SPEAKER AND LISTENER SEX FOR SPEAKER HEIGHT AND WEIGHT IDENTIFICATION

Wim A. van Dommelen

Department of Linguistics, Trondheim, Norway

ABSTRACT

This study examines the ability of listeners to judge speaker height and weight from speech samples. The results show that mainly male listeners were able to estimate male speaker height and weight. Neither male nor female listeners could judge female speaker height or weight. The data suggest that the listeners correctly used speech rate information in judging males. But low F0 and formant frequency values were wrongly taken to indicate large body dimensions.

INTRODUCTION

This investigation takes a look at the impression of speaker height and weight conveyed by the voice of an unknown speaker. The first question is whether or not such impressions are correct. Though previous investigations do not seem to give definite answers to this question, existing data indicate that listeners are very consistent when estimating speaker height or weight [1, 2]. This gives rise to the question what kind of information contained in the speech signal is used by the listener and in what way. One potentially important factor could be speaking fundamental frequency (F0). F0 acoustic scale values varying from low to high presumably correspond to perceptual scale values varying from "tall speaker" to "small speaker" [3].

In order to shed light on the issues mentioned above, for the present study listener judgements on speaker height and weight were collected. To see whether these judgements were correct, estimated height/weight was compared with actual speaker height/weight. The listeners' use of information contained in the speech signal was investigated by relating the height/weight estimates to three different parameters: F0, formant frequency and speech rate.

EXPERIMENTAL PROCEDURE

Recordings

Speech recordings were made in the studio of the Linguistics Department of Trondheim University from a group of speakers consisting of 30 volunteers (15 males and 15 females), who were mainly recruited from the university's student population. Their ages ranged between 23 and 32 years for the males (with an average of $\bar{x}$= 26.0; s= 2.8) and between 20 and 36 years ($\bar{x}$= 25.5; s= 3.8) for the females, respectively. For the occasion of the recordings, the speakers' height and weight were measured with an accuracy of 1cm and 0.5kg, respectively, using a Lindeltronic 4000 from the Department of Sport Sciences.

The first part of the speech material that was recorded consisted of a list of 10 isolated Norwegian words. Secondly, the first two paragraphs of a Norwegian fairytale were recorded. The text was of an epic character without direct speech and was slightly modified to make its style more up-to-date. The speakers were instructed to read both words and text using their natural voices.

Analysis

Analyses of the speech recordings were performed with Signalyze [4] for each of the parts that were used in the listening tests: ten isolated words and each of the two text paragraphs. For each of the 30 speakers, the following signal parameters were investigated:

(a) average F0 of the voiced signal portions. Gross errors in the analysis were corrected by hand by deleting the actual frames. Given the three text conditions, this procedure resulted in three average F0 values per speaker.

(b) F2-frequency of the neutral vowel [ə], which occurred in final position in two isolated words and in four (paragraph 1) and seven (paragraph 2) words, respectively, from the fairy tale. Taking the relation $v= f*\lambda$ (where v= velocity of sound in air, f= frequency and λ= wavelength) as a point of departure, vocal tract length l was estimated according to the formula $l= 3/4*v/F2$. Subsequently, average estimates were calculated based on n= 2, 4, and 7 measurements, respectively, for the three conditions. Henceforth, the estimates of vocal tract length will be abbreviated VT.

(c) speech rate, which was defined differently for the two text sorts. As a result of preliminary heuristic testing, for the isolated word condition the total duration of the ten words was taken as a measure, leaving out intermediate pauses. However, since pause durations in the texts obviously were highly speaker-specific, overall text durations were taken as a speech rate measure for paragraphs 1 and 2. In only a few cases, the total duration had to be corrected for speech errors. In the description, this measure will be referred to as Duration.

Listening tests

The speech material described above was used to prepare a set of 12 listening tapes, six for the groups of male and female speakers each (cf. Table 1). In preparing tapes 1-6 the original recordings were dubbed onto a second cassette; for tapes 7-12 the signals were band-pass filtered between 250Hz and 3.15kHz. All tapes were constructed according to the following pattern: each stimulus, i.e. the ten words, paragraph 1 or 2, was preceded by a voice announcing the next speaker number, and followed by a ca. 10 sec answering pause. Each tape contained each of the 15 male/female speakers once, each time with a different randomization. This means that tapes 7-12 contained the same speech material as tapes 1-6, the difference lying in randomization order and filtering condition. The main goal of presenting the same material also band-pass filtered was to collect more data, in order to be able to investigate the listeners' consistency.

