Acoustic characteristics of less-masculine-sounding male speech

Jack D. Avery
Speech Pathology Section, Veterans Administration Medical Center, Minneapolis, Minnesota 55417

Julie M. Liss
Department of Speech and Hearing Science, Arizona State University, Tempe, Arizona 85287-0102

(Received 5 July 1995; accepted for publication 7 February 1996)

This study compared samples of less-masculine-sounding (LMS) and more-masculine-sounding (MMS) male speech to identify acoustic characteristics, other than fundamental frequency, that might contribute to the perception of these categories. In the first phase, audiolocated speech samples provided by 35 males were presented to 35 female listeners in a paired-comparison perceptual experiment. Nineteen of the male speech samples were judged reliably to fall within the LMS or MMS categories. Within those 19 samples, 8 speakers (4 LMS and 4 MMS) exhibited similar distributions of habitual fundamental frequency values in connected speech and in sustained phonation. In the second phase of the experiment, various acoustic measures of these eight connected speech samples were conducted. Significant differences between measures of fundamental frequency contours, vowel formant midpoint values, and in the first, third and fourth spectral moments of two fricatives were revealed. These findings may be useful in creating stylized synthetic speech that varies on the dimension of masculinity, and they may have clinical relevance for patients wishing to modify the perception of masculinity invoked by their speech. © 1996 Acoustical Society of America.

PACS numbers: 43.70.Fq, 43.70.Gr, 43.71.Bp

INTRODUCTION

The information available in a speech signal allows listeners to make inferences about a speaker’s age, socioeconomic status, educational background, and a multitude of other personal attributes. The most salient of these characteristics is gender (Sachs, 1975). It is believed that listeners use the psychoacoustic dimensions of vocal pitch, loudness, and quality to make the dichotomous distinction of speaker sex (Oates and Dacakis, 1983; Wolfe et al., 1990). But there are regions along the continua of these dimensions that challenge culture’s expectations of male and female speech behaviors. The gender of people whose speech behaviors occupy these regions may be misidentified. Alternately, gender identification may be accurate, but the listener attributes qualities of the opposite gender to the speaker. An example is the perceptual phenomenon of male speech that is judged to be ‘‘effeminate’’ or ‘‘feminine-sounding.’’ 1

Although American culture recognizes the existence of less-masculine-sounding (LMS) male speech, the acoustic correlates of this perception are virtually unknown. Studies from related areas, such as female, transgender, and clear speech, offer clues regarding the phenomenon of LMS speech. These include aspects of vocal pitch, prosodic contours, articulatory characteristics, and speaking rate.

Fundamental frequency (F0) seems the most obvious explanation for the perception of the LMS/MMS male speech. There is substantial evidence, however, that elevated habitual fundamental frequencies are not sufficient to render male speech less-masculine-sounding, or even to distinguish male from female speech (Childers and Wu, 1991; Wu and Childers, 1991). For example, Terango (1966) found that ‘‘effeminate’’ (presumably LMS according to our terminology) men were not higher in median pitch than unselected American male values culled from other studies of male fundamental frequency. Lerman and Damste (1969) reported similar results for 13 homosexual (presumably analogous to LMS) and 13 heterosexual (presumably analogous to MMS) male speakers. They found no differences in average F0 between the two groups in any of the connected speech samples; the mean F0 for the homosexual males’ reading passage actually was lower than the mean F0 of the heterosexual males for the same passage. Other related studies have shown that male-to-female transgender individuals cannot achieve desired feminine identification merely by raising pitch (Günzburger, 1989, 1993; Mount and Salmon, 1988; Wolfe et al., 1990).

Although elevated habitual F0 may not account fully for the percept of LMS speech, gender-specific differences in F0 variations over larger segments of speech have been demonstrated. Brend (1975) associated diverse and variable changes in F0 with the speech of women. This is supported by research which shows that male-to-female transsexuals use wide and relatively rapid changes in F0 to achieve feminine-sounding speech (Günzburger, 1989, 1993; Wolfe et al., 1990). Similarly, Terango’s (1966) ‘‘effeminate’’ speakers exhibited a higher mean rate of pitch change for upward and downward inflections than the ‘‘masculine’’ speakers.

Filter characteristics of the vocal tract, as evidenced by vowel formant frequencies and consonant spectra, also may contribute to the perception of LMS speech. Men (with anatomically larger vocal tracts) tend to have lower formant frequencies than women (with smaller vocal tracts) (Childers and Wu, 1991; Coleman, 1971; Hillenbrand et al., 1995; Peterson and Barney, 1952). Likewise, speaker sex (and presumably vocal tract size) influences the height of center fre-
quencies of certain burst and fricative spectra (Ingemann, 1968; Nittrouer, 1995; Schwartz, 1968; Wu and Childers, 1991). However filter characteristics can be stylistically modified as well (Mattingly, 1966), and these modifications seem to follow culturally determined dictates for males and females (Sachs, 1975).

For example, Oates and Dacakis (1983) declared that females use “clear enunciations” (p. 146), and “more correct, standard speech sounds” (p. 144) than males. McConnell-Ginet (1983) stated that “femininity is conveyed by ... pronunciation” (p. 75). Guidelines for exploring these perceptual descriptions of pronunciation are found in the acoustic analysis of clear speech. Picheny et al. (1986) described clear speech as being slower (longer pauses between words and longer segment durations), exhibiting less modified or reduced vowel and consonant segments, and exhibiting greater intensity of obstruent sounds than conversational speech. They also found that production of the consonants /t/ and /s/ in clear speech is marked by an increase in the frequency of the maximum spectral peak of energy distribution.

Predictions regarding speaking rate differences of LMS and MMS speakers are not straightforward. If female speech is indeed characterized by hyperarticulation, one would anticipate slower articulatory rates than are found in male speech due to longer phoneme durations. However, Oates and Dacakis (1983) described feminine speech as sounding “smooth and fast” (p. 146). It is not known whether this perception derives from decreases in interword pause time or from decreases in phoneme duration. It is conceivable that decreases in interword pause time, coupled with increases in phoneme durations, may give the perceptual impression of smooth and fast speech.

The purpose of this study was to explore potential acoustic differences in the speech production patterns of two groups of men (LMS versus MMS as judged by a group of listeners). The goal was to identify two groups of men whose distributions of habitual fundamental frequencies in connected speech were identical. We then examined:

1. Prosodic contours as marked by changes in F0 values within and between words;
2. Vowel formant midpoint values (F1 and F2) for the vowels /i/, /æ/, /æ/, /u/;
3. Three of the first spectral moments of the fricatives /s/ and /ʃ/; and
4. Speech rate as measured by the ratio of articulation to interword pause time.

