli was somewhat lower than for the phonated ones, but a paired t-test rendered no significant results. Values for individual speakers amount to an average of 63% correct judgments with a range of 20% to 95%. These rates are in close agreement with those of our previous study which also used continuous speech [2].

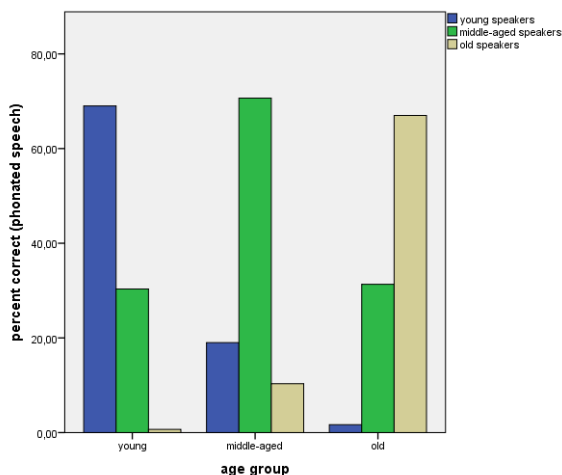


Figure 4: Age group assignments of young, middle-aged and old speakers (from left to right) based on phonated stimuli.

They are considerably higher than those reported in the literature which are based on isolated vowels rather than complete utterances [6, 10, 11]. As can be seen in Figure 4, group assignments were largely correct with only very rare occurrences of confusions between extreme age groups. An ANOVA rendered no significant differences in recognition rates between age groups ( $p = .92$ ).

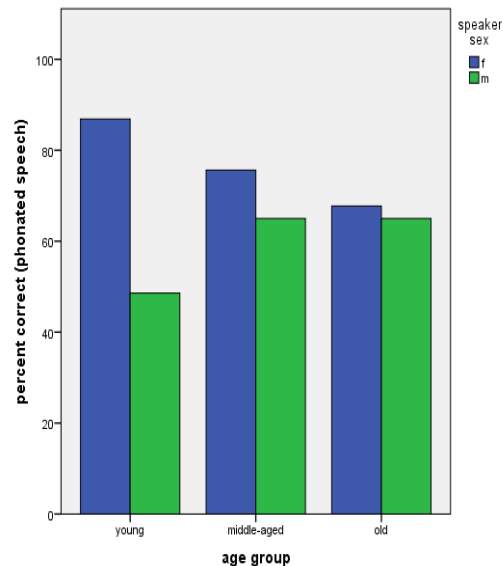


Figure 5: Percentage of correct age group assignments according to group and speaker sex (left/blue = female; right/green = male) for phonated stimuli.

Female speakers were placed more accurately than males in all three age groups. Two-tailed t-tests showed that only the difference for the young group was significant ( $p = .002$ ).

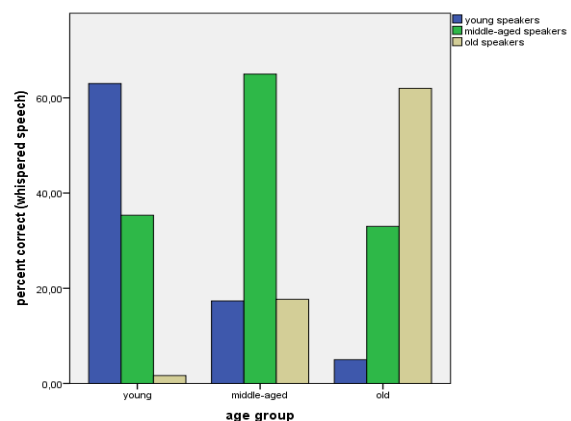


Figure 6: Age group assignments of young, middle-aged and elderly speakers (from left to right) based on whispered stimuli.

Although recognition rates for whispered speech are somewhat lower than those for phonated speech, the confusion patterns are very similar with a slight tendency for middle-aged speakers judged to be "old" more often.

An ANOVA rendered no statistical difference between age groups ( $p = .89$ ).