Table 1. Presentation order of the 12 listening tapes. Each tape contained randomized speech samples from 15 speakers.

	unfiltered	filtered
words	(1) males	(7) males
	(2) females	(8) females
paragr. 1	(3) males	(9) males
	(4) females	(10) females
paragr. 2	(5) males	(11) males
	(6) females	(12) females

A total of 20 listeners (10 males, 10 females) were asked to estimate the speakers' height and weight using prepared answering sheets with height/weight scales for each speaker. The tapes were presented over headphones in two separate sessions (tapes 1-6 and 7-12).

RESULTS

Voice and body characteristics

First of all, it was investigated whether the acoustic cues analyzed here vary systematically with actual speaker height and weight. The results of a multiple regression analysis for the male speakers showed that neither average F0 nor vocal tract length estimate VT are reliable predictors of speaker height and weight. A remarkable exception in this connection is that heavier men had a tendency to speak at slower speech rates: all three correlations between male body weight and speech rate under the different text conditions Words, Paragraph 1 and 2 were statistically significant (r= 0.592, p= 0.024; r= 0.762, p= 0.001 and r= 0.712, p= 0.004, resp.). Male body height, however, does not correlate with speech rate, in spite of a general positive correlation between height and weight. From this, it can be concluded that it is the surplus of weight going beyond the normal weight increase due to greater height that is responsible for the lower speech rate.

As far as the female speakers are concerned, only one out of 18 correlations between height/weight and acoustic cues reached statistical significance (VT for height in Paragraph 2; r= 0.516, p= 0.048). The lack of a significant correlation between female weight and speech rate reveals an interesting sex-specific difference in speech production.

Estimation of height and weight

This paragraph deals with the question whether the listeners were able to estimate the speakers' height and weight from the speech samples. The data concerning the correlation between estimated and actual height and weight are presented in Table 2. For the male speakers, a considerable number (14) of positive correlations between estimated and actual height/weight were found. Eleven out of these 14 significant correlations go back on the male listeners, indicating that it was mainly this listener group that had accurate ideas concerning male height and weight. This in contrast with the far less clear results for the group of female

Table 2
Correlation coefficients and probabilities for correlations between estimated speaker height/ weight and actual height/weight for two filtering conditions (unfiltered and band-pass filtered) and text conditions words, paragraph 1 and paragraph 2. 10 male (a) and 10 female (b) listeners. Only values statistically significant at a 5% level of probability are given. None of the correlations for the tests involving female speakers reached statistical significance.

(a) Male listeners - male speakers

		Words		Paragraph 1		Paragraph 2	
unfiltered	height	r= 0.535	p= 0.040	r= 0.526	p= 0.044	r= 0.578	p= 0.024
	weight	r= 0.591	p= 0.020	r= 0.724	p= 0.002	r= 0.840	p= 0.000
filtered	height	r= 0.530	p= 0.042	———	———	r= 0.609	p= 0.016
	weight	r= 0.865	p= 0.000	r= 0.722	p= 0.002	r= 0.790	p= 0.000

(b) Female listeners - male speakers

		Words		Paragraph 1		Paragraph 2	
unfiltered	height	———	———	r= 0.570	p= 0.027	———	———
	weight	———	———	r= 0.627	p= 0.012	———	———
filtered	height	———	———	———	———	———	———
	weight	r= 0.687	p= 0.005	———	———	———	———

listeners (compare Table 2a with 2b).

In addition to this sex-specific behaviour on the part of the listeners, a remarkable sex-related speaker factor is found: None of the correlations between estimated and actual female speaker height or weight turned out to be significant - neither for male nor for female listeners. This finding indicates that speech production differs significantly between the two groups of speakers and, in all probability, is sex-specific.

Use of acoustic cues

To shed light on the question of what kind of information the listeners used to arrive at their judgements, multiple regression analyses were performed involving F0, VT and Duration as predictor variables and estimated height/weight as dependent variables. Table 3 presents the results for male speaker height/weight as estimated by males and females. As can be seen from the table, mean F0 correlated negatively with both estimated height and weight (17 out of 24 cases significant). So, a low F0 was taken as an indication of a tall, heavy speaker, whereas high F0 values pointed to small bodily dimensions. Similarly, the data suggest that low formant frequency values (high VT values) were associated with large body dimensions. A long text duration, i.e. a slow speech rate, gave rise to the impression of a tall and (especially) heavy speaker. All the tendencies noted here are stronger for the male than for the female listeners (compare Table 3a with 3b).

