Language and Linguistics Compass 3/2 (2009): 621–640, 10.1111/j.1749-818x.2009.00125.x

Phonetic differences between male and female speech

Adrian P. Simpson*
Institut für Germanistische Sprachwissenschaft, Friedrich-Schiller-Universität Jena

Abstract
The main phonetic differences between the speech of male and female speakers are described and explanations that have been offered to account for these differences are critically discussed.

Introduction

As soon as we hear a new voice we do much more than just understanding the message it contains. We make judgements about how old somebody is or where they come from. We also try to ascertain whether we are listening to a man or a woman. Often we are only conscious that we are trying to make this judgement when we realise we have made an incorrect assignment or are confronted with an ambiguous voice.

When asked why we think we are listening to a male or a female voice, the first reason we give is the pitch of the voice, the male voice being on average lower in pitch. However, several differences have been found between male and female voices. In this article, we concentrate on describing phonetic differences between male and female voices, in other words, those differences that relate to the way in which sound is produced and perceived. Besides describing the differences, we will examine some of the explanations that researchers have offered to account for these differences. These are divided into two main types: biophysical inevitabilities and learned behaviours, more popularly referred to as nature and nurture. So, while there are a number of biological differences between males and females that have consequences for the sounds they produce, such as the dimension of the mouth, throat and vocal folds, it is also clear that we produce certain speech patterns appropriate to the gender we identify with. This is brought into particularly sharp focus when we look at the differences in the speech of preadolescent boys and girls who do not exhibit any significant anatomical differences prior to the onset of puberty (see below). Many other aspects of speech can also only be properly accounted for when we consider both biophysical inevitabilities and learned behaviours. Co-articulation is perhaps the most pervasive aspect of speech production, describing the way in which different aspects of sounds influence each other (Hardcastle and Hewlett 2006). So, for instance, the dorsal plosive at the beginning of the word 'key' is produced further forward in the mouth than the plosive in the word 'cart'. This is generally explained by referring to the different tongue positions of the vowels, that of 'key' being a front vowel, that in 'cart' being articulated further back. While many aspects of co-articulation can be explained in terms of articulatory accommodations of adjacent sounds, there are also many differences in the details of co-articulation from one language to another, such as vowel to vowel co-articulation (Manuel 1990, 2006), which can only be accounted for if we assume speakers learn different co-articulatory behaviours from one language to another.

This article will also expose certain androcentric aspects of phonetic endeavour. All too often, research has been carried out by men analysing male voices. The description of the female voice has often been carried out with reference to the male voice, not the other way round. The reader may like to read Henton (1992) for an informative corrective to this androcentricity.

To understand fully many of the phonetic differences between male and female speech and the explanations that have been offered to account for them, the reader should have a basic understanding of speech acoustics. The most appropriate introduction, especially for those with a background in the humanities, is Ladefoged (1996). This carefully takes the reader through most of the basic concepts of speech acoustics and should provide a more than adequate basis for the ideas discussed here.

Phonation and Pitch

All air entering and leaving the lungs passes through the vocal folds, which form a valve across the top of the trachea (windpipe). When brought together and appropriately tensed, air passing between the vocal folds causes them to vibrate, or more accurately, they open and close in quick succession. The frequency of vibration or fundamental frequency (henceforth F0) is tightly correlated with the perceived pitch of the voice. Male vocal folds tend to be longer and thicker than female vocal folds causing them to vibrate more slowly. Male speakers of languages such as German and English have an average F0 of 100–120 Hz (Hertz, cycles per second). Female vocal folds are shorter and lighter and vibrate at approximately twice the male frequency (200–220 Hz).

Besides vibrating at different frequencies, differences have also been found in the male and female voice quality caused by the way the vocal folds vibrate. Studies of English speakers have found two main differences in the use of voice quality: (1) male speakers use more creaky voice[1] than females, as shown by Henton and Bladon (1988) for two varieties of British English, and (2) female speakers have breathier voice quality than males, as shown by Henton and Bladon (1985) and Klatt and Klatt (1990) for British and American English, respectively. At first sight, it would seem that both differences are sociophonetic, that is, learned behaviours and Henton and Bladon have argued that females use a more breathy voice quality, which is commonly associated with increased intimacy, to make themselves more desirable to the opposite sex.

However, while there do not appear to be any anatomical or physiological reasons why males should use more creaky voice than females, the picture is less clear with breathy voice. During each cycle of normal vibration the vocal folds come together and briefly close the airway. In breathy voice, however, the vocal folds do not close completely during the cycle, so that there is a constant flow of air through the glottis. Titze's (1989) computational model of vocal fold vibration predicts that male vocal folds will completely close the glottis, briefly shutting off the airflow during each cycle. In contrast, the vertically thinner female vocal folds in the model never make complete closure during each opening-closing cycle creating a constant airflow, that is, breathy voice.

