Evidence from HNR that /s/ is a social marker of gender

Kevin Heffernan
University of Toronto

There are many studies demonstrating the phonetic cues to gender found in our speech, but for many of these cues it is not clear if they are a product of our anatomical make-up or a product of our social background. The goal of this study is to demonstrate that one of these cues to gender, the frequency spectrum of the frication produced by the sibilant /s/, contains a socially acquired component. Male and female adult speakers of two languages, Canadian English and Japanese, were recorded producing words containing /s/. The harmonics-to-noise ratio (HNR) of each subject’s speech sample was also measured. ANOVA analysis of the spectral moments of /s/ revealed significant differences between the two language groups. As well, the female English speakers showed a significant correlation between the first moment and HNR, suggesting that at least for the female English group, /s/ contains a socially acquired component.

1 Introduction

There are many studies demonstrating the plethora of robust phonetic cues to gender found in our speech. However, at this point it is not clear as to how much of this information is a product of anatomical make-up and how much of it is a product of our social background. The goal of this study is to demonstrate that one of these cues to gender, the frequency of the frication produced by the sibilant /s/, contains a social component. While the results of studies on feminine-sounding males (Avery and Liss 1996, Linville 1998) suggest that there is a social component, these results have not been clearly established for normal-sounding adults. Male and female adult speakers of two languages, Canadian English and Japanese, are recorded, and analyses performed on the spectral moments of /s/. While /s/ is the primary focus of the paper, supporting evidence is also drawn from harmonics-to-noise ratio (HNR) measurements for each of the speakers.

1 Throughout this paper, the term ‘gender’ refers to both biological sex and social gender.
2 By social component to /s/ it is meant a manner of articulating /s/ that produces an acoustic component that correlates with one’s social background, such as economic class, education, or gender.
1.1 The socially acquired component of the acoustic signal

It is a well established fact that listeners can identify the gender of a speaker with limited phonetic information. For example, listeners are able to accurately identify the gender of a speaker from isolated voiceless fricatives alone (Ingemann 1968, Schwartz 1968). Lass et al. (1980) report accuracy ratings of greater than 95% in the identification of speaker gender from both 255 Hz low-pass filter speech and 255 Hz high-pass filter speech. Even when the laryngeal component of speech is excluded by using whispered speech, gender identification is still significantly above chance; Lass et al. (1976), using isolated tokens of vowels, report a 75% accuracy rate for whispered speech, a 91% accuracy rate for 225 Hz low-pass filtered speech, and a 96% accuracy rate for unfiltered speech.

The differences between the speech of males and females at the acoustic level can to a certain extent be straightforwardly accounted for by differences in the anatomical make-up of the vocal tract. For example, males’ vocal folds are 50% larger than females’ vocal folds (Henton 1992:37). The extra length of the males’ vocal folds causes a longer oscillation period of the periodic signal produced during phonation, which in turn results in a lower fundamental frequency. Recent research in the differences between male and female speech at the acoustic level has revealed that some of the differences are acquired from one’s social environment based on one’s gender identity. The argument for this social component is based mostly on research on the speech of children, and research on the speech of gays and lesbians. To this, recently there has been the addition of a limited amount of research on cross-linguistic variation and variation across social groups within a single language community. If the acoustic differences between the speech of males and females were a consequence of anatomical differences only then in every case we would not expect see much variation between male and female children, or between speakers of the same sex but different sexual orientation, language, or social background.

Consider first the studies on the speech of children. Sachs et al. (1973) reported an 81% identification rate for the gender of children between 4 and 14 years from short sentences. Perry et al. (2001) found that when the stimuli were reduced to a single CVC syllable accuracy did not drop significantly (identification rate of 74% for children between the ages of 4 and 16 years). These results are particularly interesting in light of the fact that “there are no very salient differences between the sexes in terms of morphology or size of the vocal apparatus during the first phase [between birth and puberty]” (Beck 1997:258).