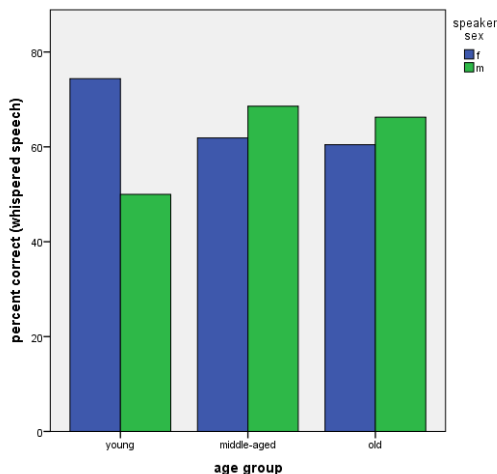


Figure 7: Percentage of correct age group assignments according to group and speaker sex (left/blue = female; right/green = male) for whispered stimuli.

As in phonated speech (cf. Fig. 5), the young female speakers were assigned correctly more often than the young males in whispered speech. In the two older groups, however, the men's age group was recognized slightly more often. Again, the difference is significant only for the young group (two-tailed t-test,  $p = .002$ ). - Finally, listener age as an influencing factor needs to be considered. The following graphs show the results for the Pearson correlation calculations:

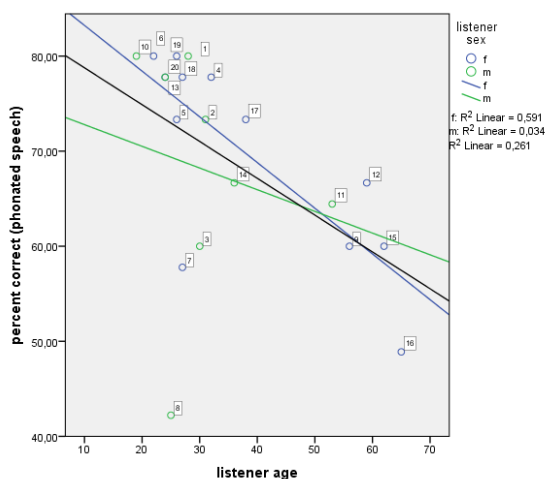


Figure 8: Correlation between listener age and the percentage of correct judgments for phonated stimuli.

Both correlations are negative, i.e. listener performance decreases with age for phonated and whispered speech alike. A calculation of the Pearson coefficient rendered  $-0.511$  for phonated speech and  $-0.515$  for whispered stimuli. Both are significant ( $p = .021$  and  $p = .020$ , respectively).

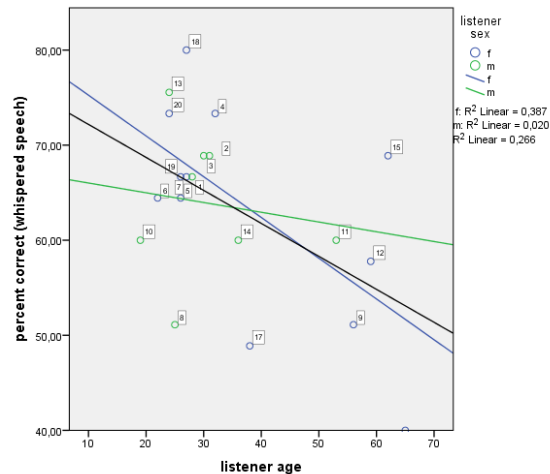


Figure 9: Correlation between listener age and the percentage of correct judgments for whispered stimuli.

#### 4. Discussion and Conclusions

With respect to the research questions asked at the outset, it can be concluded that speaker age group is much more reliably determined based on complete utterances than based on vowels only. This can be taken to imply that factors such as speech rate and breathing patterns which go well beyond resonance characteristics are utilized by listeners. We tested the utterances used in the experiment for articulation rate (ANOVA, Scheffé post-hoc testing) and found highly significant differences between all three age groups for phonated speech ( $p = .001$ ) and between the young and middle-aged groups and the young and old groups for whispered speech ( $p = .000$  for both). This means that speaking tempo has once again been established as an important factor which contributes to age estimation. This factor was also demonstrated to bear perceptual relevance by the fact that listeners placed it high on their list of criteria used. Given these circumstances, age group assignment was not found to be statistically superior for phonated speech as compared to whispered speech.

