Gender effects on phonetic variation and speaking styles
A literature study

GSLT Speech Technology Term Paper, autumn 2006

Yvonne Samuelsson
Department of Linguistics
Stockholm University

1 Introduction

Even though there has been some work done on analysis of female speech in the past, it is clear that the focal point has been the male voice. In the early eighties [Johansson et al. 1982] stated that "[c]omparatively little is known about the characteristics of the female voice as compared with the male voice." The reason for this is the high fundamental frequency range of the female voice, which "makes formant frequency estimates uncertain and, hence, information on the voice source unsafe." Almost a decade later [Titze 1989] says that "[m]uch of our knowledge about speech production comes from studies on male speakers" and [Klatt and Klatt 1990] add that "[w]omen and children have been somewhat neglected groups in the history of speech analysis by machine". Yet another decade later [Henton 1999] says that little has been done with regards to synthesis of female voices and that this is due to too little data on female speech production and that female voices have been difficult to analyze acoustically, because of "inadequacies in analytic hardware". Like in many other areas of society, men are still more well studied and they are regarded as the standard, the baseline to which everything else is compared.

This paper is a study of some of the literature on gender differences in speech. We will look at what differences there are and what kind of explanations have been given for these differences. How is speech production and perception affected by gender? Are these differences mainly biological/physiological or social/behavioural? Our focus will be adults, even though many of the studies also look at children. We will look at F_0, vowel formants and some temporal aspects.

Some of the papers use the term sex and others the term gender. In general, gender is used to refer to social categories, while sex is used for biological categories. In this paper I will mostly use the term gender, without distinguishing between these aspects, unless explicitly stated. This is because some effects on speech may be biological and some social, and it may be difficult to see which has the larger influence.

2 F_0

There has been quite a lot of work done on the frequency of the voice fundamental (F_0) in speech. [Traunmüller and Eriksson 1995] looked at a number of investigations of F_0-variation and concluded that in most cases the average F_0 was higher and the F_0-range in Hz wider for women than for men. The published data regarded several languages and various types of discourse and typical values for F_0 are 120Hz for men and 210Hz for women.

[Titze 1989] showed that analysis of female speech involves more than a simple scaling factor of the fundamental frequency. Instead there is a scaling factor based on the membranous vocal fold length, which he says accounts for several differences between male and female voices, including fundamental frequency.

[Traunmüller and Eriksson 1995], in their analysis of F_0-values, actually give a formula for transforming typical male speech into female speech (with regards to F_0), but add that this says nothing of the perception of the synthesised speech. "'Typical' female speech might still be perceived as more lively or less lively than 'typical' male speech."

This would be in accordance with [Henton 1999], who states that it is not possible to make changes to the sound spectrograph to make the patterns of men and women more similar, since "in many respects, their respective patterns are simply *not* similar. No amount of computational manipulation (or distortion) will render Eve's voice from Adam's larynx".

It seems that there are many more differences between male and female voices than just fundamental frequency values. One experiment by [Johnson et al. 1999], exploring gender expectations in vowel perception, where the expectations were manipulated both visually and acoustically, showed that judging a voice as typically female was not related to fundamental frequency, but rather to phonation type (breathiness).

However, [Traunmüller et al. 2003] come to a different conclusion. They used an articulatory model simulating the vocal tracts of speakers of varying age, varying F_0 to synthesise stimuli used in a perceptional experiment of French. The subjects hade to judge vowel identity, vocal effort and speaker age and gender. In their experiments, the perception of speaker gender appeared to be mainly based on F_0.

[Traunmüller and Eriksson 1995] say that the extent of the F_0-excursions is influenced by both linguistic factors, reflected in the language or text, and paralinguistic factors. [Traunmüller 1997] states that "the acoustic properties of speech sounds vary not only because of linguistic factors, but also as a function of organic, expressive, and transmittal factors". Some local paralinguistic factors are placement of focal and contrastive stress and some global factors are attitude and emotion. Expressed in semitones or as a percentage of F_0, the between speaker variation in average F_0 is generally reported to be higher for men than for women. [Traunmüller and Eriksson 1995] propose that one reason for this might be that there are more smokers among men than women and smoking lowers the mean F_0 and increases the F_0 range.

3 Vowels

Let us turn to vowel formants and vowel space. [Simpson 2001] says that "[i]t has long been recognized that it is not possible to derive the formant values of male vowels from their female equivalents (or vice versa) by using a single scale factor based on an average female-male vocal tract length difference of 20%." Not only would we need different scale factors for each formant and vowel category, it also turns out that the vowel space ($F_1 \times F_2$) is consistently larger for women than for men.