I. METHOD

A. Speech samples

1. Speakers

Thirty-seven men (mean age 39.9 years; range 33–50 years) from the University of Minnesota and the Minneapolis–St. Paul metropolitan area volunteered to provide speech samples for this study. The men were told only that the study focused on perceptual and acoustic analysis of male speech patterns.

Prior to speech collection, each man was screened informally to identify functional articulatory errors, stuttering, non-Midwestern accents, and previously diagnosed hearing loss. Based on these exclusion criteria, 35 men provided the corpus of speech samples for this study.

2. Recording procedures

Speech samples were audio recorded in a sound isolated booth using a Tascam 112 tape recorder, an Electro-Voice BK-1 microphone placed 8–10 in. from the speaker’s mouth, and a Shure FP11 microphone-to-line amplifier. Speakers were instructed to speak as conversationally as possible, and to use a typical level of speaking loudness, pitch, rate, and tone. The speech sample consisted of a modified version of the “Grandfather Passage” (see Table I; the passage was modified to contain more instances of the targeted fricatives /s/ and /ʃ/), the sustained vowels /i/, /æ/, /æ/, and /u/ in isolation, and as imbedded in the CVC syllables “ heed,” “ had,” “ hot,” and “ who’d.”

3. Stimulus tape

Selected portions of the audiotaped speech samples (the words presented in capital letters in Table I) were digitized for the acoustic analysis; words in bold print were the target of F0 measurements; measured vowels are underlined; measured fricatives are underlined and in italics.

You wish to know all about my GRANDFATHER. Well, he is nearly 93 years old, yet he still thinks as swiftly as ever. HE DRESSES HIMSELF IN A SHINY, OLD, BLACK, FROCK COAT, USUALLY SEVERAL BUTTONS MISSING. HIS THIN, SHORT-SLEEVED SHIRT OFFERS HIM LITTLE WARMTH. A LONG, SHAGGY BEARD CLINGS TO HIS UNSHAVED, SHARP CHIN. GIVING THOSE WHO OBSERVE HIM A PRONOUNCED FEELING OF THE UTMOST RESPECT. WHEN HE SPEAKS, HIS VOICE QUIVERS A BIT. Twice each day he plays skittishly and with zest upon a small pump organ. Except in the winter when the snow or ice prevents, he slowly takes a short walk in the open air each day. We have often urged him to walk more and smoke less, but he always answers, “ ‘banana oil!’” Grandfather likes to be MODERN in his language.

Prior to speech collection, each man was screened informally to identify functional articulatory errors, stuttering, non-Midwestern accents, and previously diagnosed hearing loss. Based on these exclusion criteria, 35 men provided the corpus of speech samples for this study.

2. Recording procedures

Speech samples were audio recorded in a sound isolated booth using a Tascam 112 tape recorder, an Electro-Voice BK-1 microphone placed 8–10 in. from the speaker’s mouth, and a Shure FP11 microphone-to-line amplifier. Speakers were instructed to speak as conversationally as possible, and to use a typical level of speaking loudness, pitch, rate, and tone. The speech sample consisted of a modified version of the “Grandfather Passage” (see Table I; the passage was modified to contain more instances of the targeted fricatives /s/ and /ʃ/), the sustained vowels /i/, /æ/, /æ/, and /u/ in isolation, and as imbedded in the CVC syllables “ heed,” “ had,” “ hot,” and “ who’d.”

3. Stimulus tape

Selected portions of the audiotaped speech samples (the words presented in capital letters in Table I) were digitized for the acoustic analysis; words in bold print were the target of F0 measurements; measured vowels are underlined; measured fricatives are underlined and in italics.

You wish to know all about my GRANDFATHER. Well, he is nearly 93 years old, yet he still thinks as swiftly as ever. HE DRESSES HIMSELF IN A SHINY, OLD, BLACK, FROCK COAT, USUALLY SEVERAL BUTTONS MISSING. HIS THIN, SHORT-SLEEVED SHIRT OFFERS HIM LITTLE WARMTH. A LONG, SHAGGY BEARD CLINGS TO HIS UNSHAVED, SHARP CHIN. GIVING THOSE WHO OBSERVE HIM A PRONOUNCED FEELING OF THE UTMOST RESPECT. WHEN HE SPEAKS, HIS VOICE QUIVERS A BIT. Twice each day he plays skittishly and with zest upon a small pump organ. Except in the winter when the snow or ice prevents, he slowly takes a short walk in the open air each day. We have often urged him to walk more and smoke less, but he always answers, “ ‘banana oil!’” Grandfather likes to be MODERN in his language.

Prior to speech collection, each man was screened informally to identify functional articulatory errors, stuttering, non-Midwestern accents, and previously diagnosed hearing loss. Based on these exclusion criteria, 35 men provided the corpus of speech samples for this study.
groups: one group sounding “less masculine,” and one group sounding “more masculine” than the modulus. It was presumed that high levels of agreement among the listeners would identify speakers who differed on the variable of “masculine-sounding,” and who fell closer to the poles of the LMS-MMS continuum for this group of men. Speakers approximately as-masculine-sounding as the modulus would generate listener agreement levels near 50%. For ease of explanation, identification, and reference, the modulus was arbitrarily named “Bob.”

The paired-comparison perceptual task was created by dubbing the third through sixth sentence of each of the remaining 34 speakers’ reading passages onto a composite stimulus tape. Additionally, 16 passages were randomly selected from the 34, renumbered, and added to the tape after the original 34 passages to assess inter- and intrarater reliability. Each of the 50 speech passages (identified as speaker 1 through speaker 50) was preceded by a dubbing of “Bob” reading the same sentences.

### B. Perceptual analysis

#### 1. Subjects

Thirty-five women (mean age 27.7 years; range 19–48 years) from the University of Minnesota served as listeners for this study. Single-sex listeners were used to minimize a possible gender bias in the judgement of the samples (Oates and Dacakis, 1983; Traunmüller and Eriksson, 1995). All subjects reported normal hearing and were compensated for participation in the study.

#### 2. Perceptual task

Subjects sat in individual cubicles and listened under headphones to the stimulus tape. Subjects adjusted their volume controls to their most comfortable listening level and were instructed not to change the setting after the experiment had commenced. They then heard “Bob” read the four-sentence passage and were instructed to use Bob as the sole criterion for rating other speakers as sounding “less masculine” or “more masculine” than he. Subjects were instructed to make their decisions quickly and to rely on first impressions; they were not directed to listen to any specific voice or speech parameters.

