Variation in /s/ and the perceived gender typicality of children's speech

Benjamin Munson

Department of Speech-Language-Hearing Sciences, University of Minnesota, Minneapolis
Munso005@umn.edu

ABSTRACT

Previous research has found an association between the acoustic characteristics of /s/ and judgments of the sex typicality of adult men's voices. The current study examined whether a similar association could be found when children's voices are used as stimuli. Two sets of productions were used, with either naturally produced /s/ tokens or tokens of /s/ produced by a trained phonetician. Listeners rated children to sound more girl-like if they produced /s/ with an especially high peak frequency, or with a diffuse spectrum suggesting a frontal misarticulation.

Keywords: Fricative, Gender, Perception, Children, Misarticulation

1. INTRODUCTION

It is axiomatic that the acoustic and perceptual characteristics of men and women's speech differ. As reviewed by [7], there are acoustic differences between the sexes in nearly every parameter that has been investigated, including vowel formants, fundamental frequency, and the spectrum of frication noise, among others. Many of these differences are likely due to sex differences in the size and shape of the vocal tract, and the mass and thickness of the vocal folds. There is ample evidence that not all sex differences in acoustic parameters are the consequence of anatomy and physiology. As shown by [3], male-female differences in vowels' formant frequencies vary across languages. [2] showed that male-female differences in /s/ acoustics cannot be predicted by male-female differences in the anatomical structures relevant for /s/ production. Moreover, some sex differences in production cannot even plausibly be explained by anatomical differences between the sexes, such as the fact that social class mediates sex differences in /s/ in Glaswegian English, as observed by [11]. These and other findings suggest that some sex differences in phonetic detail are acquired during language development.

Listeners readily assign gender to talkers based on phonetic information alone, even when the samples being rated are very short [1]. This is true even when listeners are rating the speech of children. [9] showed that listeners rate the speech of boys as young as four years old as significantly more boy-like than the speech of age-matched girls. The focus of this study is how variation in children's fricative production influences adults' perception of the gender typicality of their voices. Recently, [6] and [8] showed that variation in /s/ production predicts judgments of the gender typicality of adult men's voices. [8] examined this by exploiting the natural variation in /s/ characteristics across a group of 22 men. They found that men were rated as sounding less canonically masculine if they produced a token of /s/ that had an especially high peak frequency and compact spectrum. [6] used matched-guise stimuli to examine the perception of the same 22 men's voices. In these stimuli, the natural /s/ was replaced by various tokens of /s/ produced by a trained phonetician. Again, voices were rated as less masculine if they contained an especially high peak frequency /s/ than if they contained one with a lower peak frequency. Interestingly, [6] also found that talkers were rated to be less masculine if their productions contained a frontally misarticulated /s/, i.e., one that resembles [θ]. The mean ratings for stimuli with an especially high peak frequency /s/ and those for stimuli with a [θ]-like /s/ were nearly identical. This finding is interesting because studies of men's speech, such as [8], find non-prototypically masculine-sounding speakers are *not* more likely to produce [θ]-like /s/ than are more-prototypical speakers. However, there is a well-established social stereotype that gay men produce frontally misarticulated /s/, as discussed in [7]. The findings of [8], then, appear to show that listeners use authentic experiences hearing less-prototypically masculine men and stereotypes about male speech when making ratings of gender typicality.

The findings of [8] inspired the current investigation, which examined the relationship between /s/ variation and the perceived gender typicality of children's voices. It contains two types of stimuli, one set in which listeners rated unedited /s/-initial words produced by children, and one in which the children's /s/ was replaced with either an /s/ with an especially high peak frequency (like that of the less-masculine sounding men in [6]) or a frontally misarticulation resembling [θ]. The [θ]-like stimuli are of particular interest in this investigation, as there is reason to believe that they

will elicit different ratings from the [θ]-initial stimuli in adults. There are well-established sex differences in speech sound acquisition, with girls outpacing boys [10]. Hence, listeners might identify the [θ]-initial tokens as more boy-like because they contain a frank error. We might also predict that this tendency will interact with the age of the participant, perhaps so strongly that [θ]-initial productions are rated as less prototypically masculine when the speaker is older (and hence subject to the adult gender norms for /s/ production), than when the speaker is younger.

2. METHODS

2.1. Stimuli

The talkers were 34 boys aged 4 to 13 years. They all had full-scale IQs within normal limits, were monolingual speakers of English, and had no current speech, language, or hearing impairments. Only male talkers were used because the stereotypes about variation in adult /s/ production of [θ] are strongest for adult men. These boys produced a series of words with a variety of vowels and consonants. A single production of 14 words from each of these boys was used in this experiment.

