SPEAKER'S IDENTIFICATION BY VOICE

Antonella Giannini, Massimo Pettorino, Umberto Cinque


Since the early 70's many experimental studies have pointed out the existence of a close relationship between the acoustic characteristics of voice and the physical features of the speaker. The purpose of this study is to determine, on one hand, whether listeners are capable of identifying age, weight and height of speakers; on the other hand, to verify on the basis of perceptive tests, whether it is possible to distinguish twin's voices with better than chance-guessing accuracy. The results give rise to various considerations. Suggestions for future research are discussed.

In this study we will report the first results obtained in our research on speakers' identification by voice. Our interest will now be limited to two aspects of this topic: 1) to verify whether listeners are capable of identifying age, weight and height of speakers and 2) to verify whether, from an acoustic-perceptive point of view, it is possible for listeners to distinguish twins' voices with better than chance-guessing accuracy.

Since the early 70's many experimental studies (1) (2) (3) have pointed out the existence of a close relationship between the acoustic characteristics of voice and the physical features of the speaker. According to these studies listeners can identify many characteristics of speakers such as for instance age, sex, race, weight and height on the basis of perceptual cues obtained from recorded speech samples.

However, it must be said that, as regards height and weight, the results obtained by C. D. Gunter and Manning (4) are definitely against this hypothesis in that: "when mean values such as these are utilized, most of the discrepancy between actual and estimated measurements will be absorbed by the regression towards the mean" (p.252).

Particularly significant in this way is the test made by Cohen et al. (5) where "a panel of 24 "listeners" was selected and instructed to 1) think of any male talker and 2) estimate the weight of that talker" (p.1885). The results were compared with the actual weight of 14 men chosen at random and it is singular to see that the mean of the imagined weights is very near to the mean of the weights of "non-talker subjects". So, the authors say, "the new procedures have produced an improvement in accuracy" (p.1885).

In our opinion, however, the point is not to "guess" the speaker's age, weight or height, but to verify whether the voices of different speakers suggest different physical features to the listeners. In this case, independently of the degree of guessing accuracy, estimated values would vary by varying the actual values. The problem at this point, is to find the correlation between estimated and actual values.

The data obtained in a previous study on height and weight (6), seem to suggest that such a correlation exists. On the other hand if a close relationship exists between the acoustic characteristics of the voice and speaker age, weight and height, it should presumably be very difficult to recognize the voices of twins having an identical physical structure.

The absence of previous specific studies on twins' voices, except for Rosemberg's (7) and Lummis' (8) on the automatic recognition of the speaker, has suggested the advisability of experimental research on this subject. The results obtained in a previous
work (9) suggest that listeners have great difficulties in distinguishing between twins on the basis of their voices, although this confusion is not due exclusively to the twins’ having nearly the same physical features.

PERCEPTIVE TESTS

A total of 100 listeners participated in the study. All were students with no reported hearing difficulty. They participated in a total of 3 sessions. The tapes were presented by means of headphones to each listener seated in a sound booth.

FIRST TEST

A standard prose passage was recorded by 36 speakers, 18 males and 18 females. A paired comparison procedure was employed in the preparation of the tape. The differences in age, weight and height between the two elements of each pair were of at least 10 years, 15 cm and 15 kg (males), 10 years, 10 cm and 10 kg (females). Furthermore, 6 pairs represented the highest vs. lowest values of age, height and weight.

The listeners were asked to judge which member of each pair was the elder, the heavier or the taller.

RESULTS

The highest percentages of correct judgements for age, height and weight (94%, 94% and 75% for male; 100%, 74% and 95% for females) were obtained for pairs representing the highest vs. lowest values. Also for the other pairs however, correct judgements are always high ranging from 55% to 94%. Therefore we can say that different ages, heights and weights determine different estimations.

At this point the problem is to find the relationship between actual and estimated values. To do this, we have prepared a second test.

SECOND TEST

A tape with the 36 prose passages in a randomized order was constructed and presented to the listeners. They were asked to estimate age, weight and height for each voice.

RESULTS

We have plotted actual and estimated ages, heights and weights and we have drawn the curves of the third degree approximants according the least squares method, calculated the standard deviations and compared with the straight line $x=y$. The curves of height and weight, even if they show a slightly ascending trend, are not so meaningful as the curves of age. The latter cross the line $x=y$ in a determined point representing the average value that listeners assigned to male and female speakers respectively.

Below this average value the age tends to be overestimated, above this value the age tends to be underestimated. Therefore it would be possible to determine the age of a speaker with a high degree of accuracy by a test of this kind by finding the $x$ value corresponding to the estimated $y$ value (Fig. 1).

THIRD TEST

Three pairs of identical twins were employed as speakers: a pair of male adults /A/, 48 years old having the same height (174 cm) and weight (66 kg); a pair of female adults /B/, 33 years old having the same height (156 cm) and weight (46 kg); a pair of female children /C/ at a different stage of puberty 13 years old having different height (158 cm and 166 cm) and weight 38 kg and 46 kg).

Twelve sentences and twelve words were uttered by each pair. A tape with pairs of same sentences and different words was constructed. Listeners were asked to say whether both elements of each pair had been uttered by one only of the twins, or one element by each twin.

RESULTS

The results of the third test pointed out that, on one hand, the /A/ and /C/ twins’ voices are so similar as to generate a high degree of confusion in the listeners; on the other hand, the /B/ twins’ voices are much more correctly identified by listeners.

The average percentage of confusion, two voices falsely accepted as one, is 36%, 6% and 25% for pairs /A/, /B/ and /C/ respectively.
Fig.1. Actual and estimated age, height and weight.
The average percentage of confusion, one voice falsely accepted as two, is 32%, 10% and 26% for pairs /A/, /B/ and /C/ respectively (Fig. 2).

The percentage relative to /A/ and /C/ pairs are quite different from those obtained in a previous study (9). In fact the percentage of confusion relative to two voices falsely accepted as one has decreased, whereas it has noticeably increased for one voice falsely accepted as two. These differences must be ascribed to the fact that, contrary to the first study, listeners knew that they had to identify twins voices so they probably paid attention also to intraspeaker variations. Contrary to what we found in the previous study, percentages for sentences are not higher than those for words. This is because in the present study listeners had to judge pairs composed by different words and their accuracy dropped because of the lack of segmental differences that, however small, seem to contribute to a correct identification.

As regards Fo, the closer the two curves are, the higher is the level of confusion. As regards duration, there are no considerable differences, being in the range of 0.02 and 0.07 sec.

On the segmental level, there are slight variations of the first two formants of vowels, more relevant in the words than in the sentences.

REFERENCES


In the present paper the part on age, height and weight is mainly due to M. Pettorino, the part on twins' voices to A. Giannini and the elaboration and analysis of the data to U. Cinque.