Subjects heard the 50 speech samples with each sample preceded by Bob reading the same sample. After the first ten sets (Bob and speaker 1, Bob and speaker 2, etc.) had been completed, a smaller portion of the Bob passage was presented each time to decrease the overall duration and tediousness of the task. The listening task took approximately 1 h to complete.

Results were tallied and percentage of perceptual agreement (percentage of listeners assigning LMS or MMS label) was calculated for each speaker. Speakers who elicited 80% or higher listener agreement were categorized in this first cut as LMS or MMS. Speakers who elicited less than 80% listener agreement were judged to fall closer to the center of the distribution for this group of men and these speech samples were not subjected to additional analysis.

### 3. Listener reliability

Intra- and interjudge reliability were assessed by examining the original and repeat label assignments for the 16 samples. We expected that speakers who occupied the polar regions of the LMS and MMS continuum would elicit higher repeat agreement than those whose speech was only negligibly different from that of the modulus. This was the case for both intra- and interjudge agreement. In the initial presentation of the 16 reliability samples, ten were judged by 80% or more of the listeners as either more or less masculine-sounding than the modulus on the first assessment; eight of these ten met or exceeded 80% agreement on the repeated samples. Intrajudge reliability for these eight samples that elicited high interjudge agreement was acceptable: Eighty to 100% of the listeners agreed with themselves for seven of the eight samples.

### C. Acoustic analysis

Mean fundamental frequencies for the 19 speakers identified as LMS or MMS in the perceptual phase were computed by averaging the peak $F_0$ values of four voiced segments contained in the first sentence of each speech sample (“dress,” “self,” “old,” and “black”). This mean was selected as a gross index of habitual fundamental frequency in connected speech. The cursor function of CSpeech was used to isolate the vocalic region of each of the four syllables on a waveform display and the automatic pitch extraction function generated an $F_0$ trace across this time window. The validity of this measure as a gross index of habitual fundamental frequency is supported by a subsequent analysis which revealed similar average $F_0$ values for the speakers’ sustained /a/ productions.

With this analysis, the distributions of average peak $F_0$’s in connected speech for the two groups were assessed, and individual speakers from both groups who presented similar $F_0$ values were identified. All subsequent acoustic analysis, including measures of aspects of prosody, vowel production, sibilant production, and speech rate, were conducted on a subset of 8 of the 19 speakers.

Prosodic contour information was quantified by tracking $F_0$ movement through each speaking passage. Fundamental frequency tracings, accomplished by an autocorrelation pitch extraction algorithm (CSpeech), were acquired for each four-sentence block (see capitalized sentences in Table I). The original goal was to measure $F_0$ movement within and between all syllables of each speaker’s four-sentence passage, however it was found that the brief durations and extensive articulatory reductions of some syllables rendered certain segments unmeasurable. Therefore, 29 segments that could be reliably measured were analyzed (the segments from which these measures were taken are bolded in Table I). For each measurement, the segment of interest was operationally defined as the vowel nucleus of a syllable and any adjacent voiced phonemes that could not be readily separated from the vowel. Seventeen of the 29 segments came from single-syllable words, five were the stressed syllables of bisyllabic words, and the remaining nine segments consisted of one or more syllables of multisyllabic words.
The speech material, targeted segments, and measurement criteria were intended to be identical across speakers so that comparisons would reveal any actual differences in $F_0$ movement between the two groups. Intonations were defined as $F_0$ movement within a given vocalic nucleus, and shifts as $F_0$ movement between the end of a vocalic nucleus of one segment and the beginning of a vocalic nucleus of a subsequent segment (Fairbanks, 1940). The boundaries of the vocalic nucleus were identified as the first and last visible glottal pulses extending through at least the second formant.

Intonation values were calculated as the frequency difference between the highest and lowest discernible $F_0$ readings within a given vocalic nucleus. Depending on the sequencing of the minimum and maximum values, intonations were rising (minimum preceding maximum value), falling (maximum preceding minimum value) or flat (less than 5-Hz difference between minimum and maximum values). Because the autocorrelation pitch extraction method is prone to error near the onset of voicing (Milenkovic and Read, 1992), voiced segments less than 50 ms in duration were not measured. In the case of a flat $F_0$ track, (defined as several movements of a cursor reflecting the same frequency reading), cursors were placed to reflect the shortest time difference between the minimum and maximum reading to calculate the steepest intonation slope.

Intonation measures from the targeted voiced segments were used to calculate:

1. percentage of nonmeasurable, flat, upward, and downward intonations;
2. mean extent (in Hz) and mean slope (extent/duration) of upward intonations;
3. mean extent (in Hz) and mean slope (extent/duration) of downward intonations.

Shift values were calculated as the difference in $F_0$ between the last frequency measure of one nucleus and the first frequency measure of the subsequent nucleus. As with intonations, shifts were labelled as rising, falling, or flat, and durations between frequency measures were used for the calculations of the slope of each shift. The following values were tabulated:

1. percentage of flat, upward, and downward shifts;
2. mean extent (Hz) and slope (extent/duration) of upward shifts;
3. mean extent (Hz) and slope (extent/duration) of downward shifts.

First and second formant frequencies were measured for the four point vowels (/i/, /æ/, /a/, and /u/) in each speaker’s connected speech samples (see underlined vowels in Table I), in the hVd context, and in sustained vowel productions. The point vowels were selected as a general index of articulatory working space. The CSpeech formant tracking function on digital spectrographic displays was used in conjunction with linear predictive coding and Fourier analysis to accurately identify the formants. A 20-ms interval at the temporal middle of the vowel was isolated for formant measurement.

In the connected speech sample, words containing three productions of the four target vowels were analyzed (Table I, ‘‘grandfather, black, frock, usually, sleeved, shaggy, to, who, feeling, speaks, modern’’. Formant values were measured at an operationally defined 20-ms interval specific for each target word. Mean formant values were then calculated to reflect the speaker’s average $F_1$ and $F_2$ values for each of the four vowels.