Two of these words, *sock* and *sun*, were chosen for the analysis presented in this report. These were the only words from the set of 14 that had word-initial singleton /s/. Three versions of each token were used in the experiment. The first was the naturally produced version. The second was a version in which the /s/ was replaced by a version of /s/ produced by a trained phonetician and which had an especially high peak frequency and a very compact spectrum (henceforth *high peak frequency [HPF] /s/*). The second was a version of /s/ that was produced with a frontal misarticulation (henceforth *[θ]-like /s/*). These were combined with the naturally produced VC sequences from *sock* and *sun*. The VC sequences were normalized for RMS amplitude prior to combining them with the two fricatives. The RMS amplitude of the [θ]-like /s/ was lower than that of the HPF /s/, as is true of natural productions of /s/ variants. Together, the stimuli with HPF /s/ and the [θ]-like /s/ were called the *matched guise* stimuli.

For each stimulus, the f1, f2 (measured in Bark units) and f0 (measured in ERB units) of the midpoint of the vowel were logged. For each of the naturally produced fricatives, the first and second spectral moments were calculated for a 40 ms interval of frication noise centered at the mid-point of the fricative. The first spectral moment, peak frequency (henceforth *m1*) is thought to correspond to the location of the maximum vocal-tract constriction in the anterior-posterior dimension. It distinguishes between /s/ and /ʃ/ [4], and it characterizes the distinctively HPF /s/ associated with less-masculine sounding men [8]. The second spectral moment, spectral spread (henceforth m2) distinguishes between /s/ and /θ/. Figure 1 shows a plot of m1 by m2 for the 68 naturally produced /s/ tokens, and for the two matched-guise fricatives.

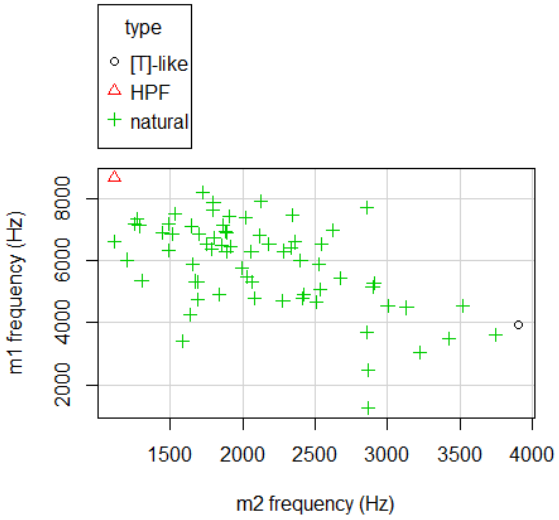

Figure 1: First (m1) and second (m2) spectral moments for the naturally produced and matched-guise high-peak frequency (HPF) and[θ]-like /s/ tokens

As Figure 1 shows, the natural fricatives were distributed well in both the m1 and m2 dimension. It also shows that the [θ]-like /s/ was just above the highest value of m2 in the naturally produced fricatives, and the HPF /s/ was just above the highest value of m1.

2.2. Listeners

The listeners were 38 people native speakers of English with no current or previous speech, language, or hearing impairment. They were recruited at the University at which this experiment was conducted.

2.3. Procedures

During the experiment listeners were presented with individual tokens over high-quality headphones while viewing a computer monitor. After each token, they were asked to rate the gender typicality of the child's speech on the six-point scale used by [9], where 1=*positively a female*, 2=*appeared to be a female*, 3=*unsure; may have been a female*,

4=*unsure; may have been a male*, 5=*appeared to be a male*, 6=*positively a male*. Because we were interested in the possible mediating role of children's age on responses, and because children's age is not unambiguously discernible from speech, listeners were told the child's age in years on each trial. The talker-word combinations were fully randomized. Listeners heard only one token of *sock* and *sun* for each talker, so that listeners never rated the natural production, the HPF /s/, and the [θ]-like /s/ for the same word by the same talker. There were six different stimulus sets across the 38 listeners, arranged so that an approximately equal number of listeners rated each token. The full set of 14 words by each talker was presented, for a total of 476 stimuli per listener, 68 of which per listener are analysed in this report.

3. RESULTS

3.1. HPF and [θ]-like /s/ tokens

The first analysis examined the influence of /s/ type for the matched-guise stimuli only. The ratings for these were used as the dependent measures in a linear mixed-effects model. The dependent measure was modelled as a Poisson distribution, given that there were only six possible responses. A series of progressively more complex models were built. The simplest model included only random intercepts for listeners. Subsequent models introduced new independent variables, random slopes for the influence of that variable on individual listeners, and, for models with more than one independent measure, interactions between independent measures. Model building ceased when a more complex model did not have a statistically significantly better fit than the next-least complex model.

