A comparison of the contributions of two vocal characteristics to the perception of maleness and femaleness in the voice

Coleman, R. O.

journal: STL-QPSR
volume: 14
number: 2-3
year: 1973
pages: 013-022

http://www.speech.kth.se/qpsr
II. SPEECH PERCEPTION

A. A COMPARISON OF THE CONTRIBUTIONS OF TWO VOCAL CHARACTERISTICS TO THE PERCEPTION OF MALENESS AND FEMALENESS IN THE VOICE*

R. O. Coleman

Abstract

Two experiments were carried out in which a comparison was made between the contributions of \( F_0 \) on one hand, and vocal tract resonances on the other, to a perception of maleness and femaleness in the adult voice. In the first experiment, in which natural voice was used, the frequency of the \( F_0 \) was found to be very highly correlated \( (r_s = .94) \) with the degree of maleness and femaleness in the voice. The vocal tract resonances were less highly correlated and it is apparent that in the presence of the natural laryngeal tone, these perceptions are based on the frequency of the \( F_0 \).

In the second experiment a tone produced by a laryngeal vibrator was substituted for the normal glottal tone at simulated \( F_0 \)'s representing both males (120 Hz) and females (240 Hz). When listeners were asked to identify the sex of the speakers some inconsistency with the findings of the first experiment was seen. The female \( F_0 \) was a weak indicator of female voice quality when combined with male vocal tract resonance although the male \( F_0 \) retained the perceptual prominence seen in the first experiment. This finding may be indicative of some basic difference in the normal glottal characteristics of males and females.

Introduction

It has long been recognized that the shorter overall vocal tract length in females results in an upward shift in the frequency of female vowel formants. The degree of this shift has been estimated to be on the order of about 20% (Peterson and Barney, 1952), although individual differences in vocal tract length can be expected to affect the exact amount of shift occurring between any two individuals.

The perceptual significance of these differences has been studied by Coleman (1971) who reported that in the absence of normal differences in \( F_0 \), the sex of the speaker can be recognized with little difficulty in most instances. Confusion occurred when the speaker's vocal tract resonances were more like that of the opposite sex than of his own. The degree of maleness or femaleness in the voice was also found to be significantly correlated with the frequency of the resonances of the person's speech.

* Guest researcher from the University of Oregon Medical School, Portland, Oregon, USA from 10 February to 7 August, 1973.
In other words, the higher the vowel formant frequencies, the more female-like the voice was felt to be and vice versa. This finding would suggest that under certain circumstances listeners are quite sensitive to the relatively small differences in vocal tract resonances that occur between speakers. It appears that listeners are also sensitive to resonance differences in certain consonants. Schwartz (1968) and Ingemann (1968) have found, for example, that the sex of a speaker can be accurately identified on the basis of isolated, voiceless fricatives.

On the basis of the perceptual significance of the vocal tract resonances, it has been hypothesized that a perception of a speaker's vocal "pitch" and subsequently the maleness and femaleness of his voice, results from the combining of the information conveyed by both the laryngeal fundamental and the resonances of the vocal tract (Coleman, 1971). This hypothesis has implications for the automatic recognition of speech as well as for the understanding of certain types of voice disorders and two experiments designed to test it were carried out. In both, the purpose was to compare the contributions of the laryngeal fundamental and vocal tract resonances to a perception of maleness and femaleness in the voice.

In the first experiment, utilizing natural voice, correlation coefficients were computed between measures of vocal tract resonance, laryngeal fundamental, and judgments of the degree of maleness or femaleness in the voice. The degree of correlation is considered to indicate the contribution of each of the two vocal characteristics to the listener judgments. In the second experiment, the two vocal characteristics were contrasted so that listeners were presented with speech composed of vocal tract resonances characteristic of one sex in combination with an $F_0$ characteristic of the opposite sex and asked to identify the sex of the speaker. In such a forced-choice situation, listener identifications would presumably be based on the more perceptually prominent of the two vocal characteristics.