Spectral moments analysis, specifically the statistical measures of mean, skewness and kurtosis have been used successfully to investigate voiceless obstruents (Forrest et al., 1988), and developmental characteristics of fricative production (Nittrouer, 1995; Nittrouer et al., 1989). Nittrouer’s (1995) ancillary findings included gender differences in the production of /s/ and /ʃ/, in which mean center frequencies were higher for women than men. In the present study, spectral characteristics of the sibilants /s/ (self, missing, sleeved, respect, speaks), and /ʃ/ (shiny, short, shirt, shaggy, sharp) were examined (Table I). Moments analysis was conducted for a 50-ms window at the temporal midpoint of the frication noise. The 50-ms window produced five readings (fast Fourier transforms calculated every 10 ms using a 20-ms Hamming window) whose averages were taken to be representative of the sibilants’ spectral characteristics. The measures of interest included the mean, skewness, and kurtosis of each distribution [see Forrest et al. (1988), Nittrouer et al. (1989), and Nittrouer (1995) for detailed descriptions of the computation of spectral moments by CSpeech].

The duration of each speaker’s four-sentence passage and all pause times were measured. Pause times were defined as any interword silent segment, and therefore included closure durations preceding word-initial stop consonants. No intraword pauses were produced by these speakers. The summed pause times were subtracted from the duration of the entire passage to estimate articulation time.

### D. Reliability

All measurements were performed by the first author, following the establishment of acceptable interjudge reliability with the second author. A minimum of 20% of all measures were repeated to calculate intrajudge reliability. In each case, absolute difference values between the original and the remeasurement were evaluated.

For intonation, 95.3% of the remeasurements differed by 5 Hz or less from the original; 79.1% of the these were identical. For shifts, 95.0% of the remeasurement discrepancies were equal to or less than 5 Hz; 71.7% of these comparisons were identical. Levels of reliability for all measures of prosodic contours were regarded as highly acceptable.

One remeasurement of each of the four vowels in each speaking condition was made for each speaker. In connected speech, 96.9% of the $F_1$ values fell within 45 Hz, and 90.6% of the $F_2$ values fell within 60 Hz. For CVC productions, 88% of the $F_1$ values and 100% of the $F_2$ values fell within 40 and 60 Hz, respectively. For sustained phonations, 100% of the $F_1$ and 87.5% of the $F_2$ frequencies fell within 40 and 60 Hz, respectively. The magnitude of these discrepancies are within the expected range for these types of measurements (Liss and Weismer, 1994).
TABLE II. Speakers who elicited at least 80% labeling agreement among the listeners and their corresponding average fundamental frequencies in connected speech.

<table>
<thead>
<tr>
<th>LMS#</th>
<th>% Agreement</th>
<th>Mean F0</th>
<th>MMS#</th>
<th>% Agreement</th>
<th>Mean F0</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>85.7</td>
<td>96.6</td>
<td>1</td>
<td>80.0</td>
<td>100.8</td>
</tr>
<tr>
<td>9</td>
<td>85.7</td>
<td>99.0</td>
<td>5</td>
<td>85.7</td>
<td>87.5</td>
</tr>
<tr>
<td>11</td>
<td>100.0</td>
<td>111.9</td>
<td>7</td>
<td>91.4</td>
<td>102.0</td>
</tr>
<tr>
<td>14</td>
<td>88.6</td>
<td>100.8</td>
<td>8</td>
<td>82.9</td>
<td>133.0</td>
</tr>
<tr>
<td>15</td>
<td>85.7</td>
<td>99.0</td>
<td>13</td>
<td>91.4</td>
<td>93.0</td>
</tr>
<tr>
<td>16</td>
<td>100.0</td>
<td>117.4</td>
<td>17</td>
<td>85.7</td>
<td>122.8</td>
</tr>
<tr>
<td>24</td>
<td>88.6</td>
<td>112.3</td>
<td>21</td>
<td>100.0</td>
<td>84.1</td>
</tr>
<tr>
<td>26</td>
<td>100.0</td>
<td>109.0</td>
<td>30</td>
<td>80.0</td>
<td>102.6</td>
</tr>
<tr>
<td>29</td>
<td>91.4</td>
<td>89.1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>33</td>
<td>100.0</td>
<td>134.4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>34</td>
<td>100.0</td>
<td>103.1</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

E. Analysis

Descriptive and inferential statistics were used to analyze the acoustic data. The complete data set for the eight speakers consisted of 104 vowel F1 and F2 measures; approximately 480 fundamental frequency measures for the intonation and shift analysis; 80 fricative spectra for moments analysis; 45 peak fundamental frequencies for estimates of habitual pitch (N=19 speakers); and approximately 300 pause duration measures for the rate analysis. Means and standard deviations were computed for each group and for all eight speakers (grand mean) on the measures of vowel formant frequencies in connected and /hVd/ environments, intonations and shifts (percent rising, extent rising, percent falling extent falling, and percent flat), habitual fundamental frequencies, and articulation-to-pause time ratios. The t tests and nonparametric statistics were conducted to identify group differences (p<0.05) in characteristics of vowel and fricative production, and in measures of prosody and speaking rate.

II. RESULTS

A. Perceptual analysis

Calculation of listener agreement revealed that 19 speakers (11 LMS and 8 MMS) met or exceeded the 80% criterion (Table II). Speech samples that did not elicit this level of agreement were not subjected to further analysis.

B. Acoustic analysis

1. Mean habitual fundamental frequency

The 19 speech samples were analyzed for average fundamental frequency. LMS speaker values ranged from 89.1 to 134.3 Hz, and MMS values ranged from 84.1 to 133.0 Hz.

Results are summarized in Table II. A t test revealed that the mean F0 of the LMS group (106.6 Hz, S.D. 12.34) was not statistically different from the mean F0 of the MMS group (103.23, s.d. 16.89).

Recall that the goal of the perceptual phase was to identify two groups of speakers who elicited high interjudge correspondence and whose distributions of habitual F0 did not differ. Examination of the mean F0 of the 19 speech samples revealed a near-identical match of four LMS and four MMS speech passages (see Table III). This resulted in highly similar habitual fundamental frequency distributions for the two groups. Comparisons of the values from connected speech with those F0 values from the sustained phonation samples confirmed the similarity of the groups on this dimension.

Because these eight speech samples met our criteria, they were used for all subsequent acoustic comparisons. As added evidence of the validity of the perceptual results, listeners made these particular eight judgements with exceptionally high inter-and intrarater reliability. For ease of data presentation and discussion, these eight passages will be referred to as sequential numbers of 1 to 8 (LMS 1–4 and MMS 5–8).