At first glance, it is somewhat surprising that no significant differences were found with respect to speaker age (group). But in contrast to earlier studies which used college-age listeners, most of our listeners are within the age range of the middle-aged group, thus eliminating the young-speaker bias as reported by [5]. Still, listener age had a significant effect, thus confirming the results of earlier studies [5, 11]. - Effects of speaker sex were found for the young group only in both phonated and whispered conditions. This is clearly something which will merit further attention in future work.

The present results look encouraging from a forensic angle because there is no significant decline in listener performance in the assessment of whispered speech as compared to phonated speech as long as full utterances are available and the task is to identify an age group only. However, in the forensic setting an exact estimation is called for, and it remains to be established just how well this can be done compared to phonated speech.

## 5. References

- [1] Braun, A., "Age estimation by different listener groups", in *Forensic Linguistics* 3, 65-73, 1996.
- [2] Braun, A., Diehl, N., "Age estimation based on whispered speech", Abstracts IAFPA, Trier, 2010.
- [3] Braun, A., Rietveld, A.C.M., "The Influence of Smoking Habits on Perceived Age", in *Proceedings of the XIIIth International Congress of Phonetic Sciences*, Stockholm, vol. 2, 294-297, 1995.
- [4] Harnsberger, J.D., Shrivastav, R., Brown, W.S., Jr., Rothman, H., Hollien, H., "Speaking rate and fundamental frequency as speech cues to perceived age", in *Journal of Voice* 22, 58-69, 2008.
- [5] Huntley, R., Hollien, H., Shipp, Th., "Influences of listener characteristics on perceived age estimations", in *Journal of Voice* 1, 49-52, 1987.
- [6] Jacques, R. D., Rastatter, M. P., "Recognition of speaker age from selected acoustic features as perceived by normal young and older listeners", in *Folia Phoniatica* 42, 118-124, 1990.
- [7] Kahane, J.C., "Anatomic and physiologic changes in the aging peripheral speech mechanism", in D.S. Beasley / G.A. Davis [Eds], *Aging. Communication Processes and Disorders*, 21-45, New York etc.: Grune & Stratton, 1981.
- [8] Kent, R.D., Burkard, R., "Changes in the acoustic correlates of speech production", in D.S. Beasley / G.A. Davis [Eds], *Aging. Communication Processes and Disorders*, 47-62, New York etc.: Grune & Stratton, 1981.
- [9] Linville, S.E., "Vocal Aging". San Diego: Singular, 2001.
- [10] Linville, S.E., Fisher, H. B., "Acoustic characteristics of perceived versus actual speaker age in controlled phonation by adult females", in *The Journal of the Acoustical Society of America* 78, 40-48, 1985.
- [11] Linville, S.E., Korabic, E.W., "Elderly listeners' estimates of vocal age in adult females", in *The Journal of the Acoustical Society of America* 80, 692-694, 1986.
- [12] Ptacek, P.H., Sander, E.K., "Age recognition from voice", in *Journal of Speech & Hearing Research* 9, 273-277, 1966.
- [13] Ramig, L.A., "Aging speech: physiological and sociological aspects", in *Language & Communication* 6, 25-34, 1986.
- [14] Ryan, W.J., Burk, K.W., "Perceptual and acoustic correlates of aging in the speech of males", in *Journal of Communication Disorders* 4, 181-192, 1974.
- [15] Shipp, Th., Hollien, H. "Perception of the aging male voice", in *Journal of Speech & Hearing Research* 12, 703-710, 1969
- [16] Shipp, Th., Qi, Y., Huntley, R., Hollien, H. "Acoustic and temporal correlates of perceived age", in *Journal of Voice* 6, 211-216, 1992.