**Speakers**

The persons from whom the speech samples were obtained were twenty normal adult males and a like number of females selected from an American university population. Speakers were essentially unselected except those whose first language was not English or who had some obvious speech defect or regional accent were not included.
Recording procedure

Tape recordings were made on a high fidelity array that included an Ampex 601 recorder and an Electro-Voice 666 microphone. The following spoken material was recorded by each speaker: Two repetitions of a standard six-word sentence, three productions of each of the vowels /i/, /u/, /ɛ/, and /a/, and three versions of a portion of the "Rainbow Passage". These particular vowels were selected since they are representative of the extreme tongue positions during production. This is an important consideration since the difference between the vowel formant frequencies of males and females can be expected to vary from one vowel to the next. This is because, according to Fant (1966), it is the pharyngeal cavity that is shorter in females with the oral cavity being about the same length in both sexes. The change from males to females, therefore, would be more pronounced in those vowels having formants associated with the pharyngeal cavity, for instance F₂ of /i/.

The sample sentences and one version of the Rainbow Passage were produced with a normal voice. Two other versions of Rainbow Passage were produced with a laryngeal vibrator (Western Electric, Model 5 Electro-larynx) substituting for the glottal sound at frequencies of 120 Hz and 240 Hz. These were selected as being representative of the typical male and female F₀. The isolated vowels were also produced with the laryngeal vibrator producing a 100 Hz sound. This was done in order to enhance the spectrographic analysis of the vowels since it is sometimes difficult to identify the higher formants of female speakers, particularly those with a high F₀. Subjects were instructed to shape their mouths to produce each target vowel and allowed to practice with the laryngeal vibrator until easily recognizable vowels could be produced.

A problem inherent in using a vibrator produced tone to substitute for the normal laryngeal tone is that if the trachea is coupled to the rest of the vocal tract by a glottal opening, the resonant characteristics of the vocal tract are altered. This has been discussed by Fant et al. (1972) who point out that introducing an exteriorly based sound source near the level of the open glottis results in the introduction of the resonances and antiresonances that originate in the trachea. These are capable of bringing about a fundamental alteration in the transfer characteristics of the vocal tract and in the vowels produced by that vocal tract.
This effect is illustrated in Fig. II-A-1. A spectrogram of the vowel /i/ produced by the author using a laryngeal vibrator with the glottis both open and closed is shown. As a result of the coupling of the trachea to the rest of the vocal tract, \( F_2 \) of this vowel has been virtually eliminated and \( F_1 \) shifted upward in frequency. The perceptual change accompanying this is equally marked and with the glottis open the vowel is unrecognizable. It is important, therefore, that the glottis remains closed when producing vowels with a laryngeal vibrator in order to insure a normal formant structure.

In this study this problem was handled retrospectively by reviewing all vowels and eliminating those which did not have a quality easily recognizable as being that of the target vowel. Virtually all the vowels had an acceptable quality and it is likely that persons normally keep their glottis closed when producing vowels in this way. At the same time it is impossible to verify for certain that some degree of coupling between the trachea and the rest of the vocal tract did not exist. Vowel formant values have, therefore, been reported as rankings within the sample population rather than as specific formant frequencies in Hz. The assumption is that the speakers would be relatively consistent in their performance and whatever shift in formant frequency may be occurring will also be consistent for this group of speakers.

**Formant measures**

Formant frequencies for the four vowels under study were obtained from hand-drawn lines-of-best-fit around intensity-by-time spectrographic sections of each of the isolated vowels. The frequency of the most intense harmonic of each formant was designated as the frequency of that particular formant. Individual values for formants 1, 2, and 3 (\( F_1 \), \( F_2 \), and \( F_3 \)) for each vowel were averaged and the resulting figure used to represent the relative vocal tract resonance characteristic of each speaker. It was found that it was not always possible to clearly distinguish between \( F_1 \) and \( F_2 \) in the /a/ produced by several speakers and the /a/ was subsequently eliminated from consideration. Rankings of vocal tract resonance characteristics are based, therefore, on measures of /i/, /ɛ/, and /u/, only.
Fig. II-A-1. Vowel /i/ produced using a laryngeal vibrator with speaker's glottis open and closed.
Fundamental frequency \((F_0)\)

Fundamental frequency was measured for each speaker by determining the number of vertical striations appearing on intensity-by-time spectrograms in two standard sentences and dividing by the duration of the voiced portion of the sentences. The resulting \(F_0\) values were averaged for each speaker and ranked within the forty speaker group.