Table 3
Number of times the pos(itive) or neg(ative) correlations between estimated male speaker height/weight (EMH/EMW) and F0, VT, and Duration reached statistical significance (p< 0,05%). max= 6.

a) male listeners

	F0	VT	Duration
EMH	6 neg	2 pos	1 pos
EMW	5 neg	5 pos	4 pos

b) female listeners

	F0	VT	Duration
EMH	3 neg	0	1 pos
EMW	3 neg	3 pos	2 pos

Comparison of the data from Table 3 with the corresponding ones for female speaker height/weight (Table 4) reveals an effect of speaker sex also with respect to the use of acoustic cues. Only male listeners show a weak tendency to associate low F0 values with greater height/weight (3 out of 12 cases significant). VT, however, correlates more often with perceived height/weight (9 out of 24 cases significant). In strong contrast with the results for male speakers, varying speech rate appears to have no influence on the height/weight ratings, neither for men nor for women.

Table 4
Number of times the pos(itive) or neg(ative) correlations between estimated female speaker height/weight (EFH/EFW) and F0, VT, and Duration reached statistical significance (p< 0,05%). max= 6.

a) male listeners

	F0	VT	Duration
EFH	1 neg	4 pos	0
EFW	2 neg	2 pos	0

b) female listeners

	F0	VT	Duration
EFH	0	2 pos	0
EFW	0	1 pos	0

DISCUSSION

In line with previous investigations [5, 6], the present correlations between mean speaking F0 and speaker height/weight were statistically nonsignificant. Estimated vocal tract length, too, did not vary systematically with body dimensions. Since mean F0 and formant frequencies are largely dependent on the dimensions of the laryngeal and supralaryngeal parts of the speech mechanism, this outcome suggests that these dimensions vary independently of the rest of the human body.

An unexpected result was the observed tendency for male speakers to have reduced speech rate with increasing body weight. This raises the question whether this phenomenon is due to biological or to socio-cultural factors. At present, it is also an open question why the women in this study behaved differently in this respect.

Interestingly, the sex-specific behaviour in speech production is also reflected in the identification of speaker height/ weight in that the listeners' ability to estimate these dimensions was confined to the group of male speakers. It was mainly the group of male listeners who were able to do so. This suggests that this group exploited the typically male speech rate behaviour.

The results have shown that speech rate is not the only source of information used by the listener. The final impression of the speakers' body characteristics on the listener turned out to be multifacetted. Obviously, also F0 and spectral information are important factors. Differently from the correct use of speech rate for estimating male speaker weight, F0 and spectral parameters were exploited inappropriately. Low F0 as well as low low formant frequency values were taken to indicate large body dimensions. High values were interpreted as originating from a small, light speaker. The incorrect use of these parameters, however, had no crucial impact on the (correct) use of speech rate information. In all probability, the reason for this must be sought in the fact that there was no correlation between speech rate on the one hand and mean F0 or formant frequencies on the other. This means that possible modifications of speech rate based judgements were stochastical, rather than systematic.

References

[1] Lass, N.J., Phillips, J.K. & Bruchey, C.A. (1980), "The effect of filtered speech on speaker height and weight identification", *Journal of Phonetics*, vol. 8, pp. 91-100

[2] Dommelen, W.A. van (1993), "Speaker height and weight identification: a re-evaluation of some old data", *Journal of Phonetics*, vol. 21, pp. 337-341

[3] Ohala, J.J. (1984), "An ethological perspective on common cross-language utilization of F0 of voice", *Phonetica*, vol. 41, pp. 1-16.

[4] Keller, E. (1993), *Signalyze™. Signal Analysis for Speech and Sound.* Network Technology Corporation, Charlestown.

[5] Hollien, H., Green, R., and Massey, K. (1994), "Longitudinal research on adolescent voice change in males", *Journal of the Acoustical Society of America*, vol. 96, pp. 2646-2654.

[6] Künzel, H.J. (1989), "How well does average fundamental frequency correlate with speaker height and weight?", *Phonetica*, vol. 46, pp. 117-125.