2. Prosody

Intonation (F0 movements within the defined segments) results for the eight speakers are summarized in Table IV. Percent of rising intonations for each speaker ranged from 17.9% to 45.2%. Mean extent of excursion of rising intonations ranged from 10.73 to 25.69 Hz, and mean rising slope (extent/duration) for each speaker was 0.140 to 0.264. Percent of falling intonations from 19.4% to 71.4%. Mean extent of excursion of falling intonations was 9.33 to 24.99 Hz, and mean falling slope ranged from 0.120 to 0.223. Statistical comparisons of the group means revealed only that the MMS speakers had a significantly greater extent of downward intonations than the LMS speakers (Mann–Whitney rank sum test: T=2434, p=0.046).

Shift (F0 movement between target segments) results for each of the eight speakers are summarized in Table V. Percent of rising shifts for each speaker ranged from 20.0% to 74.1%. Mean extent of excursion of rising shifts was 14.64 to 26.7 Hz, and mean rising slope for each speaker ranged from 0.044 to 0.104. Percent of falling shifts ranged from 11.1% to 22.67%. Mean extent of excursion of falling shifts ranged from 9.53 to 22.67 Hz, and mean falling slope ranged from 0.028 to 0.099. Group comparisons revealed significant differences between the mean extent of downward shifts between the LMS and MMS groups (Mann–Whitney...
TABLE IV. Intonation summary—individual and group mean values are presented. The table contains the number of measures taken for each subject, the percentage of intonation values that were rising, the mean extent of the rise, the mean slope of the rise, and the analogous information for the intonation values that were falling. Standard deviations are presented in parentheses. The percentage of intonation values that did not exceed differences of 5 Hz are in the last column. *Significant difference per Mann–Whitney rank sum test (p<0.05).

<table>
<thead>
<tr>
<th></th>
<th>n</th>
<th>% Rising</th>
<th>Extent rising</th>
<th>Rising slope</th>
<th>% Falling</th>
<th>Extent falling</th>
<th>Falling slope</th>
<th>% Flat</th>
</tr>
</thead>
<tbody>
<tr>
<td>LMS  1</td>
<td>29</td>
<td>37.9</td>
<td>13.2 (7.2)</td>
<td>0.180 (0.115)</td>
<td>55.2</td>
<td>16.4 (13.9)</td>
<td>0.158 (0.101)</td>
<td>6.9</td>
</tr>
<tr>
<td>LMS  2</td>
<td>29</td>
<td>27.5</td>
<td>17.1 (9.5)</td>
<td>0.167 (0.121)</td>
<td>34.5</td>
<td>12.7 (3.4)</td>
<td>0.194 (0.095)</td>
<td>24.1</td>
</tr>
<tr>
<td>LMS  3</td>
<td>31</td>
<td>25.8</td>
<td>25.9 (14.2)</td>
<td>0.222 (0.112)</td>
<td>61.3</td>
<td>19.3 (11.6)</td>
<td>0.198 (0.110)</td>
<td>9.7</td>
</tr>
<tr>
<td>LMS  4</td>
<td>31</td>
<td>45.2</td>
<td>12.6 (8.3)</td>
<td>0.162 (0.076)</td>
<td>19.4</td>
<td>9.3 (3.4)</td>
<td>0.145 (0.067)</td>
<td>29.0</td>
</tr>
<tr>
<td>Mean</td>
<td>30</td>
<td>34.2</td>
<td>16.2 (10.6)</td>
<td>0.180 (0.102)</td>
<td>42.5</td>
<td>15.9*(11.1)</td>
<td>0.178 (0.1)</td>
<td>17.5</td>
</tr>
<tr>
<td>MMS  5</td>
<td>30</td>
<td>36.7</td>
<td>10.8 (4.2)</td>
<td>0.264 (0.212)</td>
<td>40.0</td>
<td>14.3 (9.9)</td>
<td>0.171 (0.078)</td>
<td>7.1</td>
</tr>
<tr>
<td>MMS  6</td>
<td>28</td>
<td>37.9</td>
<td>10.7 (5.4)</td>
<td>0.140 (0.099)</td>
<td>41.4</td>
<td>15.1 (7.7)</td>
<td>0.120 (0.057)</td>
<td>20.7</td>
</tr>
<tr>
<td>MMS  7</td>
<td>29</td>
<td>22.2</td>
<td>15.7 (9.3)</td>
<td>0.141 (0.037)</td>
<td>44.4</td>
<td>18.7 (7.7)</td>
<td>0.212 (0.128)</td>
<td>25.9</td>
</tr>
<tr>
<td>MMS  8</td>
<td>27</td>
<td>17.9</td>
<td>16.1 (6.8)</td>
<td>0.233 (0.193)</td>
<td>71.4</td>
<td>25.0 (14.6)</td>
<td>0.223 (0.193)</td>
<td>20.7</td>
</tr>
<tr>
<td>Mean</td>
<td>28.5</td>
<td>28.9</td>
<td>12.5 (6.3)</td>
<td>0.174 (0.117)</td>
<td>49.1</td>
<td>19.4* (11.6)</td>
<td>0.190 (0.139)</td>
<td>16.7</td>
</tr>
</tbody>
</table>

TABLE V. Shift summary—Individual and group mean values are presented. The table contains the number of measures taken for each subject, the percentage of shift values that were rising, the mean extent of the rise, the mean slope of the rise, and the analogous information for the shift values that were falling. Standard deviations are presented in parentheses. The percentage of shift values that did not exceed differences of 5 Hz are in the last column. *Significant difference per Mann–Whitney rank sum test (p<0.05).

