THE PERCEPTION OF MALENESS AND FEMALENESS IN THE VOICE AND ITS RELATIONSHIP TO VOWEL FORMANT FREQUENCIES

RALPH O. COLEMAN

The relationship between sex and laryngeal fundamental frequency, which is a lower fundamental for males, has long been recognized and it has been traditional to assume that a listener identifies the sex of a speaker on the basis of the laryngeal fundamental. Recent studies, however, suggest that listeners may be able to recognize speaker sex when the laryngeal fundamental is absent. Weinberg (1971) very recently reported that speaker sex identity was retained by a group of laryngectomized speakers. Schwartz (1968) and Schwartz and Rine (1968), also observed this in whispered vowels and voiceless fricatives produced by normal speakers. A possible explanation for this may lie with the differences in vocal tract resonance characteristics, as evidenced by vowel formant frequencies, that have been found to exist between males and females. These have been described by Peterson and Barney (1952) and others (Ladefoged and Broadbent 1957). It was the purpose of this study to explore the possibility that these resonance differences are a cue to speaker sex. Specifically the study was designed to answer these questions: (1) is it possible to distinguish between male and female speakers when the variations in fundamental frequency between sexes have been eliminated, and (2) if these distinctions can be made, are they a function of vocal tract resonance differences between male and female speakers as revealed by vowel formant frequencies?

The study was carried out in two parts. In the first part a panel of listeners was asked to determine the sex of a mixed group of adults articulating the sound produced by an electro-larynx having a fundamental frequency of 85 Hz. In the second part of the study listener judgments were compared with the frequencies of the formants of the vowels /i/ and /u/ produced in isolation by these speakers.

1. PROCEDURE

The subjects were 20 normal speaking adults: 10 males and 10 females.

The vocal sound source used in the study was a commercially produced electronic larynx which produced a steady buzz of 85 Hz $\pm$ 3 Hz. A standard speech sample

plus the isolated vowels /i/ and /u/ were tape recorded for each speaker using the electro-larynx buzz rather than the laryngeal tone as the sound source.

A listening tape was prepared and presented to a judging panel of 15 university students who were instructed simply to listen to the tape and determine the sex of each speaker. They were told only that some of the speakers were male and some female, without specifying the number of each. In addition they were asked to indicate the confidence with which their selection for each speaker was made on a seven-point scale where a rating of 1 would indicate a 'guess' and a 7 would represent complete confidence in their choice. Intermediate values on the scale would, of course, indicate degrees of certainty between these extremes. When the listener judgments were actually obtained, the panel was allowed to listen to the entire tape without attempting any speaker identifications. This was done in order to provide the panel with a perceptual frame of reference and their judgments were then based on a second playing of the tape which followed immediately.

In the second part of the study, vowels which had been produced in isolation by each subject were analyzed spectrographically and the frequency of the first three formants obtained for each vowel. These values were averaged and the resulting individual averages were considered to represent the over-all vocal tract resonance characteristics of each subject. These were then compared with the degree of male or female quality in each voice as indicated by the extent to which judges agreed on the sex of the speaker and over-all certainty with which these identifications were made.

2. RESULTS

The results of the listener judgments are shown in Table 1. The average certainty for each speaker as shown in the right column was obtained by assigning pluses to correct identifications and minuses to incorrect identifications. The 15 ratings for each subject were then summed algebraically and an average computed. As can be seen 18 of the 20 subjects had 'plus' averages indicating that they were judged to have a voice quality appropriate to their sex although the degree of this quality varies from one subject to the next. The exceptions were two female subjects with minus averages, indicating that the judges frequently mistook them for male speakers and these two females would be considered to have slightly male voice quality.