Listener judgments - Experiment I

A listening tape was prepared which consisted of 5 sec samples taken from each of the forty speakers' normal voice production of the Rainbow Passage. These were presented to a group of 17 young adults who were asked to 1) determine the sex of each speaker, and 2) to estimate, on a scale of 1-7, how much of the quality they associate with that particular sex each voice contained. In order to reduce the influence of possible male/female differences in rate, juncture, and inflection, the five-second samples were presented backwards. The judgments should, therefore, be based solely on vocal tract resonance and fundamental frequency information which would be unaffected by the backward presentation.

The ratings given each subject were averaged and ranked within the forty subject group. The ranking of each subject is considered to represent the relative degree of maleness or femaleness \((M-F\) voice quality) in each speaker's voice. Correlation coefficients \((\text{Spearman} \ r_s)\) were then computed between each of the three rankordered measures. The assumption is that the closest correlation would indicate the vocal characteristic that is making the greatest contribution to the perception of \(M-F\) voice quality. If the two characteristics under study were equally prominent this should be reflected by a similarity in the correlation coefficients.

Results - Experiment I

The results of the various rank-order correlations are summarized in Table II-A-I. These same results are also shown in Figs. II-A-2, II-A-3, and II-A-4. In these figures the ranking of "1" represents the lowest \(F_0\), the lowest VTR figure, and the Most Male sounding voice, respectively.
RANKINGS OF THE RATINGS OF MALE-FEMALE VOICE QUALITY

Fig. II-A-2. Rankings of listener ratings of degree of male-female voice quality compared with rankings of $F_o$. 
RANKINGS OF THE RATINGS OF MALE–FEMALE VOICE QUALITY

Fig. II-A.3. Rankings of individual vocal tract resonances compared with rankings of listener ratings of degree of male–female voice quality.
RANKINGS OF VOCAL TRACT RESONANCES

Fig. II-A-4. Rankings of individual vocal tract resonances compared with rankings of $F_0$. 
TABLE II-A-I. Spearman rank-order correlation coefficients ($r_s$) between degree of Male/Female voice quality, fundamental frequency ($F_0$), and vocal tract resonances (VTR).

<table>
<thead>
<tr>
<th>Comparison</th>
<th>Combined Males and Females</th>
<th>$r_s$ Males only</th>
<th>$r_s$ Females only</th>
</tr>
</thead>
<tbody>
<tr>
<td>M-F Voice Quality with $F_0$</td>
<td>.94*</td>
<td>.65*</td>
<td>.88*</td>
</tr>
<tr>
<td>M-F Voice Quality with VTR</td>
<td>.59*</td>
<td>.00+</td>
<td>.27+</td>
</tr>
<tr>
<td>VTR with $F_0$</td>
<td>.56*</td>
<td>.14+</td>
<td>.17+</td>
</tr>
</tbody>
</table>

* p > .01  
+ p N.S.

It can readily be seen that the listeners were basing their judgments of the degree of maleness or femaleness in the voice on the frequency of the laryngeal fundamental. The correlation coefficient of .94 represents an almost perfect one-to-one correspondence between the rankings of how male or female sounding a person's voice was judged to be and the frequency of his laryngeal fundamental. A significant correlation of .59 was also observed between VTR and M-F voice quality. This is interpreted as being simply an expression of the relationship between VTR and $F_0$ which showed a similar correlation of .56.

When the correlations are compared separately for the two sexes, however, somewhat different results are seen. A significant, though somewhat reduced, correlation is again seen between the rankings of $F_0$ and M-F voice quality while the rankings of the other two measures show no correlation. That is to say, rankings of $F_0$ were unrelated to those of VTRs, and VTR rankings were unrelated to those of M-F voice quality within males and females as sub-groups. This is a reflection of the dichotomous nature of the two sexes in their rankings of $F_0$ and M-F voice quality. This dichotomy would contribute to the relationship between the rankings of these measures when both sexes are considered as a single group but would not when they are considered as two subgroups. As can be seen in Figs. II-A-2, II-A-3, and II-A-4 there is no overlap between males and females in their rankings of $F_0$ and M-F voice quality.
Listener judgments - Experiment II

In this experiment voices which consisted of a male characteristic in combination with a female characteristic were presented to a group of listeners who were asked to determine the sex of each speaker. As a control, the listeners were also asked to make a similar judgment for voices containing vocal characteristics consistent for one sex.