<table>
<thead>
<tr>
<th></th>
<th>n</th>
<th>% Rising</th>
<th>Extent rising</th>
<th>Rising slope</th>
<th>% Falling</th>
<th>Extent falling</th>
<th>Falling slope</th>
<th>% Flat</th>
</tr>
</thead>
<tbody>
<tr>
<td>LMS  1</td>
<td>28</td>
<td>42.9</td>
<td>19.4 (16.0)</td>
<td>0.028 (0.016)</td>
<td>28.6</td>
<td>13.4 (4.5)</td>
<td>0.050 (0.048)</td>
<td>28.6</td>
</tr>
<tr>
<td>LMS  2</td>
<td>29</td>
<td>37.9</td>
<td>18.5 (11.7)</td>
<td>0.040 (0.034)</td>
<td>41.4</td>
<td>18.7 (9.9)</td>
<td>0.075 (0.037)</td>
<td>20.7</td>
</tr>
<tr>
<td>LMS  3</td>
<td>30</td>
<td>56.7</td>
<td>17.3 (8.2)</td>
<td>0.052 (0.036)</td>
<td>23.3</td>
<td>22.7 (10.2)</td>
<td>0.099 (0.063)</td>
<td>20.0</td>
</tr>
<tr>
<td>LMS  4</td>
<td>30</td>
<td>20.0</td>
<td>26.7 (18.0)</td>
<td>0.035 (0.027)</td>
<td>36.7</td>
<td>19.1 (10.8)</td>
<td>0.070 (0.079)</td>
<td>43.3</td>
</tr>
<tr>
<td>Mean</td>
<td>29.3</td>
<td>39.3</td>
<td>19.6 (12.7)</td>
<td>0.046 (0.043)</td>
<td>32.5</td>
<td>18.6* (9.4)</td>
<td>0.073* (0.058)</td>
<td>28.2</td>
</tr>
<tr>
<td>MMS  5</td>
<td>29</td>
<td>44.8</td>
<td>14.6 (7.4)</td>
<td>0.030 (0.025)</td>
<td>20.7</td>
<td>13.8 (9.4)</td>
<td>0.040 (0.014)</td>
<td>34.5</td>
</tr>
<tr>
<td>MMS  6</td>
<td>28</td>
<td>32.1</td>
<td>18.8 (15.4)</td>
<td>0.050 (0.044)</td>
<td>28.6</td>
<td>11.9 (6.5)</td>
<td>0.028 (0.018)</td>
<td>39.3</td>
</tr>
<tr>
<td>MMS  7</td>
<td>28</td>
<td>42.9</td>
<td>21.5 (17.5)</td>
<td>0.061 (0.071)</td>
<td>28.6</td>
<td>13.3 (11.3)</td>
<td>0.041 (0.034)</td>
<td>28.6</td>
</tr>
<tr>
<td>MMS  8</td>
<td>27</td>
<td>74.1</td>
<td>23.4 (14.8)</td>
<td>0.104 (0.093)</td>
<td>11.1</td>
<td>9.5 (4.1)</td>
<td>0.039 (0.050)</td>
<td>14.8</td>
</tr>
<tr>
<td>Mean</td>
<td>28.0</td>
<td>48.2</td>
<td>20.1 (14.2)</td>
<td>0.068 (0.074)</td>
<td>22.3</td>
<td>12.5* (8.4)</td>
<td>0.037* (0.027)</td>
<td>29.5</td>
</tr>
</tbody>
</table>

rank sum test: T=596, p =0.004), and between group means for the downward slope (Mann–Whitney rank sum test: T=588, p =0.003).

Thus the two groups of speakers, whose average fundamental frequencies in connected speech did not differ, displayed several significant differences in prosodic measures: extent of downward intonations, and extent and slope of downward shifts.

Although the pattern of statistically significant differences is revealing, additional information is gained by examining the data in holistic terms (Young, 1993). For example, given the finite pitch range of humans, one can hypothesize that rising intonations generally are followed by falling shifts (rising-intonation-falling-shift, or RIFS pattern) and that falling intonations generally are followed by rising shifts (falling-intonation-rising-shift, or FIRS pattern). Obviously simplistic in nature, this set of patterns can nonetheless account for the majority of the data for the two groups of speakers of this study. Figure 1 utilizes this model to chart the percent of occurrence, and mean extent and slope of both sets of patterns for both groups of speakers. Flat intonations and shifts obviously complicate this model, yet the occurrence of flat intonations and shifts did not differ between the groups and therefore is not addressed here. Although both F0 patterns were exhibited by the speakers, there was a slight preference for the FIRS pattern (49.1% of MMS intonations were falling and 48.2% of shifts were rising; 42.5% LMS intonations were falling and 39.3% of shifts were rising). The RIFS pattern occurred less frequently for both groups, however the LMS group exhibited a higher percentage of its occurrence than did the MMS group (LMS, 34.2% rising intonation and 32.5% falling shift; MMS, 28.9% rising intonation and 22.3% falling shift).

3. Vowel formants

F1 and F2 values of the four connected speech vowels for each speaker and for the LMS and MMS groups are shown in Fig. 2. As illustrated by the error bars which represent one standard deviation about the mean, individual formant values varied substantially within and between speakers. The majority of F1 and F2 LMS values (with the exception of /u/ F2) were higher than the comparable MMS values. As a result, the LMS quadrilateral is shifted up and to the right of the MS quadrilateral. Despite the visual disparity of the two sets of values, statistical comparisons of the two groups revealed that only the difference between the /i/ F2 values was significant (Mann–Whitney rank sum test: T=80, p =0.023).

First and second formant frequencies of the CVC vowels...
were measured as a reference point determine the amount of difference between those vowels produced in connected speech and in near isolation (hVd context). Mean formant frequencies and standard deviations for the CVC productions are illustrated in Fig. 3. As with the connected speech values, the LMS vowel quadrilateral is displaced upward and to the right of the MMS quadrilateral, and again, statistical comparisons revealed that the difference between the /i/ F2 values for LMS and MMS speakers was significant \( T = 5.46, p = 0.013 \).

Finally, Fig. 4 depicts the degree to which the formant frequencies of the point vowels are reduced or centralized in connected speech by superimposing the connected speech quadrilaterals onto the CVC quadrilaterals. The absolute areas of the CVC quadrilaterals and connected speech quadrilaterals do not appear to differ substantially between the two groups. However, absolute differences between each pair of mean formant values (e.g., connected speech /a/ F1 minus CVC /a/ F1) revealed less difference between the LMS group means for /a/, /u/, and the F2 of /i/ than that produced by the MMS group. This suggests less vowel reduction among the LMS group in their connected speech.

**4. Sibilant spectra**

Average energy distributions of the five tokens of /s/ and five tokens of /ʃ/ productions for each speaker as interpreted by the peak frequency (kHz), skewness, and kurtosis of the distributions are summarized in Table VI. This table also presents composite LMS and MMS energy distribution data. As a group, the energy of the LMS speakers’ /s/ productions centered around a higher frequency (5.857 kHz) than the MMS speakers’ productions (5.183 kHz). This difference was significant (Mann–Whitney rank sum test, \( T = 264, p = 0.001 \)). Additionally, the LMS and MMS /s/ productions significantly differed in the skewness of the energy distributions (Mann–Whitney rank sum test: \( T = 289, p = 0.001 \)) with the LMS energy distributions being significantly more negatively skewed than the MMS energy distributions. No significant difference was found in the kurtosis of the distributions.