Whiteside and Marshall (2000) found marked sex differences in the voiceless / voiced contrast in terms of voice onset time (VOT) for voiceless obstruents in the speech of 9- and 11-year old boys and girls. They found that by the age of 11, girls had significantly longer VOT values than boys, which agrees with the data on both English-speaking adults (Ryalls et al. 1997, Swartz 1992, Whiteside and Irving 1997) and German-speaking adults (Scharf and Masur 2002). Whiteside and Marshall suggest that these differences stem from not only sex-linked developmental changes in anatomy and physiology, but from sociophonetic factors as well, namely that women tend to articulate speech more carefully (2000:206-7).

Next consider sexual orientation. Avery and Liss (1996) reported significant differences between measures of vowel midpoint values for male voices that were
perceived as ‘less-masculine-sounding’ and male voices that were perceived as ‘more-masculine-sounding’.

Crist (1997) discovered that /s/ frication in /sk/ and /sp/ clusters was lengthened in stereotyped gay speech. Although the results were not very robust (the effect was not found for /st/ clusters), they still indicate that there are phonetic cues to sexual orientation.

Lastly, consider speech community. Blandon et al. (1984) calculated in Bark the mean difference between normalized vowel tokens from males and females for seven linguistic groups: RP English, French, General American English, Modified Northern British English, Swedish, Standard Dutch, and Utrecht. The differences ranged from 1.2 Bark for RP English to 0.55 Bark for Utrecht Dutch.

Ryalls, Zipprer, and Baldauff (1997) measured VOT for 10 African Americans and 10 Caucasian Americans, and found significant differences for both gender and race. The African Americans tended to use more prevoicing (i.e. a negative VOT) for voiced stops than their Caucasian American counterparts.

All of these studies point to the same conclusion: there are many robust phonetic cues to our gender in our speech.

1.2 The social component of /s/

The pronunciation of /s/ is one of the many cues to social gender. The best documented difference between the production of /s/ by females and males is that females produce /s/ with concentrations of energy at higher frequencies than males (see Flipsen et al. 1999 for an overview of the literature). To give two examples of results, Schwartz (1968) reported mean peak frequencies for /s/ produced in isolation for 9 adult males of approximately 5.5 KHz and for 9 adult females of approximately 6.5 kHz (taken from their Figure 1). Flipsen et al. (1999) report centres of gravity for /s/ embedded in a carrier phrase for 12 adolescent males of approximately 6.2 kHz and for 14 adolescent females of approximately 7.5 kHz.

Linville (1998) measured /s/ duration, /s/ peak frequency, average speaking fundamental frequency, speech rate, and long-term average spectra of 5 openly gay men and 4 straight men. She had naïve listeners identify sexual orientation based on speech samples, and reports an identification rate of 80%. The gay judgements correlated with higher peak frequency values and longer duration for /s/.

Avery and Liss (1996) found significant differences between measures of the first, third, and fourth spectral moments of two fricatives for male voices that were perceived as ‘less-masculine-sounding’ and male voices that were perceived as ‘more-masculine-sounding’. Clearly these differences cannot be accounted for by anatomical determinism alone.

Starks (2000), using impressionistic phonetics, measured the distribution of two variants of /s/, one with the tongue tip lowered, and the other with the tongue tip raised to a position close to the back of the teeth, among four ethnic groups in New Zealand, Pacific Islanders, New Zealand Europeans, Maori, and Asians. Significant differences in the percentage of speakers using the raised variant were found both across the groups (ranging from 61% for the Pacific Islanders to 23% for the Asians) and within each group across gender (on average 64% of the females used the raised variant compared to 41%
of the males). Unfortunately, Starks does not report what the acoustic correlates of ethnic identity are in this study.

Stuart-Smith et al. (2003) carried out a spectral analysis of /s/ produced by 31 speakers of Glaswegian. The speakers were stratified by age (teenagers or 40-60 years), social class (working class or middle class) and sex (male or female). They found not only a consistent effect for sex, but also effects for age and class as well. Of interest is the finding that working class girls group with males speakers in their values for mean and peak spectral frequencies.