It is apparent that it was easier for the judges to correctly identify the male speakers. Seven of the ten male subjects were correctly identified by all 15 judges and the remaining three were correctly identified by 14 judges. All judgments were made with a high degree of confidence, as indicated by the low average of 3.5 and the high of 6.7. A 7.0, as you may recall, indicated complete confidence on the part of the judges. The results for the female subjects were generally similar. There were however, more incorrect identifications and somewhat less confidence over-all on the part of the judging. For the two groups together, 88% of the identifications were correct

TABLE 1

Listener identifications of the sex of speakers and the average of the certainty with which these identifications were made. A higher average signifies a greater degree of certainty on the part of the 15 judges and is interpreted as signifying a more distinct male or female voice quality.

Subject	Times Sex Correctly Identified		Times Sex Incorrectly Identified		Average Certainty on Scale 1-7	
	Males	Females	Males	Females	Males	Females
1	15	15	0	0	6.7	6.6
2	15	15	0	0	6.6	6.4
3	15	14	0	1	6.2	5.7
4	15	15	0	0	5.9	5.1
5	15	15	0	0	5.8	3.7
6	15	12	0	3	5.4	2.3
7	15	12	0	3	5.0	1.7
8	14	10	1	5	4.0	1.6
9	14	6	1	9	3.8	—0.5
10	14	4	1	11	3.5	—1.9
Total	147 (98%)	118 (79%)	3 (2%)	32 (21%)		

and 12% were incorrect, with most of the incorrect identifications accounted for by three female subjects.

These findings would indicate that the voices, particularly the male voices, contained a distinct male or female quality in the absence of inter-subject variation in the fundamental frequency. At the same time, the range in the certainty ratings indicates that this quality exists on something of a continuum from strongly female with scatter in between these extremes.

The spectrographic analyses of /i/ and /u/ were correlated with the listener judgments and are shown on the next three slides. The abcissa represents a continuum between male voice quality, on the left side, and female voice quality, on the right side, with subject identifications located appropriately. The degree of maleness or femaleness in a voice, therefore, is indicated by how far to that side of the midpoint the subject's average falls. The two female subjects who were judged to have a slightly male voice quality, appear, as a result, on the male side of the abcissa. It should be pointed out that the means of the three formants in each vowel were used as a basis for comparison with listener judgments rather than individual formant values because it was felt that the average more closely represented the listener's actual experience in perceiving the vowels and would be more representative therefore, of individual vocal tracts. Figure 1 is for /i/, Figure 2 is for /u/, and Figure 3 is for /i/ and /u/ combined.

These figures show that, while the relationship between male and female voice

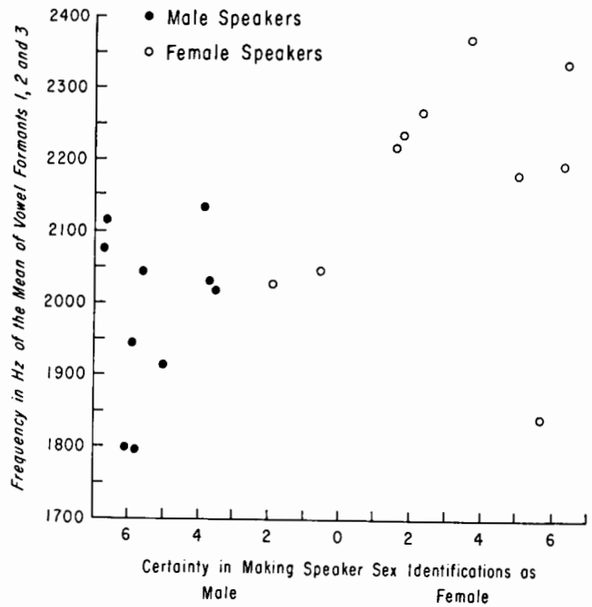

Fig. 1. Average of the frequencies of Formants 1, 2, and 3 for the vowel /i/ versus the certainty with which speaker sex identifications were made. The certainty figures represent the average of the ratings given by all 15 judges. A higher average signifies greater certainty and is interpreted as signifying a more distinct male or female voice quality.