No significant difference was found between the LMS and MMS /ʃ/ peak frequency or distribution skewness, however, kurtosis did differ between the two groups. LMS group kurtosis was \(-0.536\) while MMS group kurtosis was at

![Diagram of INTONATION and SHIFT patterns](image-url)
This indicates a more diffuse energy distribution of the MMS speakers’ /l/ productions.

5. Rate

Speaking rate measures of the four-sentence speaking passages are presented in Table VII. Total durations of the passages ranged from 17.1 to 21.9 s. Passage duration differences between LMS and MMS speakers were not found to be statistically significant. Calculations of the percentage of total duration occupied by articulation (total duration minus pause time, divided by total duration) revealed that three of the LMS speakers spoke slightly more rapidly than the MMS speakers. LMS speakers 1 and 2 were the most rapid speakers, with articulation comprising 79% of their total durations. MMS 6, the slowest speaker, had 84% of his total duration occupied by speech.

To gain an appreciation for the multifactorial nature of the perception of the LMS-MMS continuum, we generated acoustic profiles for each of the eight speakers by comparing individual data to the grand means. Values that fell one standard deviation or more beyond this grand mean were considered to be part of each speaker’s most salient acoustic feature pattern. For example, if a speaker produced a mean rising slope for intonations that was outside one standard deviation of the grand mean, rising intonation was recorded as part of that speaker’s profile. For ease of presentation, speech profiles are presented for each F0-matched LMS-MMS pair.

a. Pair #4: LMS 4 and MMS 8. Average F0’s for these speakers were 134.4 and 133.0 Hz, respectively. LMS 4’s speech was marked by extensive inflectional patterning in-
including an RIFS pattern that was remarkable for its extent and slope. Additionally, his speech exhibited one of the greater ranges of high to low pitch over the course of his passage. Vocal tract filter characteristics were remarkable for consistently high vowel format values and the negative skewness of the /s/ and /ʃ/ energy distributions. LMS 4 most closely resembles the prototypical LMS speaker of this study (i.e., when all factors are combined), and it is of note that he was rated as less-masculine-sounding than the modulus by 100% of the listeners.

In contrast to LMS 4, MMS 8 spoke with a very consistent F1RS pattern. Seventy-one percent of his intonation contours was the slowest of the entire group. Perhaps the slow rate characteristic of precise enunciation of final obstruents and monotonous pattern was associated with the impression of MMS. MMS 7's speech was closer to monotone than most other speakers as evidenced by the diminished intonation and shift slopes and high percent of flat shifts of his prosodic contours. Interestingly, MMS 6 produced some formant values that fell substantially above the grand mean, a pattern more in accordance with LMS expectations. With 84% of his total passage duration taken up by articulation, this speaker was the slowest of the entire group. Perhaps the slow rate and monotonous pattern was associated with the impression of MMS.

c. Pair #2: LMS 2 and MMS 6. LMS 2 and MMS 6 had F0's of 96.6 and 93.0 Hz, respectively. LMS 2's inflectional contours were unremarkable when compared to the grand mean with the exception of high percent of downward shifts and the percent of unmeasurable intonation segments. This likely is a byproduct of his rapid articulation rate. The most notable aspect of LMS 2's profile was in measures of sibilant energy distributions. LMS 4 most consistently high vowel format values and the negative skew, and high kurtosis values.

d. Pair #1: LMS 1 and MMS 5. LMS 1 and MMS 5 had average F0's of 89.1 and 87.5 Hz, respectively. It is interest that LMS 1's speech was unremarkable for virtually all of the areas investigated (i.e., he did not fall outside one standard deviation from the grand mean on any of the measures), yet he was judged LMS by 91% of the listeners. It is of note that this speaker was the most rapid of the entire group, with long pause times and brief articulation times. An informal review of his speech sample points toward the presence of other LMS perceptual cues that were not measured here, including voice quality (i.e., breathiness) and the hyperspeech characteristic of precise enunciation of final obstruents (Picheney et al., 1986).

MMS 5's acoustic profile was marked by steep slopes of rising intonations. In line with MMS predictions, he produced formant values that were substantially lower than the grand mean. His speech rate was the second slowest of the

### Table VI. Individual and groups mean values for the first three moments of /s/ and /ʃ/ burst spectra (frequency, skewness, and kurtosis) and standard deviations. *Significant differences per Mann–Whitney rank sum test (p<0.05).

<table>
<thead>
<tr>
<th></th>
<th>/s/ kHz (s.d.)</th>
<th>/s/ Skew (s.d.)</th>
<th>/s/ Kurtosis (s.d.)</th>
<th>/ʃ/ kHz (s.d.)</th>
<th>/ʃ/ Skew (s.d.)</th>
<th>/ʃ/ Kurtosis (s.d.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>LMS 1</td>
<td>5.90 (0.242)</td>
<td>0.278 (0.681)</td>
<td>0.779 (1.763)</td>
<td>4.262 (0.283)</td>
<td>0.215 (0.429)</td>
<td>1.158 (0.496)</td>
</tr>
<tr>
<td>LMS 2</td>
<td>6.49 (0.294)</td>
<td>-0.91 (0.500)</td>
<td>1.521 (0.891)</td>
<td>4.997 (0.277)</td>
<td>-0.436 (0.235)</td>
<td>-0.070 (0.636)</td>
</tr>
<tr>
<td>LMS 3</td>
<td>5.15 (0.381)</td>
<td>1.265 (0.474)</td>
<td>2.563 (2.825)</td>
<td>4.435 (0.249)</td>
<td>0.640 (0.369)</td>
<td>-0.368 (0.713)</td>
</tr>
<tr>
<td>LMS 4</td>
<td>5.90 (0.190)</td>
<td>0.057 (0.509)</td>
<td>2.838 (2.130)</td>
<td>4.959 (0.341)</td>
<td>-0.450 (0.389)</td>
<td>-0.672 (0.812)</td>
</tr>
<tr>
<td>Mean LMS</td>
<td>5.86* (0.572)</td>
<td>0.353* (0.699)</td>
<td>1.926 (1.851)</td>
<td>4.664 (0.397)</td>
<td>-0.009 (0.532)</td>
<td>-0.536* (0.571)</td>
</tr>
<tr>
<td>MMS 5</td>
<td>5.22 (0.218)</td>
<td>1.043 (0.311)</td>
<td>2.174 (1.607)</td>
<td>4.361 (0.358)</td>
<td>0.316 (0.413)</td>
<td>-1.089 (0.393)</td>
</tr>
<tr>
<td>MMS 6</td>
<td>5.20 (0.343)</td>
<td>1.245 (0.888)</td>
<td>3.034 (5.298)</td>
<td>4.731 (0.449)</td>
<td>0.014 (0.705)</td>
<td>-0.583 (0.115)</td>
</tr>
<tr>
<td>MMS 7</td>
<td>5.06 (0.272)</td>
<td>1.178 (0.335)</td>
<td>0.841 (0.926)</td>
<td>4.495 (0.415)</td>
<td>-0.237 (0.376)</td>
<td>-1.225 (0.262)</td>
</tr>
<tr>
<td>MMS 8</td>
<td>5.28 (0.260)</td>
<td>0.749 (0.322)</td>
<td>0.517 (0.838)</td>
<td>4.428 (0.177)</td>
<td>0.088 (0.323)</td>
<td>-0.879 (0.269)</td>
</tr>
<tr>
<td>Mean MMS</td>
<td>5.18* (0.245)</td>
<td>1.052* (0.390)</td>
<td>1.670 (2.124)</td>
<td>4.506 (0.305)</td>
<td>0.017 (0.424)</td>
<td>-0.937* (0.429)</td>
</tr>
</tbody>
</table>