The above studies suggest that at least for feminine-sounding males and children, there is a socially acquired component. How exactly is it that the feminine-sounding men produce a higher frequency frication during the production of /s/? The acoustic noise of /s/ is produced when air, after passing through a narrow channel at the point of constriction, hits an obstruction (the teeth) downstream. Shadle (1985), using a set of mechanical models demonstrated that “the source spectrum depended only on the flow rate and the constriction-obstacle distance, not on the length of the surrounding vocal tract” (as reported in Shadle 1991:414). Furthermore, Ladefoged and Maddieson state that “in a fricative a variation of one millimetre in the position of the target for the crucial part of the vocal tract makes a great deal of difference” (1996:137). To summarize, the crucial factor in determining the spectrum of /s/ is the distance from the point of constriction to the teeth, and a small amount of variation in that distance will produce perceivable changes in the frication. It must be that feminine-sounding men are producing /s/ with a more fronted point of constriction. If this distance is significantly shorter for women than men, then this would explain their high peak frequencies. But, “general male-female vocal tract differences, however, exist mainly behind the area of constriction and the obstacle” (Strand 1999:88). Strand implies that it is doubtful that the slight differences in the vocal tract that do exist in front of the point of constriction between men and women are capable of accounting for the robust differences seen in the frequency of the frication. This point will be picked up again in the discussion of the concluding remarks.

1.3 The role of breathiness in the construction of gender

Breathiness is the introduction of noise in the higher frequencies of the speech signal due to the incomplete adduction of vocal folds during phonation as a result of low muscular effort (Laver 1980). Henton and Bladon (1985) measured the difference in amplitude of the first and second harmonics of open vowels taken from the citation speech samples of 36 Received Pronunciation speakers (twenty females, sixteen males) and 25 Modified Northern accent speakers (twelve females, thirteen males). Henton and Bladon point out that breathiness is used as a phonemic contrast in some languages of the world, and therefore a priori we do not expect enough of an anatomically determined difference between men and women to interfere with the perception of this phonemic contrast. Regardless, they discovered that the female speakers had significantly greater amplitude in the first harmonic than the second harmonic (an indication of a breathy voice signal) than their male counterparts. They argue (p. 226) that (at least in these
dialects of English) women intentionally use a breathy voice to create desirability in the opposite sex through its association with sexual arousal:

If a woman can manage to sound as though she is sexually aroused, she may be regarded as more desirable or with greater approbation by a male interlocutor than if she speaks with an ordinary, modal voice. At an ethological level, breathy voice may be seen as apart of the courtship display ritual, as important as bodily adornment and gesture.

Given this argument, it holds that breathiness plays a key role in the marking of gender (at least in English), and therefore should be a good indicator of the social gender of the speaker.

Greater breathiness in the voice of females was also found for American English speakers (Klatt and Klatt 1990), Dutch (Günzburger 1991) and Spanish (Trittin and Santos y Lleó 1995), although the results for Spanish were not significant. There is no data comparing men and women in Japanese that I know of, but Ito (2003), looking only at Japanese males, reports that breathiness contributes to the expression of positive politeness in Japanese.

Breathiness also varies from one speech community to the next. In their study, Henton and Bladon found that the Modified Northern speakers (both men and women) had more breathy speech than the Received Pronunciation speakers. Significant differences in HNR measurements were also found between African American English and Standard American English (Purnell et al. 1999), and between Italian and Polish male speakers (Wagner and Braun 2003). These results strongly support the claim that breathiness is not only a product of physiology differences in the vocal folds of men and women.

We are now in a position to state the central hypothesis of the paper. If both the frication of /s/ and the breathiness of the voice signal are primarily the product of a social component that is used to express the social gender of the speaker, then the spectral moments of /s/ and the HNR will pattern together; that is, speakers with a high frequency frication will have a breathy voice (i.e. a low HNR), and vice versa.

2 Methodology

Flipsen et al. (1999) addressed a number of issues concerning the measurement and classification of the spectra of /s/. They found that the optimal acoustic characterization of /s/ is (a) obtained from the midpoint of frication, (b) represented in a linear scale, (c) reflected in the statistics of the 1st and 3rd spectral moments, (d) controlled for phonetic context, (e) collapsed across the age range, and (f) described individually by gender. These points were taken into consideration when the experimental methodology was designed.
2.1 Speakers

22 speakers, 10 Japanese (five female and five male), 12 English (six female and six male), were recruited from the university’s student population. Besides the recording, each speaker was asked to fill out a brief questionnaire in order to obtain background information. Every speaker reported that the language under investigation was their mother tongue, all were monolingual as a child, and in the case of the Japanese, they speak Japanese at least 25% of the time now as an adult. The speakers’ ages ranged from 21 to 36 years old (mean = 26.3). The speakers were compensated with a small sum of money for their participation.