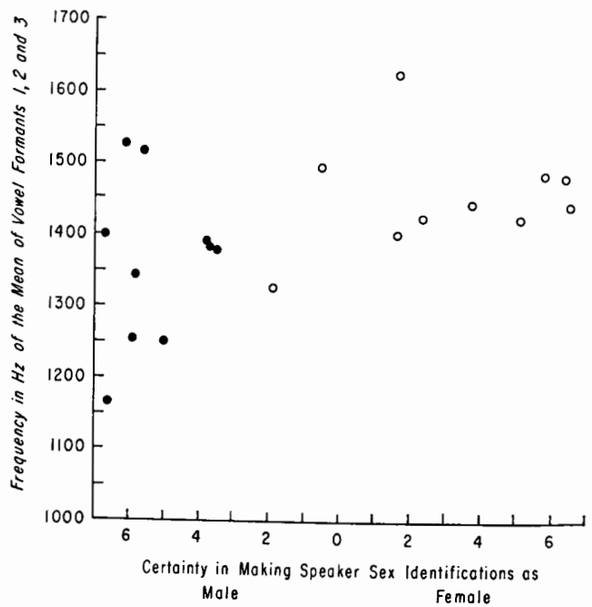

Fig. 2. Same as Fig. 1 for the vowel /u/.

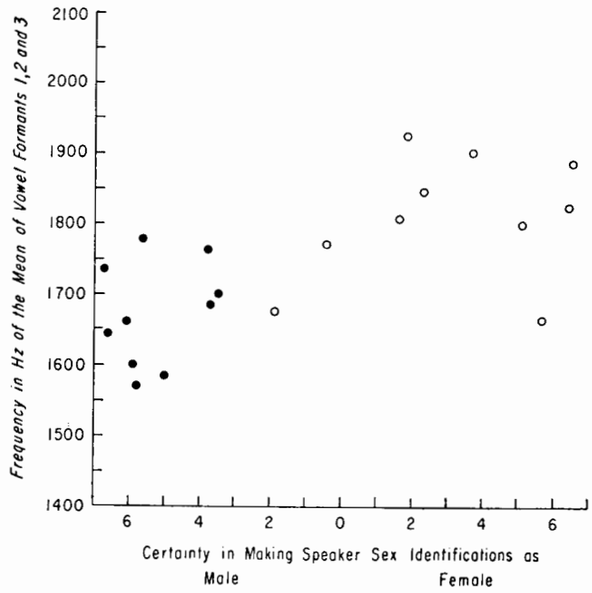

Fig. 3. Same as Fig. 1 for the vowels /i/ and /u/ combined.

quality and formant frequencies was not a perfect one, there was a statistically significant degree of correspondence between them.

3. DISCUSSION

The results of the listener judgments indicate that it is possible to recognize the sex of a speaker when a single frequency sound source is substituted for the laryngeal tone and it is clear that these identifications can be made with considerable confidence in many instances. Whether vocal tract resonances, as manifested by formant frequency averages, are the only cue to speaker sex under these experimental conditions was not completely demonstrated. It was apparent, however, that for most subjects a male quality in the voice was associated with lower vowel formants and a female quality in the voice was associated with higher formants. Where the formant frequency averages for males and females are close together, the cue to speaker sex appears to be considerably less distinct.

The listener judgments in the study more clearly distinguished between the sexes than did the formant frequency averages. In view of the fact that virtually all phonemic elements have spectra that are affected by the resonance qualities of the vocal tract, this should not be surprising. These include the noise burst following a stop-plosive, the peak frequency of the noise in sibilance and affrication, and the formant structure of vowels, semi-vowels, and nasals. It is likely that cues to vocal tract size are con-

tained in virtually all phonemic elements, though the more intense elements, such as vowels, might be expected to provide stronger cues. The spectrum analysis in the present study was made on only two isolated vowels, while the listener judgments were based on a conversational segment containing 175 phonemic elements in addition to the two vowels. Had the entire conversational segment for each speaker been analyzed, it is likely that the correlations between resonance peak frequencies and male and female voice quality would have been higher.