The first analysis used /s/ type (HPF, [θ]-like) and talker age in years as predictors, entered in that order. The model that best fit the data was one with both of these factors included, with a random slope for the influence of /s/ type on individual listeners, and no interaction between /s/ type and age. Both independent factors were significant (for /s/ type [reference: HPF /s/]: $\beta = 0.143$, SEM=0.023, $z=6.164$, $p<0.001$, for age: $\beta = 0.017$, SEM=0.004, $z=4.228$, $p<0.001$). This indicates that the stimuli with [θ]-like /s/ were given higher (i.e., more prototypically boy-like) ratings than stimuli with HPF /s/.

A second set of models examined the ratings, this time including the stimuli with naturally produced /s/ in addition to the matched-guise stimuli. This comparison violated statistical assumptions somewhat, as the /s/ type was a fixed effect for the matched-guise stimuli and a random effect for the naturally produced /s/ tokens. The same model-building procedure was used, and again, the best fitting model had significant effects of /s/ type and age in years on ratings. When the [θ]-like ratings were used as the reference level, the effects for both of the other factors were significant. The factor was positive for the natural stimuli and negative for the HPF /s/, indicating that the natural /s/ stimuli were rated as more boy-like than the stimuli with [θ]-like /s/, which were more boy-like than those with HPF /s/.

The next set of models examined whether additional improvements in model fit could be found when adding acoustic characteristics of the VC sequences of the matched-guise tokens. The f1, f2, and f0 of the VCs were added to then model using the same general scheme as earlier. We tested whether these variables, alone or in interaction with the fixed effect of /s/ type, improved model fit. The best-fitting model was one in which f1 and f2 frequencies were included. Models with f0 and with an interaction between the fixed effect of /s/ type and f1 and f2 did not improve model fit. As expected, the coefficient for f2 was negative ($\beta = -0.117$, SEM=0.008, $z=-14.36$, $p<0.001$): higher values were associated with more girl-like speech. However, the coefficient for f1 was positive ($\beta = 0.026$, SEM=0.008, $z=3.19$, $p=0.0014$): higher values of f1 were associated with ratings of more boy-like speech. While the specific reason for this finding is not clear from these data, they show that listeners' ratings in this case did not merely reflect judgments based on adult sex differences in vocal-tract resonances.

3.2. Naturally Produced Tokens

The next analysis examined the influence of /s/ type for the natural tokens. As in the analyses in section 3.1, progressively more complex models were built, and models were retained only if they improved fit beyond the next simplest model. The predictors m1, m2, f1, f2, and f0 were added in that order. The most complex model beyond which no further improvement in fit was noted included m1, m2, f2, and all possible interactions among those variables. Adding f1 or f0, or interactions between f1, f0, and the other variables did not improve model fit. The coefficients for the full model are shown below. All of the coefficents in this model are significant at $p<0.001$ except m1*m2, for which $p=0.059$.

Table 1: Coefficients for the full model predicting gender judgments from acoustic characteristics of the naturally produced stimuli

Factor	β	SEM
m1	-0.098	0.013
f2	-0.097	0.013
m2	-0.108	0.014
m1*f2	-0.060	0.018
m1*m2	0.021	0.011
f2*m2	-0.061	0.015
m1*f2*m2	-0.080	0.018

All of the significant coefficients are negative: higher peak frequency /s/ (as indicated by a higher m1), more spectrally spread /s/ (i.e., more [θ]-like /s/, as indicated by a higher m2) are associated with judgments of more girl-like speech. Moreover, higher f2 is associated with more girl-like speech, as in section 3.2. Combinations of these factors only served to strengthen these relationships. That is, the negative coefficient in the interaction between m1 and f2 indicate that the association between m1 and more girl-like speech is greater with increasing f2 values.

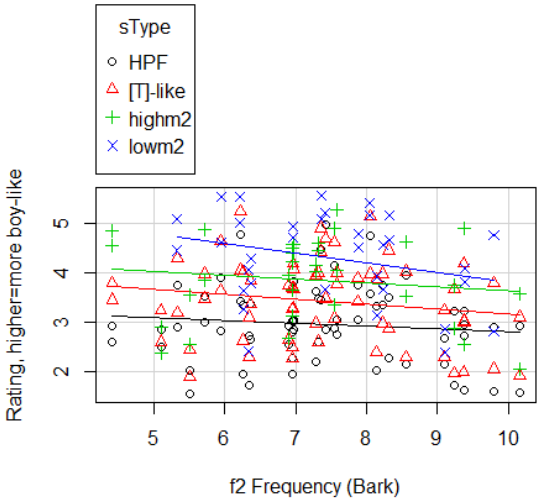

Figure 2: The relationship between f2 frequency and perceived gender. The naturally produced stimuli with a relatively higher m2 frequency (*highm2*) are plotted separately from those with a relatively lower m2 frequency (*lowm2*). The ratings for the matched-guise stimuli with the high-peak frequency (HPF) /s/ (*HPF*) and [θ]-like /s/ (*[T]-like*) are shown for comparison.