Simpson ([Simpson 2000, Simpson 2001]) says that several reasons for this non-uniform scaling have been proposed, like anatomical reasons (gender-specific ratios of oral-pharyngal cavity length), various interactions between F_0 and formant frequencies

and sociophonetic reasons.

[Maurer et al. 1992] suggest a direct relationship between formant patterns and F_0. They examined whether differences in formant patterns were due to differences in sex (and age) or due to differences in F_0. If the differences in patterns for men and women are connected to the difference in F_0 then the differences should disapperar when F_0 is identical. In general, they found no differences in formant patterns for the same F_0, but there were differences within a speaker group for different F_0. Their experiments showed that formant frequencies generally rise with F_0. For German they could not detect any gender differences for F_1 when F_0 was the same, and only some differences for F_2 (not in back vowels). There were, however, differences for F_3. Above 2kHz all the formant frequencies of women showed higher means than the men's. This means that there are gender differences in higher formant frequencies.

[Diehl et al. 1996] agree that there is a non-uniform scaling between male and female vowels, stating that adult female vocal tracts tend to be shorter than those of adult males, which leads to higher frequencies for female formants. They try to investigate what the basis of this non-uniform scaling is, with the hypothesis that all else being equal, higher F_0 gives reduced vowel identifiability because of sparser harmonic sampling. Women do not aim to sound as if they have smaller vocal tracts than they do, but to produce vowels more dispersed in the $F_1 \times F_2$ space than those of males because of this identifiability problem. Two experiments showed that an increase of F_0, at least in the region beyond 150Hz, reduced the vowel labeling accuracy.

In [Henton 1995] data from six phonetic studies on seven languages and dialects are discussed. She concludes that women produce more open-mouthed variants of vowels than men, which means that female speech is more phonetically explicit. She sees this in a socio-phonetic light, where greater articulatory distinctions may be the standard or prestige forms, which women try to guard, while men use more non-standard forms.

[Diehl et al. 1996], however, say that even though clearer speech, particularly in the form of more dispersed vowels, is considered more 'feminine' in most cultures, this may be more of an effect than a cause. [Simpson 2000] agrees, in his discussion of the possibility that male speakers may have to travel greater articulatory distances than female speakers to reach analogous phonetic targets. Similar articulatory distances between vowel categories will be acoustically further apart for women. Experiments with lingual pellets showed that women cover a greater acoustic space both in linear (Hz) and non-linear (Bark) terms. Males in some respects cover a greater articulatory distance. He concludes that male and female articulatory spaces have different size and this stands in an inverse relationship to the size of their acoustic products. Men and women use different articulatory speeds to produce similar perceptual forms. "The implications of such dynamic differences are wideranging. Phonetic correlates of clarity, often attributed to female speakers, such as more widely distributed acoustic vowel spaces and greater vowel duration, may be nothing more than by-products of reconciling differences in articulatory dimension."

[Johnson et al. 1999] carried out experiments using auditory-visual mismatches to compare several approaches to speaker normalization in speech perception. Speaker normalization is the mechanism in speech perception which recovers the 'same' words or sounds from speech produced by different talkers. Their first experiment explored gender expectations in vowel perception, where the expectations were manipulated both visually

and acoustically and showed, as mentioned before, that a voice was perceived as female based on phonation type rather than fundamental frequency. It also turned out that vowel boundaries shifted due to visual stimuli.

The results from their second experiment suggested that listeners integrate audio visual information for both vowels and talkers. Their third experiment explored the effects of an abstract talker representation in the perceptual identification of gender-ambiguous vowels, by asking one group of listeners to imagine a female talker and one group to imagine a male talker. The boundary of the vowels was sensitive to instructions about the identity of the talker, suggesting that some part of talker normalization is due to expectations about male and female voices that listeners have in speech perception.

[Johnson et al. 1999] conclude that the results indicate that speaker normalization in speech perception is based on abstract talker representations and that talker identity is perceived from several cues in the listening situation, like direct acustic cues for vocal tract length (formants), indirect cues like F_0 and mode of vocal fold phonation, visual cues and imagined talker characteristics.

4 Duration and temporal effects

There is a growing amount of data from a number of languages, which show gender differences with regard to duration. There are two main patterns, that female vowels are longer than male and that women produce greater differences between long and short vowel categories ([Simpson and Ericsdotter 2003, Simpson 2003]).

Some studies try to explain this with sociophonetics, that longer vowels and greater durational differences are associated with speaking more clearly, which is regarded as female (see [Simpson and Ericsdotter 2003, Simpson 2003] and [Henton 1995]). There have also been several studies that attribute this phenomenon to physiology.

[Simpson 2000] says that since men may have to travel greater articulatory distances than women to reach analogous phonetic targets, this would mean that women can maintain shorter vowel durations and an increase in overall utterance tempo. Men and women want to achieve similar perceptual products, which may require differences in articulatory speed. It would make men move faster and women move slower.