From forty speakers used in Experiment I, the five females with the highest vocal tract resonances and the five males with the lowest vocal tract resonances were selected for use in this experiment. A listening tape was prepared which consisted of five-second segments of each of the ten speakers articulating the tone produced by a laryngeal vibrator at each of the two pitches: 240 Hz and 120 Hz. The listening tape consisted, therefore, of an equal number of each of the following combinations: low VTR with low $F_0$, high VTR with high $F_0$, and high VTR with low $F_0$. Half of the samples, therefore, contained two vocal cues consistent with one sex and half contained two contrasting vocal cues.

This 20 item series was presented to a listening group composed of 25 young adults. Listeners were asked simply to determine whether each speaker was a male or a female. Since speech produced with a laryngeal vibrator has an unusual quality the group was allowed to listen to the entire series and their judgments were based on a second playing of the tape which followed immediately.

In those speech samples where the voices contained two contrasting vocal characteristics, listener identifications of the speaker sex would be based on the more prominent of the two characteristics. If the cues were perceptually equal, identifications would be distributed randomly with each sex identified about an equal number of times. In those voices where the two vocal characteristics were consistent, speaker sex identifications should be appropriate to that sex.

Results - Experiment II

The results of Experiment II are summarized in Table II-A-II and illustrated in Figs. II-A-5 and II-A-6. When the two vocal characteristics were consistent for one sex, shown in the left side of the table, the sex of the speaker was correctly identified 245 out of 250 times and it is clear that this was an easy identification for the listeners to make.
Fig. II-A-5. Speaker sex identifications based on voices in which two vocal qualities characteristic of the same sex are combined. \( F_0 \) = Fundamental Frequency; VTR = Vocal Tract Resonance.
Speaker Voices combining a male $F_0$ with a Female VTR
Voices combining a Female $F_0$ with a Male VTR
Listener identifications of speakers as MALE
Listener identifications of speakers as FEMALE

Fig. II-A-6. Speaker identifications based on voices in which a vocal quality characteristic of one sex is combined with a vocal quality characteristic of the other sex. ($F_0 = $ Fundamental Frequency; $VTR = $ Vocal Tract Resonance).
TABLE II-A-II. Distribution of listener identifications of speakers as male or female in response to two vocal characteristic combinations: I - Fundamental frequency ($F_0$) and Vocal Tract Resonance (VTR) representative of the same sex: II - A vocal quality characteristic of one sex combined with a vocal quality characteristic of the other sex. Speakers 1 - 5 are female, 6 - 10 are male.

<table>
<thead>
<tr>
<th>Speaker</th>
<th>Vocal Characteristic:</th>
<th>Times each Speaker perceived as:</th>
<th>Vocal Characteristic:</th>
<th>Times each Speaker perceived as:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$F_0$</td>
<td>VTR</td>
<td>Male</td>
<td>Female</td>
</tr>
<tr>
<td>1.</td>
<td>Female</td>
<td>Female</td>
<td>2</td>
<td>23</td>
</tr>
<tr>
<td>2.</td>
<td>&quot;</td>
<td>&quot;</td>
<td>0</td>
<td>25</td>
</tr>
<tr>
<td>3.</td>
<td>&quot;</td>
<td>&quot;</td>
<td>1</td>
<td>24</td>
</tr>
<tr>
<td>4.</td>
<td>&quot;</td>
<td>&quot;</td>
<td>1</td>
<td>24</td>
</tr>
<tr>
<td>5.</td>
<td>&quot;</td>
<td>&quot;</td>
<td>1</td>
<td>24</td>
</tr>
<tr>
<td>Total</td>
<td>5</td>
<td>120</td>
<td>88</td>
<td>37</td>
</tr>
<tr>
<td>6.</td>
<td>Male</td>
<td>Male</td>
<td>25</td>
<td>0</td>
</tr>
<tr>
<td>7.</td>
<td>&quot;</td>
<td>&quot;</td>
<td>25</td>
<td>0</td>
</tr>
<tr>
<td>8.</td>
<td>&quot;</td>
<td>&quot;</td>
<td>25</td>
<td>0</td>
</tr>
<tr>
<td>9.</td>
<td>&quot;</td>
<td>&quot;</td>
<td>25</td>
<td>0</td>
</tr>
<tr>
<td>10.</td>
<td>&quot;</td>
<td>&quot;</td>
<td>25</td>
<td>0</td>
</tr>
<tr>
<td>Total</td>
<td>125</td>
<td>0</td>
<td>84</td>
<td>41</td>
</tr>
</tbody>
</table>