Figure 2 and shows the relationship between F2 frequency and ratings for the natural stimuli and for the same productions with matched-guise /s/ tokens. As these figures show, higher F2 frequency was associated with judgments that the talker's voices sounded more female. They also illustrate the influences of /s/ type discussed previously. In the case of m2, they show an interaction between m2 and f2 for the natural tokens. Specifically, they show that f2 frequency was more strongly related to gender judgments for stimuli that had an especially compact spectrum, i.e., an especially low m2. The combination of low f2 and low m2 led to talkers being the most likely to be judged as male sounding.

4. DISCUSSION

The results of this study show that adults' judgments of the gender typicality of boys' voices are affected by the acoustic characteristics of /s/. The specific way that /s/ affects judgments resembles the patterns seen in studies of /s/ variation and judgments of adult men's voices, such as [6] and [8]. Specifically, they show that variants of /s/ with an especially high frequency and variants of /s/ that resemble [θ] cue listeners to judge the voice as more girl-like, much in the way that such variants cued listeners in previous studies to judge male adult voices as less masculine sounding. A similar result was found in stimuli that varied naturally between a HPF-/s/ and a [θ]-like /s/. The results do contrast with those [6] in one interesting way. In that study, the ratings of gender typicality were equivalent for HPF and [θ]-like /s/, while in the currently study, [θ]-like /s/ was given a rating intermediate between the naturally produced stimuli (even those with very [θ]-like natural productions of /s/) and HPF /s/. One possibility is that this reflects the different scales used. The scale in [6] asked about masculinity of adult voices, while the scale in this study asked listeners to guess the talker's gender. The differences across studies might reflect [θ]-like productions more strongly cueing non-prototypically male voices than to cueing female gender. The differences between the ratings of the [θ]-like /s/ tokens and the ratings of the HPF /s/ tokens have interesting implications for the models that children receive during remediation for frontal misarticulation. Previous research (i.e., [5]) has shown that talkers hyperarticulate /s/ during the production of intentionally clear speech. If this type of hyperarticulation is present in the input that children get during speech therapy, then children with [θ]-like /s/ productions are effectively being taught to produce a HPF /s/. Given that HPF /s/ is associated with more female-sounding speech than is [θ]-like /s/, we might imagine that boys with a strong male gender identity might resist learning this variant. Future research should examine this directly in sound-learning experiments and *in situ*.

5. REFERENCES

[1] Bachorowski, J., & Owren, M. (1999). Acoustic correlates of talker sex and individual talker identity are present in a short vowel segment produced in running speech. *Journal of the Acoustical Society of America, 106*, 1054-1063.

[2] Fuchs, S., & Toda, M. (2010). Do differences in male versus female /s/ reflect biological or sociophonetic factors? In S. Fuchs, M. Toda, & M. Zygis (Eds.), *Turbulent Sounds: An Interdisciplinary Guide* (p. 281-302). New York: Walter de Gruyter.

[3] Johnson, K. (2006). Resonance in an exemplar-based lexicon: the emergence of social identity and phonology. *Journal of Phonetics, 34,* 485-499.

[4] Jongman, A., Wayland, R., &Wong, S. (2000). Acoustic characteristics of English fricatives. *Journal of the Acoustical Society of America, 108*, 1252-1263.

[5] Julien, H and Munson, B. (2012). Modifying Speech to Children based on their Perceived Phonetic Accuracy. *Journal of Speech, Language, and Hearing Research, 55*, 1836-1849.

[6] Mack, S., & Munson, B. (2012). The association between /s/ quality and perceived sexual orientation of men's voices: implicit and explicit measures. *Journal of Phonetics, 40*, 198–212.

[7] Munson, B., & Babel, M. (2007). Loose lips and silver tongues: projecting sexual orientation through speech. *Linguistics and Language Compass 1*, 416-44

[8] Munson, B., McDonald, E.C., DeBoe, N.L., & White, A.R. (2006). The acoustic and perceptual bases of judgments of women and men's sexual orientation from read speech. *Journal of Phonetics, 34*, 202-240.

[9] Perry, T.L., Ohde, R., & Ashmead, D. (2001). The acoustic bases for gender identification from children's voices. *Journal of the Acoustical Society of America, 109*, 2988-2998.

[10] Smit, A. (1993). Phonologic error distributions in the Iowa-Nebraska articulation norms project: consonant singletons. *Journal of Speech and Hearing Research, 36*, 533-547.

[11] Stuart-Smith, J. (2007). Empirical evidence for gendered speech production: /s/in Glaswegian. In J. Cole & J. Hualde (Eds.), *Laboratory Phonology 9* (pp. 65-86). Berlin: Mouton de Gruyter.