2.2 Articulation of /s/ in Japanese

During the articulation of /s/ in Standard Japanese, the blade of the tongue is raised towards the ridge of teeth to form a small opening. Whether the tongue points to the upper teeth or the lower teeth does not matter (Akamatsu 1997). Akamatsu classifies /s/ as lamino-alveolar.

2.3 Articulation of /s/ in English

According to Ladefoged and Maddieson (1996:146), English /s/ “usually has a constriction in the middle of what we refer to as the alveolar region (i.e. the forward part of the alveolar ridge). It can be formed either by the tip of the tongue, or by the blade with the tip behind the lower front teeth.” Video imaging of the tongue during articulation of /s/ shows that the mid-section cross-sectional profile of the tongue looks like an upside-down W, with a groove running down the centre and sloping sides (Stone and Vatikiotis-Bateson 1995). It is unclear if this groove also occurs in the Japanese production of /s/. Nor is it clear how much the actual shape of the tongue influences the frication produced during articulation.

2.4 Materials

A set of word cards was created containing 65 disyllabic CVCV words in natural orthography. Among these words were 25 tokens of /s/: 15 word-initial tokens, and 10 word-medial tokens. Each token occurs before either the vowel /a/ or the vowel /e/ to avoid secondary articulation effects of rounded vowels (Shadle and Scully 1995). The environment /i/ was excluded because /s/ occurs as the allophone [ʃ] before /i/ in Japanese. The other tokens were added as distractors. The words were read from the cards one at a time and recorded with a Marantz CDR300 digital CD recorder (16 bit resolution, 44.1 kHz sampling frequency) and an Audio-Technica omnidirectional condenser clip-on microphone (30 ~ 20,000 Hz frequency response, S/N ratio > 65dB) attached approximately 15 cm away from the speaker’s mouth. The set of word cards was
shuffled and then reread two more times for a total of 25x3 /s/ tokens per speaker. This resulted in three recording files for each speaker.

2.5 Analysis of the spectral moments of /s/

Spectral analysis was performed on a personal computer using PRAAT. Following Flipsen et al. (1999), the centre of each token was visually identified by the researcher and a 40 ms Hamming window extracted. The spectrum was derived via Fast Fourier Transform (FFT). For each spectrum, all four moments were calculated using the PRAAT software. The moment statistics for a fricative token are derived by treating the FFT spectrum as if it were a random probability distribution (see Forrest et al. 1998). The first moment, Centre of Gravity, is a measure of where the centre of the energy concentration produced by the frication lies. The second moment, Standard Deviation, is a measure of how much the energy spreads out from the Centre of Gravity. The third moment, Skewness, is a measure of how much the shape of the distribution below the Centre of Gravity differs from the shape of the distribution above the Centre of Gravity. The fourth moment, Kurtosis, is a measure of how much the distribution differs from a Gaussian shape in terms of the peakedness of the central section.

2.6 Analysis of harmonics-to-noise ratio

The HNR is the ratio of the magnitude of the periodic signal to the magnitude of the aperiodic signal, and was originally introduced as a measure of voice hoarseness (Yumoto et al. 1982). Basically, the greater the noise component of the speech signal, the lower the ratio is. The first step in measuring HNR is to isolate a speech sample that is completely voiced. Following Biemans (2000), each speech recording was sampled at 10 ms intervals, and a voiced / unvoiced decision made for each sample using the forward cross-correlation method built into PRAAT (Boersma 1993). All voiced samples were concatenated to form a continuous stretch of voiced speech, on which the HNR calculations were then performed. The values reported are the mean of the HNR values calculated by the PRAAT HNR algorithm using a sampling frame of 10 ms for each file.

3 Results and discussion

3.1 Spectral moments of /s/

Tables 1 and 2 give the mean and standard deviation for the four spectral moments for all tokens by sex.
Table 1 - Mean values and standard deviations for the four moments for Japanese speakers

<table>
<thead>
<tr>
<th></th>
<th>Centre of Gravity</th>
<th>Standard Deviation</th>
<th>Skewness</th>
<th>Kurtosis</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>mean</td>
<td>s.d.</td>
<td>mean</td>
<td>s.d.</td>
</tr>
<tr>
<td>women, n=5</td>
<td>8411.9</td>
<td>530.2</td>
<td>3082.7</td>
<td>443.4</td>
</tr>
<tr>
<td>men, n=5</td>
<td>8000.8</td>
<td>477.7</td>
<td>2766.3</td>
<td>320.4</td>
</tr>
</tbody>
</table>

Table 2 - Mean values (and standard deviations) for the four moments for English speakers

<table>
<thead>
<tr>
<th></th>
<th>Centre of Gravity</th>
<th>Standard Deviation</th>
<th>Skewness</th>
<th>Kurtosis</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>mean</td>
<td>s.d.</td>
<td>mean</td>
<td>s.d.</td>
</tr>
<tr>
<td>women, n=5</td>
<td>8276.2</td>
<td>702.2</td>
<td>2680.6</td>
<td>395.5</td>
</tr>
<tr>
<td>men, n=5</td>
<td>7487.5</td>
<td>607.9</td>
<td>2932.9</td>
<td>334.0</td>
</tr>
</tbody>
</table>

An ANOVA analysis was run with two between-subjects factors (Language, Sex) and two within-subjects factors (Repetition number, Position) for each of the moments. Language was either Japanese or English. Sex was either male or female. Repetition number was one through three, since the word list was repeated three times. Position was either word-initial or word-medial. The significance level was 0.05. For the first moment, there were significant effects for Position F(1,18)=5.90, p=0.026, and Sex F(1,18)=8.54, p=0.009. There was one significant interaction PxL F(1,18)=15.01, p=0.001. For the second moment, there was a significant effect for Position F(1,18)=14.76, p=0.001. There was a significant interaction for PxLxS F(1,18)=9.45, p=0.007. For the third moment, there was there was a significant effect for Position F(1,18)=7.08, p=0.016., and a significant interaction, PxL F(1,18)=11.99, p=0.003. For the fourth moment, there was only one significant interaction, PxL F(1,18)=5.69, p=0.28.

As expected, no main effect was found for Repetition number, and a main effect was found for Sex, but only for the first moment. A Newman-Keuls post-hoc test was run on each of the interactions, with the results given in Table 3.
moment 1: Position x Language

<table>
<thead>
<tr>
<th></th>
<th>initial</th>
<th>medial</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Japanese</td>
<td>8071.1</td>
<td>8369.8</td>
<td>p&lt;0.001</td>
</tr>
<tr>
<td>English</td>
<td>7909.3</td>
<td>7840.8</td>
<td>n.s.</td>
</tr>
<tr>
<td>p</td>
<td>0.027</td>
<td>p&lt;0.001</td>
<td></td>
</tr>
</tbody>
</table>

moment 2: Position x Language x Sex

<table>
<thead>
<tr>
<th>sex</th>
<th>initial</th>
<th>medial</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>men</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Japanese</td>
<td>2747.8</td>
<td>2801.0</td>
<td>n.s.</td>
</tr>
<tr>
<td>English</td>
<td>2898.8</td>
<td>3052.7</td>
<td>n.s.</td>
</tr>
<tr>
<td>p</td>
<td>0.003</td>
<td></td>
<td></td>
</tr>
<tr>
<td>women</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Japanese</td>
<td>2976.8</td>
<td>3241.5</td>
<td>p=0.001</td>
</tr>
<tr>
<td>English</td>
<td>2671.3</td>
<td>2664.6</td>
<td>n.s.</td>
</tr>
<tr>
<td>p</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
<td></td>
</tr>
</tbody>
</table>

moment 3: Position x Language

<table>
<thead>
<tr>
<th></th>
<th>initial</th>
<th>medial</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Japanese</td>
<td>0.383</td>
<td>0.176</td>
<td>p&lt;0.001</td>
</tr>
<tr>
<td>English</td>
<td>0.386</td>
<td>0.416</td>
<td>n.s.</td>
</tr>
<tr>
<td>p</td>
<td>0.014</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

moment 4: Position x Language

<table>
<thead>
<tr>
<th></th>
<th>initial</th>
<th>medial</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Japanese</td>
<td>1.702</td>
<td>1.176</td>
<td>p&lt;0.001</td>
</tr>
<tr>
<td>English</td>
<td>1.631</td>
<td>1.478</td>
<td>n.s.</td>
</tr>
<tr>
<td>p</td>
<td>&lt;0.001</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 3 - Significance results from the Newman-Keuls post-hoc tests

The post-hoc tests show that there are significant differences between the moments of word-initial and word-medial /s/ in Japanese only (with the exception of the second moment, in which case it is the Japanese women only). I am not sure why there are significant results for position in Japanese. It may be that in that language the word-medial /s/ is heavily influenced by the preceding syllable, but in English the co-articulatory effects are much weaker. Further pursuit of this possibility is beyond the scope of this research.

The relevant results to the research at hand are the differences between languages. As just discussed, any difference in the medial position may be due to unrelated causes such as the effects of co-articulation. If we look only at word-initial position, we still see that there is a significant difference between languages for the first moment. This is also the only moment that shows sex as a main effect. As such, if speakers and listeners are to use /s/ as cues to gender, they must be using the first moment (i.e. the frequency of the energy concentration) as their primary cue. That this moment also shows significant differences for language (at least in word-initial position) supports the hypothesis that there is a social component to the articulation of /s/. Of course, it may still be argued that...
this difference is due to the fact that the Japanese speakers are more diminutive in size, and therefore have smaller vocal tracts (although recall Shadle’s argument against this mentioned earlier). Any lingering doubts about the claim that there is a social component to the production of /s/ will be eliminated by the HNR results reported in the following section.

3.2 Harmonic-to-noise ratio

Table 4 gives the HNR results. For both languages, women have a greater HNR than men. That women have greater HNR values than men is consistent with the results found by Biemans (2000:71) for Dutch speakers, and by Ferrand (2000) for English speaking prepubescent children.3

<table>
<thead>
<tr>
<th></th>
<th>Japanese Women</th>
<th>Japanese Men</th>
<th>English Women</th>
<th>English Men</th>
</tr>
</thead>
<tbody>
<tr>
<td>HNR</td>
<td>13.80</td>
<td>11.95</td>
<td>15.22</td>
<td>12.35</td>
</tr>
<tr>
<td>s.d.</td>
<td>1.70</td>
<td>0.93</td>
<td>1.36</td>
<td>1.80</td>
</tr>
</tbody>
</table>

Table 4 - Mean HNR values and standard deviations in dB

In order to test the hypothesis that breathiness of the voice correlates with the production of /s/, a Pearson’s correlation coefficient was calculated on the HNR ratios and the moment means by Position (word-initial, word-medial) for the Japanese women, Japanese men, English women, and English men, with the significance level set at 0.005. For the word-initial position measures, significant results were found for the first moment measures of the English women r(10)=0.889, p=0.018, and for the third moment measures of the Japanese women r(8)=0.895, p=0.043. No significant results were found for the men. For the word-medial position, a significant result was only found for the first moment measures of the English women r(10)=0.914, p=0.011.

A priori, there is no reason why the breathiness of the voice should correlate with the first moment of /s/; one is determined by the supralaryngeal configuration of the vocal tract while the other is determined by the laryngeal settings. Rather, it must be that both breathiness and the first moment of /s/ are robust markers of social gender, at least to the female English speakers. If one is going to assert femininity with a breathy voice, then it only makes sense that if there was a social component to /s/, they would also assert femininity with a high frequency /s/. The strong correlation between HNR values and the first moment of /s/ shows that this is most likely the case, lending further support to the hypothesis that there is a social component to /s/.

4 Conclusions and further questions

The larger difference between the mean values for the first spectral moment for the English men and women and the strong correlation between the first moment and HNR for the female English speakers strongly suggesting that the acoustic signal

---

3 Neither Biemans nor Ferrand indicate why this should be.
produced by the frication of /s/ contains a socially acquired component. If the frication of /s/ was mostly determined by anatomical make-up then we would expect to see a higher group mean for the first moment of the Japanese speakers, who are roughly of smaller stature than the English speaking subjects.

The relationship between HNR and the first spectral moment of /s/ in English for the female speakers clearly shows that both are working together to portray information about the social gender of the speaker. Most likely there is a social component to /s/ in Japanese as well, but just not as robust as in English. Thus we see still the outlier JF1, but JF1’s mean value does not even approach the extreme shown by the English outlier EF4. The difference in the importance of /s/ as a marker of social gender in English and Japanese may be because the Japanese language has number of other ways of marking social gender, ranging from the blatant, such as first-person pronoun choice and sentence-final particle choice (Sturtz 2001) to the subtle, such as the choice of hedge words (Lauwereyns 2002), the ellipsis of topic and subject markers wa and ga (Takano 1998), and the ellipsis of the object marker wo (Matsuda 1995), to name but a few. On the other hand, it may be that /s/ is more of a robust marker in English than in Japanese because it is more common due to its exceptional phonological distribution (it violates the sonority hierarchy in words such as stop and spoon), and its occurrence as the second person morpheme on verbs and the plural morpheme on nouns.4

A question that now remains is how much, if any at all, anatomy plays a role in the determination of the spectral moments of /s/. We saw that for both languages, the females as a group had higher first moment mean frequency than the male group. I assume that this is a universal trend, but this needs to be empirically tested. If this assumption holds, then clearly anatomy plays same role. At the same time, we have female speakers such as JF1 and EF4 whose mean first moment values are less than the group mean for the corresponding males. There is nothing to suggest that their low means are anatomically determined. How then do we explain the existence of such outliers? I propose that among the population there are a small number of petite women and large men who actually produce higher and lower frequency frication of sibilants due to the anatomy of the their vocal tracts. The population as a whole then picks up on these differences in frequency and creates a corresponding stereotypical perception of gender (see Biemans 2000:13 for an elaboration of this model with respect to voice quality). That such stereotypes do truly exist is demonstrated clearly by Strand (Strand and Johnson 1996, Strand 1999) in her work on the role of gender stereotypes in speech perception. This stereotype then in turn gets amplified in the production stage as women conform to the social perception of femininity and men to the social perception of masculinity. However, given that the production targets reflect socially-determined stereotypes, there is nothing preventing a woman from acquiring a masculine-sounding /s/ or a man from acquiring a feminine-sounding /s/, and this is exactly what we see here and in the studies on feminine-sounding men.

To end the discussion, I would like to consider these results in the greater context of sociolinguistics studies on sound change in general. Given that coronal sibilants do not occur in every language of the world (Maddieson 1984), and that in those that do have a

---

4 Furthermore, /s/ occurs with limited frequency in Japanese because it is restricted to occurring only before back non-front vowels.
coronal sibilant(s) show variation in the phonological status of those sibilants (see Evers, Reetz and Lahiri 1998), theoretically it is possible for sibilants to participate in diachronic language change. Indeed we do see their participation at the phonological level in the form of changes between the categories in the voiceless/voiced and the fricative/affricate dichotic phonological contrasts (for example, Wolf 1984). But is it possible for this stereotyped perception of gender, and hence the articulatory target that generates the appropriate frication, to participate in a sound change? Considering that it is a construct of our society, and that our society is something that changes, then I don’t see why it would not be possible. If so, then is there currently a change-in-progress in English (at least at the phonetic level) with speaker EF4 an innovator, leading the change? If a change is under way, then does it correlate with the changing attitudes and practices towards the roles of women in society?

Finally, is the central frequency of the frication of sibilants one of a class of phonetic cues to gender (such as breathiness, creakiness, VOT, pitch variability, speech rate, etc.) that behave in similar ways, that in anatomical make-up plays an initial role in seeding the social perception of gender associated with the phonetic marker, but does not limit the actual production of that marker by individual speakers? If such a class exists, is there then another class of phonetic cues to gender, such as fundamental frequency, that are governed by anatomical make-up alone, and therefore behave in strikingly different ways? In the past, technological limitations forced sociolinguists to focus on more salient markers of sociolinguistic variables such as syntactic and phonological variation. While these technological limitations still do exist to certain extent, they are quickly disappearing. The time is now ripe for a re-examination of sociolinguistic theory at the detailed level of phonetics, and a good place to start is by looking for the answers to the questions posed here.

References