When the two vocal characteristics were contrasted within one voice, shown on the right side of the table, different results are seen. It should be kept in mind in evaluating the distribution of these judgments that there are no "correct" or "incorrect" identifications since, at this stage, both vocal characteristics are presumed to be of equal perceptual prominence. Rather, a choice as to whether a speaker is male or female indicates simply that the listener's perception was based on one vocal characteristic or the other and not whether his judgment was right or wrong.

It is clear from the distribution of identifications in response to the contrasted cues that both male characteristics are perceptually more prominent than their female counterpart. Speakers were identified as
females 78 times and as males 172 times indicating that listeners based their identification on the male voice characteristic at a better than two to one rate. Both male characteristics appear to be about equally dominant over the female vocal characteristic in the same voice. Of the 172 male identifications 88 were based on a low $F_0$ and 84 on a low VTR characteristic. There is no question, however, that the presence of female vocal characteristic reduces the perceptual prominence of the male characteristic. This is apparent in the reduction of male identifications from 100% when two male characteristics were combined to 69% when a male and a female characteristic were combined.

**Discussion**

The results of these two experiments considered together appear to be somewhat inconsistent. In the experiment using live, unaltered voices the frequency of the speaker's $F_0$ was found to be the primary determiner of how male or female a voice sounded. The presence of vocal tract resonances that were more characteristic of the opposite sex in these voices did not influence the judges' estimates of degree of maleness or femaleness and it is clear that in natural speech, a perception of vocal pitch is a product of the frequency of the $F_0$.

In the second experiment, on the other hand, speakers were more likely to be judged as male even in the presence of a female-like $F_0$ and it is apparent in these cases that the male vocal tract resonance characteristic had considerable perceptual importance. The male $F_0$, however, retained its predominance over the female vocal tract resonance characteristic which would be consistent with the findings of Experiment I.

The somewhat different results of the two experiments may be a result of the use of the laryngeal vibrator as a vocal sound source in Experiment II. It may be easier to produce a more natural sounding male than female fundamental with the particular laryngeal vibrator used in this study. This could account for the perceptual weakness of the female $F_0$ in the second experiment that was not seen in the first. It is also possible that the glottal source in females differs from males in some basic way besides simply that of pitch. This needs to be examined more closely since it may be desirable to produce female sounding speech synthetically that can be easily recognized as such either by humans or by automatic speech recognition.
devices. If a good approximation of the natural female glottal tone is not provided, speaker sex recognition may be confounded by the frequency of the vocal tract resonances with which it is combined.

Conclusions

In natural speech, the degree of male or female quality in the voice is a function of the frequency of the laryngeal fundamental. Individual vocal tract resonance characteristics, whether male or female, contribute little or nothing to the perception of this vocal quality. When a laryngeal vibrator is substituted for the normal glottal tone, however, the female $F_0$ is perceptually weaker than male vocal tract resonance characteristics, while the male $F_0$ retains its perceptual prominence.

Acknowledgments

This study was supported by a grant from the Selling Research Foundation of the University of Oregon Medical School and by Dept. of H.E.W. Grant 72-260-5089. A portion of this research was carried out while the author was a guest researcher at the Dept. of Speech Communication, Speech Transmission Laboratory of the Royal Institute of Technology (KTH), Stockholm. The valuable suggestions provided by many of the STL staff are gratefully acknowledged. Appreciation is also extended to those persons who served as willing subjects and listeners in the two experiments.

References:


