

THE SPEAKER'S AGE: A PERCEPTUAL STUDY

Massimo Pettorino & Antonella Giannini

University of Naples L'Orientale, Italy
mpettorino@unior.it; agiannini@unior.it

ABSTRACT

The results of this experimental research on Italian language show that it is possible to infer a speaker's age on a perceptual basis with good approximation. The procedure involves a series of comparative tests where the unknown voice is compared to voices belonging to different age ranges. Then, on the basis of the answers, listeners are asked to indicate the age of the unknown voice in a reduced interval. The procedure has been tested on two voices of different age. Furthermore the study shows that young listeners give the best results and neither different kinds of speech (spontaneous/read) nor the variety of regional Italian have any relevant influence on the identification of age. Such a procedure could be useful for several applications, as for instance the identification of the speaker's age in a forensic field and the protection of the children in web navigation.

Keywords: perception, age, voice, recognition

1. INTRODUCTION

Many experimental studies prove that a relationship exists between a speaker's age and his voice. The major changes that the voice shows in relation to age are:

- lowering of breathing functions [2]
- muscle relaxation, hardening of vocal folds [9]
- progressive tonal lowering [7]
- lowering of speech rate [1, 5]
- increase of jitter and shimmer [9, 11]
- lowering of formant frequencies [8]
- longer vowels and stop consonants [10]
- increased standard deviation of f_0 [6, 8, 14]

Furthermore, in a study conducted on Italian language [3], a well-known TV reporter read the same text he had read 40 years before during a TV news segment, and the results of the spectroacoustic analysis showed that, both on the segmental and suprasegmental levels, the "old" speech was more isochronic, the f_0 was about 30% higher, with a larger tonal range, with longer and

more frequent silent pauses. All these changes explain why, on the perceptual level, a voice is sufficient to allow the listener to infer the age of the speaker. The accuracy of the judgment depends on many factors, as for instance the type of speech (read/spontaneous), the speech rate, the listener's age, the speaker's dialect and so on.

In Ptacek e Sanders [6], ten listeners had to recognize whether the speakers were younger than 35 or older than 65. The correct answers went from 78% (in the case of prolonged vowels) to 99% (for a short reading). In 1969, Shipp and Hollien [13] conducted a similar experiment with 175 male speakers of different ages (20-89). Three listener groups had to: 1) define each voice as "young/old/neither young nor old"; 2) identify the age range; 3) guess the age of the speaker. Each task had different results: very high percentage of correct answers in the first case; less in the second task; even less in the third, with a tendency to overestimate young voices (20-40) and to underestimate old ones (70-90). In 1976, Hartman e Danhauer [4] conducted a perceptual test on male and female voices of different ages. Listeners were able to identify the age ranges on the basis of different voice qualities (clear/hoarse voice, slow/fast articulation rate, high/low tonal range, frequency of silences). Finally, Schöz [12] finds that on the perceptual level the recognition of the speaker's age depends also on non-phonetic factors:

- speaker-related factors: physical and physiological state (size, weight, health), language, sociolect, dialect and native accent psychological state, experience/wisdom society, culture typical/atypical of chronological age;
- speech-sample related factors: speech type (vowel, spontaneous, read, etc.), speech sample length (duration), technical sound quality (recording, medium, playback);
- task-related factors: instructions and motivation, precision demanded, time allowed and number of repetitions, number of stimuli, tendency to avoid extreme age estimations (centring).

In conclusion, all of these studies show that there are too many factors involved as people's voices change as they age, and therefore, it is extremely difficult to derive the age of a speaker on the basis of an acoustic/perceptual analysis of speech signals. The aim of this research, which is part of a wider project seeking to develop a standard procedure of speaker's age identification, is to verify whether, and with which degree of approximation, it is possible to identify a speaker's age exclusively on a perceptual basis.