### Table VII. Speaking rate measures for the four target sentences expressed in seconds, and the percent of total duration occupied by speech (% articulation).

<table>
<thead>
<tr>
<th></th>
<th>Total duration</th>
<th>Pause time</th>
<th>Articulation time</th>
<th>% Articulation</th>
</tr>
</thead>
<tbody>
<tr>
<td>LMS 1</td>
<td>20.2</td>
<td>4.3</td>
<td>15.9</td>
<td>79</td>
</tr>
<tr>
<td>LMS 2</td>
<td>17.6</td>
<td>3.6</td>
<td>14.0</td>
<td>79</td>
</tr>
<tr>
<td>LMS 3</td>
<td>17.9</td>
<td>3.6</td>
<td>14.3</td>
<td>80</td>
</tr>
<tr>
<td>LMS 4</td>
<td>20.3</td>
<td>3.9</td>
<td>16.4</td>
<td>81</td>
</tr>
<tr>
<td>Mean LMS</td>
<td>19.0</td>
<td>3.9</td>
<td>15.1</td>
<td>79.5</td>
</tr>
<tr>
<td>MMS 5</td>
<td>21.9</td>
<td>3.9</td>
<td>18.0</td>
<td>82</td>
</tr>
<tr>
<td>MMS 6</td>
<td>20.7</td>
<td>3.3</td>
<td>17.4</td>
<td>84</td>
</tr>
<tr>
<td>MMS 7</td>
<td>17.1</td>
<td>3.2</td>
<td>13.9</td>
<td>81</td>
</tr>
<tr>
<td>MMS 8</td>
<td>17.7</td>
<td>3.3</td>
<td>14.4</td>
<td>81</td>
</tr>
<tr>
<td>Mean MMS</td>
<td>19.4</td>
<td>3.4</td>
<td>16.0</td>
<td>82.0</td>
</tr>
</tbody>
</table>
III. DISCUSSION

This study investigated the acoustic correlates of less-masculine-sounding male speech. Because the phenomenon is a perceptual one, care was taken to identify men who represent the polar regions of the proposed LMS-MMS continuum for this group of speakers. Acoustic analysis revealed differences between the LMS and MMS speakers in terms of their prosodic (F0) contours, vowel formant values, sibilant energy distributions, and articulation rates. These differences generally are in accord with predictions derived from literature of female, transgender, and clear speech.

The hypothesis that LMS speakers exhibit more frequent and extensive changes in F0 within and across speech segments than the MMS men is partially supported by the data here. The results suggest, however, that the pattern of F0 changes may be an important consideration. The FISTS pattern was more common for all speakers, but the LMS speakers produced the RIFS pattern more frequently (based on percent of occurrence of overall intonation productions) and with greater extent and slope than the MMS speakers. The LMS speakers then appear to compensate for the rapid and extensive rise of their intonations by quickly reducing their F0 during shifts. The LMS group produced significantly more extensive downward shifts with steeper slopes than those produced by the MMS group.

The LMS group vowel quadrilateral was displaced upward and to the right of the MMS group vowel quadrilateral, with the second formant of /s/ statistically higher for the LMS than the MMS group. Since habitual pitch did not differ between groups, this implies that the frequency differences were either stylistically produced (i.e., by articulation or tongue and mandible positioning), or as a result of differences in vocal tract size. Physical size was not measured or controlled, but the investigators did note that the speakers were not greatly disparate in height. It also was found that vowel reduction in connected speech was greater for MMS speakers than for LMS speakers. Less vowel reduction among the LMS speakers supports the notion that they produced more clear speech than the MMS speakers.

Sibilant production characteristics differed between the two groups, with LMS speakers producing higher /s/ center frequencies surrounded by a negatively skewed distribution of energy. The energy distributions of the MMS productions of /ʃ/ were more diffuse than those produced by the LMS speakers, as indicated by kurtosis values. High center frequencies are characteristic of clear speech caused by either an increase in the center frequency of the turbulent noise at the constriction, or because of an increase in the resonant frequency as a result of decreasing the size of the oral cavity in front of the constriction by lip retraction (Picheny et al., 1986). Lip retraction may also explain the elevated vowel formant frequencies exhibited by the LMS speakers (Sachs, 1975). In addition, the LMS speakers had a wider discrepancy between their center frequencies for /s/ and /ʃ/ than did the MMS speakers. This finding supports the notion of LMS clear speech and may reflect a greater lip spreading and rounding for /s/ and /ʃ/, respectively.

The comparison of each speaker’s profile to the grand mean suggests a multivariate explanation for the perception of LMS-MMS speech. Although there are trends in the data that would suggest typical LMS and MMS profiles, any given acoustic feature may vary in importance, depending on the context in which it occurs.

These findings have clinical relevance for patients wishing to modify the perception of masculinity invoked by their speech. Moreover, the characteristics of the segmental and suprasegmental parameters measured here may be useful in developing stylized synthetic speech that varies on the dimension of masculinity.

ACKNOWLEDGMENTS

This research was part of a masters thesis completed by the first author and directed by the second author at the University of Minnesota, and portions were presented at the American Speech-Language and Hearing Association convention in Anaheim, 1994. Support for this research was provided in part by an award from the Bryngelson Fund, Department of Communication Disorders, University of Minnesota.