[Simpson 2003] examined diphthong durations for English and found that the mean acoustic duration of the female diphthong is 10% greater than the male, even though the utterances containing these tokens were not significantly different for men and women. The mean duration of the tongue body movement was slightly longer for male speakers.

[Byrd 1992] discusses vowel reduction, which is known to be affected by speech rate. Her experiments show that men, who speak faster, tend to reduce their vowels to schwa more often than women. She adds that there is a possibility that women use a different set of reduced vowels. The experiments of [Whiteside 1995, Whiteside 1996] showed lower rates of syllables per second for women, realising consonant clusters more fully. She also found that men tend to elide or reduce vowels and consonants, which leads to shorter sentence durations. She interprets the fact that female speech segments were on average longer than those of men as evidence that women tended to realise speech segments more fully, which would support the thesis that women enunciate more clearly.

When it comes to sentence duration, different experiments seem to give different re-

sults. [Simpson and Ericsdotter 2003] investigated American English and Central Standard Swedish. They found that differences in whole-utterance duration were the exception (and when they occured the male sentence duration was longer). Female vowels in focus (accented) where significantly longer than the male vowels in their experiments for Swedish, but men had greater durations in much of the consonantal material.

They discuss the sociophonetic explanation of females speaking more clearly and conclude that this would mean that women restrict their clarity to places of prominence and that they compensate for that with shorter durations at other places in the utterance. They say that another possible reason for gender-specific durational patterns is the consequences of differences in male and female articulatory dimensions. Men traverse greater distances to get to the same phonetic targets, but they do this with greater articulatory speeds. This may give a temporally shorter movement for men. They conclude that sex-specific durational differences at one point in an utterance must be compensated for elsewhere, because of the lack of differences in sentence duration. This means that there is not one single factor to derive the durations of one gender to the other. They add that their results are contrary to other findings for English and German, where female sentence durations were longer than male.

[Byrd 1992] investigated American English and found that gender had a significant effect on speaking rate (measured for two sentences). Women spoke reliably more slowly than men. [Whiteside 1995] looked at temporal gender differences in a northern British accent. The experiments showed that the sentence durations were longer for women than for men. The material also shows that women tend to read at a slower rate compared to men and the women displayed greater variability in their sentence duration ([Whiteside 1996]).

[Whiteside 1995] found a link between gender and pausing in some of her material. There was also a link between the occurence of pauses and longer duration values when words precede a pause. Further analysis, in [Whiteside 1996], showed a significant connection between speaker sex and pausing, women tending to pause more than men.

Men appeared to use F_0 shifts (declination in the F_0 pattern) to mark syntactic boundaries, while women used pauses more and when a pause was present there was an increase in the duration of the word and phonetic segments preceding the pause. Women signal syntactic boundaries through pausing and where no pauses are present they use phrase-final lengthening. Even in sentence final position women tend to produce longer phonetic realisations than men. In [Whiteside 1995] she says that this could point to men tending to dominate a conversation by not pausing, since it reduces turn taking and interruptions, while women pause more, allowing for interruptions.

5 Conclusions

We have looked at some of the literature about gender effects on phonetic variation and speaking styles, with focus on F_0, vowel formants and some temporal aspects.

F_0 is generally lower for men (around 120Hz) and higher for women (around 210Hz). Women also seem to have a larger F_0-range than men. Some experiments show a relationship between F_0 and perception of gender, while others do not. There seems to be a connection between F_0 and vocal fold length, but all in all other factors than F_0 also

seem to influence the perception of gender. This shows that F_0 is not the only distinction between male and femal voices.

Formant frequencies seem to generally rise with F_0, but there is no uniform scaling. We would need different factors for each formant. Additionally, the vowel space is larger for women. Some researchers say that this is because women speak more clearly and articulate more because, in a Labovian perspective, this is seen as the prestige or standard form, which women guard. Some experiments however showed that a higher F_0 makes it harder to distinguish vowels, which could be one explanation for a larger vowel space (more dispersed vowels) for women. Other experiments showed that men cover a greater articulatory distance while women cover a greater acoustic space to produce the similar acoustic products. There is some evidence that the perception of vowels is based on both visual and auditory cues, but also on our expectations about the gender of the speaker. This suggests that we use different ways of decoding the speech of men and women.

There seems to be a consensus on that women produce longer vowels than men, but some experiments show longer consonants for women while others do not (or even show longer consonants for men). The results differ also when it comes to sentence duration and gender, some experiments showing that women speak or read more slowly and have overall longer sentence durations, others not showing this relation at all. Some experiments showed that women tend to pause more than men, which was attributed to social behaviour, that women allow for interruptions by pausing, while men do not.