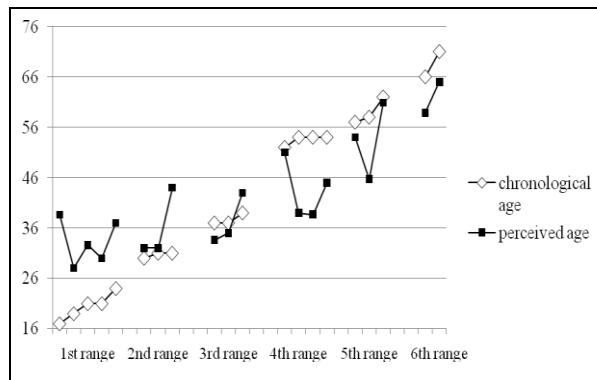
2. FIRST PHASE

In the first phase of our research we tried to identify the most suited typology of listeners in order to recognize a speaker's age. The corpus consisted of a sentence taken from a newspaper, read by 23 Italian male speakers belonging to six different age ranges, with the first being from 16 to 25, and the sixth consisting of speakers from 66 to 75. The same corpus has been used also in the next phases of the research.

Recordings were made in a silent room. Firstly, we made a tape with a sequence of 10 pairs of voices (one was repeated twice, as a control) with an interval of 5 seconds between one pair and the other. The two elements of each pair differed with at least 1 age range. On a second tape, we made a sequence of 20 voices (6 from the first tape and 14 new voices) in random order. The two tapes were the object of two perception tests administered to 177 listeners of different ages, 90 female and 87 male. For the first test, the listeners had to indicate, on a given form, which one of the two voices of each pair was younger. The results showed that the listeners of the first two ranges (aged 16-35) gave the best results (94% and 91% correct answers). The worst results came from listeners from the highest range (aged 66-75). There were no differences between male and female listeners.

The second tape was then administered in a second test. After each voice the listeners had to write, on a form, the speaker's age. The results confirmed that young voices are overestimated and old voices underestimated (see also [13]). The best results were obtained for voices of the 2nd and 3rd ranges (aged 26-45). It is relevant to note that, in any case, chronological and perceived ages differ for not more than one range (fig. 1). Nevertheless, we have to say that in this second test the correct identification of age range was generally under 60% of the total answers given.

Figure 1: Chronological vs. perceived age of the speaker.



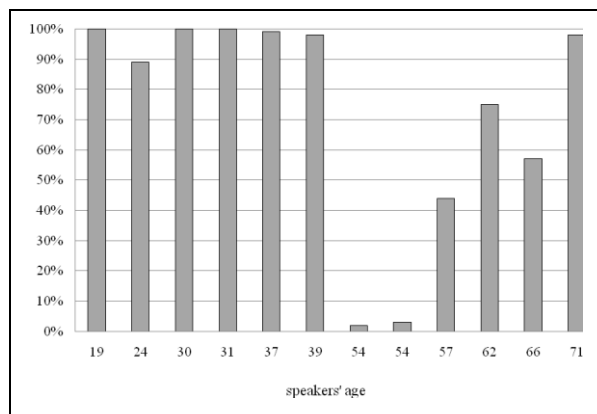
3. SECOND PHASE

The results of the first phase showed that the best listeners are those from 16-25 years, both male and female. Furthermore they proved that the comparative test is quite reliable, but also that it is very difficult to estimate a speaker's age in a totally free-choice situation.

At this point we decided to make a listening tape in which X-voice (in our case, 52 years old) was compared with 12 voices (2 for each range). The aim of this step was to try to reduce the answer range of the second test.

The perception test was then administered to 93 Italian university students (aged 19-25). They had to indicate, for each pair of voices, which one was younger. The results are shown in figure 2.

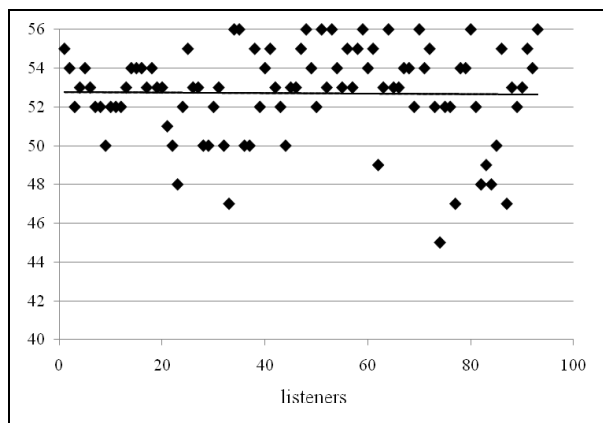
Figure 2: Percentage values of correct answers from the comparative test.