That /i/ and /u/ alone are insufficient to characterize the vocal tract of a speaker is evidenced in other ways. There is overlap between the formant frequencies of some female and male subjects whose sex was correctly identified. If the only cue to speaker sex was contained in the formant frequencies of the vowels studied, the populations should have been as dichotomous on the basis of formant frequency averages as they were on listener judgments.

The results of the two parts of this experiment considered together indicate that the perception of male and female voice quality is probably based on some auditory sensation of 'vocal pitch' which is the result of a combination of acoustic cues. While the most obvious acoustic component is likely the difference in the fundamental frequency of males and females, this study demonstrates that a second component may be the location in the frequency spectra of vocal tract resonances. These may provide a cue to the size of the vocal tract which could in turn indicate the probable head and neck size of the speaker. The extent of male or female quality in a particular voice then, would likely be determined by the interaction of these two acoustic cues, assuming that the perception of vocal pitch does result from the combining of this information. For instance, a low frequency laryngeal fundamental in combination with low frequency vocal tract resonances would likely result in a voice that would be perceived as strongly male. Other combinations would result in either a strongly female voice or one with varying degrees of maleness or femaleness. A continuum of this kind was indicated when the judges demonstrated varying degrees of certainty in determining the sex of different speakers.

Since separate physical structures determine the characteristics of the laryngeal fundamental and the vocal tract resonances, it would be possible for conflicting cues to speaker sex to be present in one individual. The fact that the dimensions of both the larynx and the vocal tract are probably influenced by the individual's physical stature would tend to reduce the likelihood of this occurring, but it would not prevent it.

University of Oregon
Portland, Oregon

REFERENCES

Ladefoged, P. and D. Broadbent
1957 "Information Conveyed by Vowels", *Journal of the Acoustical Society of America* 29: 98-104.

Peterson, G. and H. Barney
1952 "Control Methods Used in a Study of the Vowels", *Journal of the Acoustical Society of America* 24:175-184.

Schwartz, M.
1968 "Identification of Speaker Sex from Isolated Voiceless Fricatives", *Journal of the Acoustical Society of America* 43:1178-1179.

Schwartz, M. and H. Rine
1968 "Identification of Speaker Sex from Isolated Whispered Vowels", *Journal of the Acoustical Society of America* 44:1736-1737.

Weinberg, B. and S. Bennett
1971 "A Study of Talker Recognition of Esophageal Voices", *Journal of Speech and Hearing Research* 14:391-395.

DISCUSSION

GUIRAO (Buenos Aires)
Did you correlate your scale of uncertainty with your data from the spectrograms?

COLEMAN
Yes, the certainty ratings were correlated with the vowel formant frequencies taken from the spectrograms.

GUIRAO
It would be interesting to see data for more vowels and take a look at the correlation between uncertainty and acoustical spectrographic data.

COLEMAN
Weighting of the different formants was considered by me, but I know of no information that would indicate the importance to perception of the vowels of the different formants, particularly to the perception of maleness or femaleness. Without this information a meaningful weighting would be impossible.

LEIDNER (Brookline, Mass.)
1. About 15 years ago, Peterson and Barney did a study of the various formant frequencies of vowels of men, women, and children. They implied that vowel formant frequencies may be one of the cues for the perception of sex distinction. How does your study and its conclusion differ from theirs?

2. You mentioned that there has never been a study before which was aimed at demonstrating that vowel formant frequencies are a cue to vocal tract size, but I would call your attention to a paper read by Mr. Timothy Rand at the April, 1971 meeting of the Acoustical Society of America, in which he reports just this kind of

research. An abstract is contained in the issue of the *Journal of the Acoustical Society of America* published some time after that meeting.

COLEMAN

Peterson and Barney were concerned with vowel specification and utilized vowel formant frequencies for a large sample of men, women and children. He was not, as far as I know, concerned with any of the perceptual aspects of the vowels he analyzed. My study is concerned with the relationship between perceptions of speaker sex and vocal tract resonances.