There are a few problems or reasons for caution when investigating gender differences in speech. One is that several of the experiments draw conclusions based on a few informants. This means that tendencies might be individual rather than gender-based. Other aspects that need to be discussed are differences that are due to differences between the type of speech investigated, like read and free speech, or differences between languages and dialects.

All in all there are many reasons for differences between male and female speech. In general it seems that women tend to have greater variation in their speech than men, on several levels. This is very interesting and is most likely the explanation for why female speech has been seen as more difficult to analyse. I would think that some of these differences definitely are based on physical sex and general differences in the vocal organs of men and women. But I also think that there are additional aspects based on social gender. All of this interacts and that is why it is so difficult to pinpoint what the differences between male and female voices really are.

References

[Byrd 1992] Dani Byrd. 1992. Preliminary results on speaker-dependent variation in the TIMIT database. *Journal of the Acoustical Society of America*, 92(1):593–596.

[Diehl et al. 1996] Randy L. Diehl, Björn Lindblom, Kathryn A. Hoemeke, and Richard P. Fahey. 1996. On explaining certain male-female differences in the phonetic realization of vowel categories. *Journal of Phonetics*, 24:187–208.

[Henton 1995] Caroline Henton. 1995. Cross-language variation in the vowels of female and male speakers. In *Proceedings of the XIIIth International Congress of Phonetic Sciences (ICPhS 95)*, pages 420–423.

[Henton 1999] Caroline Henton. 1999. Where is female synthetic speech? *Journal of the International Phonetic Association*, 29(1):51–61.

[Johansson et al. 1982] C. Johansson, J. Sundberg, and H. Wilbrand. 1982. X-ray study of articulation and formant frequencies in two female singers. *STL-QPSR*, 23(4):117–134.

[Johnson et al. 1999] Keith Johnson, Elisabeth A. Strand, and Mariapaola D'Imperio. 1999. Auditory-visual integration of talker gender in vowel perception. *Journal of Phonetics*, 27(4):359–384.

[Klatt and Klatt 1990] Dennis H. Klatt and Laura C. Klatt. 1990. Analysis, synthesis, and perception of voice quality variations among female and male talkers. *Journal of the Acoustical Society of America*, 87(2):820–857, February.

[Maurer et al. 1992] Dieter Maurer, Norman Cook, Theodor Landis, and Christian D'Heureuse. 1992. Are measured differences between the formants of men, women and children due to f0 differences? *Journal of the International Phonetic Association*, 21:66–79.

[Simpson and Ericsdotter 2003] Adrian P. Simpson and Christine Ericsdotter. 2003. Sex-specific durational differences in English and Swedish. In *Proc. XVth ICPhS*, pages 1113–1116, Barcelona.

[Simpson 2000] Adrian P. Simpson. 2000. Gender-specific differences in the articulatory and acoustic realization of interword vowel sequences in American English. In Phil Hoole, Masaaki Honda, and Christine Mooshammer, editors, *5th Seminar on Speech Production: Models and Data*, pages 209–212, Kloster Seeon.

[Simpson 2001] Adrian P. Simpson. 2001. Dynamic consequences of differences in male and female vocal tract dimensions. *Journal of the Acoustical Society of America*, 109(5):2153–2164.

[Simpson 2003] Adrian P. Simpson. 2003. Possible articulatory reasons for sex-specific differences in vowel duration. In *Proc. 6th International Seminar on Speech Production*, Sydney.

[Titze 1989] Ingo R. Titze. 1989. Physiologic and acoustic differences between male and female voices. *Journal of Acoustical Society of America*, 85(4):1699–1707.

[Traunmüller and Eriksson 1995] Hartmut Traunmüller and Anders Eriksson. 1995. The frequency range of the voice fundamental in the speech of male and female adults. Manuscript.

[Traunmüller et al. 2003] Hartmut Traunmüller, Anders Eriksson, and Lucie Ménard. 2003. Perception of speaker age, sex and vowel quality investigated using stimuli produced with an articulatory model. In *Proceedings of the XVth International Congress of Phonetic Sciences (ICPhS 03)*, pages 1739–1742.

[Traunmüller 1997] Hartmut Traunmüller. 1997. Perception of speaker sex, age, and vocal effort. *Phonum*, 4:183–186.

[Whiteside 1995] Sandra P. Whiteside. 1995. Temporal-based speaker sex differences in read speech: A sociophonetic approach. In *Proceedings of the XIIIth International Congress of Phonetic Sciences (ICPhS 95)*, volume 3, pages 516–519.

[Whiteside 1996] Sandra P. Whiteside. 1996. Temporal-based acoustic-phonetic patterns in read speech: Some evidence for speaker sex differences. *Journal of the International Phonetic Association*, 26(1):23–40.