The X-voice is clearly perceived older than the voices of the first three ranges (19-39 years), with 97% of the answers being correct. Then, after a surprising drastic fall for voices belonging to the same range of the X-voice, the correct recognition starts going up for the subsequent ranges, reaching 98% for the oldest range. In the second part of the

test we asked listeners to indicate the speaker's age in a fixed interval. At the beginning, we thought of using a 20-year range interval, from 40 to 60, but our X-voice would then occur in the middle of the range, and we wanted to avoid this possibility. Thus, we decided to ask the listeners to indicate an age between 40 and 56. Figure 3 shows the answers of the 93 listeners. As we can see, the mean perceived age corresponds to 52.60. It is interesting to note that no answers occur below 45. This means that the lower threshold could be fixed at least 5 years after, with a 97% rate of correct recognition (see figure 2). As regards the higher threshold, we noticed that 10 answers (10% of the total) were on the 56-year-line and some would still have possibly been placed at a higher level. So, as aforementioned, the range could reach a higher value (for instance, 61).

Figure 3: Perceived age of the speaker.

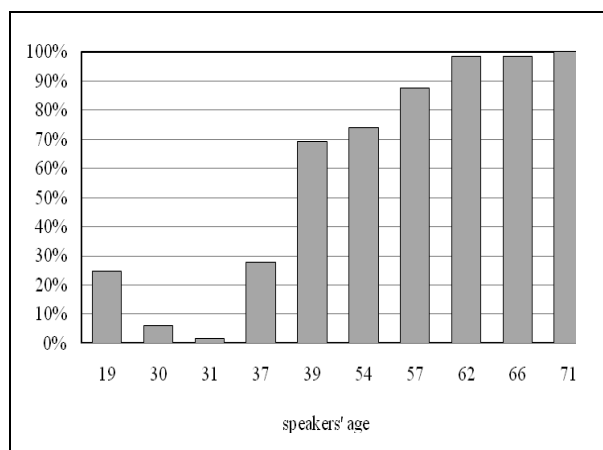


4. THIRD PHASE

The aim of the third phase of our research is to verify whether variables related to the speech sample and to the regional area of the speakers affect, and at which degree, the reliability of our procedure.

Following the same procedure from the 2nd phase of experiments, we made a listening tape in which a short sample of spontaneous speech (8 seconds), uttered by a 38 year-old speaker (X-voice), of a regional area different from the listener's area, is compared to read speech uttered by 10 speakers of different ages (from 19 to 71). The 10 pairs of voices were administered to 66 listeners, all university students (aged 18-26). In the first session, the listeners were asked to recognize the younger voice of each pair. The percentage values of recognition of the X-voice as younger, for each pair, are shown in figure 4.

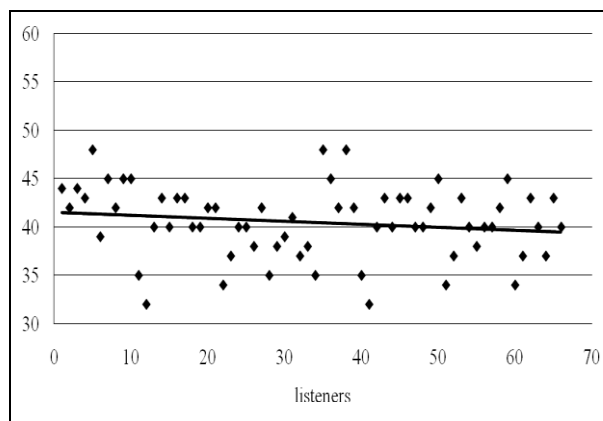
Figure 4: Percentage values of perceived age.



As we can see, almost 100% of the listeners recognized the X-voice as younger than 62 years and almost 90% younger than 57. This allows us to fix the upper threshold at 56 years. On the contrary, the values related to the 30 and 31 year-old voices (6% and 2%) testify that they are perceived as younger than the X-voice. On the basis of these considerations, we decided to fix the age range of the X-voice between 32 and 56 years.

Then, we went on with the second test. We administered the X-voice to the same listeners and asked them to write an age within that interval. The results of the test are shown in figure 5. As we can see, the perceived range is narrower than the given one: only one case corresponded to the lower threshold (32 years) and not one was higher than 48 years. The mean value is 40.5. This seems to be a good result, considering that the X-voice was 38 years old.

Figure 5: Perceived age of the speaker.



5. CONCLUSION

Although further research on a larger sample is needed, the results of the perceptual tests show that

it is possible to infer, on a perceptual basis and with good approximation, the speaker's age. Furthermore, neither the different kind of speech (spontaneous/read) nor the different variety of regional Italian have any relevant influence on the identification of age.

The final goal of our research is to develop a semi-automatic procedure so that, given an unknown voice, it is possible to identify the age of the speaker following these steps:

1. comparative test with n -couples of voices (i.e. X-voice / voice of n -range);
2. to administer the test to young listeners (19-25);
3. on the basis of the results, to identify the minimum and maximum threshold of the X-voice;
4. to perform the absolute test on the identified range.

Such a procedure could be useful for several applications, as for instance the identification of the speaker's age in a forensic field, the protection of the children in web navigation, or in all those cases where the subject is not able to give his personal data (medical, social, immigrants welfare fields).

6. REFERENCES

- [1] Amerman, J.D., Parnell, M.M. 1992. Speech timing strategies in elderly adults. *Journal of Phonetics* 20, 65-76.
- [2] Awan, S.N. 2006. The aging female voice: acoustic and respiratory data. *Clinical Linguistics & Phonetics* 20(2-3), 171-180.
- [3] Giannini, A., Pettorino, M. 2009. L'età della voce. *La Fonetica Sperimentale: Metodo e Applicazioni, Atti del IV Convegno Nazionale AISV* 4, 165-178.
- [4] Hartman, D.E., Danhauer, J.L. 1976. Perceptual features of speech for males in four perceived age decades. *Journal of the Acoustical Society of America* 59(3), 713-715.
- [5] Hollien, H., Shipp, T. 1972. Speaking fundamental frequency and chronological age in males. *Journal of Speech and Hearing Research* 15, 155-159.
- [6] Jacques, R., Rastatter, M. 1990. Recognition of speaker age from selected acoustic features as perceived by normal young and older listeners. *Folia Phoniatica* 42, 118-124.
- [7] Lindblad, P. 1992. *Rösten*. Lund: Studentlitteratur.
- [8] Linville, S.E. 1987. Acoustic-perceptual studies of aging voice in women. *Journal of Voice* 1, 44-48.
- [9] Linville, S.E. 2004. The aging voice. *The American Speech-Language-Hearing Association (ASHA) Leader* 12-21.
- [10] Ptacek, P.H., Sander, E.K. 1966. Age recognition from voice. *Journal of Speech and Hearing Research* 9, 273-277.
- [11] Ramig, L.A., Ringel, R.L. 1983. Effects of physiological aging on selected acoustic characteristics of voice. *Journal of Speech and Hearing Research* 26, 22-30.
- [12] Schütz, S. 2006. Perception, analysis and synthesis of speaker age. *Travaux de l'Institut de Lund* 47, 1-186.
- [13] Shipp, T., Hollien, H. 1969. Perception of the aging male voice. *Journal of Speech and Hearing Research* 12, 703-710.
- [14] Traunmüller, H., van Bezooijen, R. 1994. The auditory perception of children's age and sex. *ICSLP* 1171-1174.

¹ The sentence was "Il Presidente della Repubblica ha ricevuto il Ministro degli Esteri per esaminare lo stato della situazione internazionale" (The President of the Republic has received the Minister of Foreign Affairs to examine the state's international situation).