Another well-documented and at the same time poorly understood difference between male and female voice is the use of pitch. While it is uncontroversial that the average male pitch is lower than the average female pitch, there is disagreement as to whether males and females differ in pitch range, in other words, how large the difference is between the maximum and minimum pitch used in an utterance. There are two main reasons for this controversy. The first resides in the way in which we represent F0. The second resides in the types of speech activity we analyse.

F0 is often expressed in Hertz by calculating the fundamental frequency of vocal fold vibration, that is, the number of times the vocal folds open and close in one second. While this is a simple physical measurement, it is less appropriate as a way of describing how a listener perceives the pitch of the sound produced. This is because sound perception, in particular the way in which the ear analyses sound, is non-linear. Put simply, what we perceive as equal jumps in pitch, are physically different. The higher the pitch becomes the greater the physical difference between two tones has to be for the perceived difference to remain the same. For this reason, Hertz measurements are often converted into other units of measurement that more appropriately reflect our perception of frequency. One unit of measurement that is often used in pitch studies is the semitone. This is a logarithmic measure. Specifically, a doubling of frequency, that is, 100–200 Hz or 200–400 Hz is represented by an equal semitone interval. The diagrams in Figure 1 illustrate well the difference between the representation of F0 in (a) linear units (Hertz) and (b) non-linear units (semitones). The curves in the diagram represent the change in fundamental frequency over time for the German utterance 'Riecht ihr nicht die frische Luft?' ('Can't you smell the fresh air?') produced by a female speaker (green) and a male speaker (red). The upward movement of the

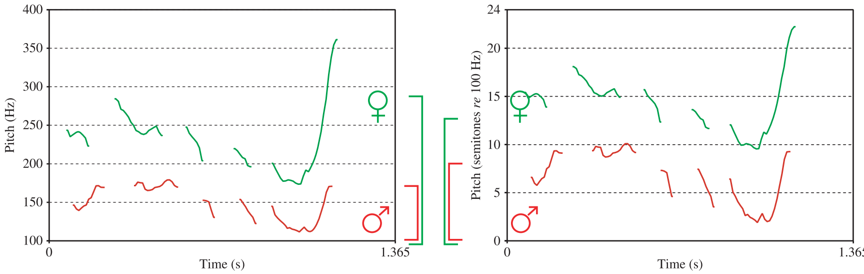

Fig. 1. Linear Hertz (left) vs. non-linear semitone (right) representation of the fundamental frequency contour for the German sentence 'Riechst du nicht die frische Luft?' spoken by a female speaker (green) and a male speaker (red). Vertical lines in the centre of the figure compare the female and male pitch range in both representations.

curve at the end reflects the sharp rise in F0 over the last syllable 'Luft' ('air'). Gaps in the curve represent stretches of voicelessness, that is, when the vocal folds are not vibrating. The chief difference to notice in Figure 1 is that in the linear Hertz representation in (a) the vertical excursions of the male F0 movements appear to be very much smaller than the female ones. In contrast, in the perceptually more appropriate semitone representation in (b), the male and female F0 movements have become more similar in size. The vertical red and green lines in the middle of the two diagrams show the F0 range for the male and female speakers in Hertz (left) and semitones (right). These lines have been aligned to facilitate comparison. However, this representation not only provides us with a better approximation to the way in which we perceive male and female pitch, it is important to provide an appropriate representation of male and female pitch for a balanced discussion of the speech differences between the sexes. Henton (1989) shows that an inappropriate (i.e. linear Hertz) representation of F0 has frequently been used in scientific studies to substantiate claims about typical female voice stereotypes, such as speaking more emotionally, or having 'swoopy' pitch movements. By using results from her own investigation as well recalculating Hertz values from older studies Henton shows in terms of semitones, males and female speakers of English do not exhibit any significant differences in pitch range. The results of a large, but little known study of more than a thousand male and female speakers present a similar picture for German (Herbst 1964, 1969), showing, in fact, that it is male speakers who have a larger pitch range.

But even the appropriate numerical representation of F0 is not the whole story. As we can see from Figure 1, the female F0 range for this utterance is larger than the male's in both representations. This difference is due chiefly to the F0 rise over the last syllable 'Luft'. Is this just chance variation? Maybe not. Henton's (1989) empirical study, as well as the data from the other studies that she recalculated, were mainly restricted to carefully elicited declarative speech involving a falling pitch pattern. In a detailed study of gender-related intonation differences in Dutch (Haan and van Heuven 1999; Haan 2002), a large sample of interrogative utterances involving rising pitch patterns was also studied. The main reason for widening the analysis to include interrogative utterances are gender-specific communicative differences. Specifically, 'female speech has been shown to be more expressive, more involved, more listener-directed [. . .] than male speech' (Haan and van Heuven 1999: 1581). The analysis of statements, and different types of questions in both controlled elicitation and from spontaneous patient-doctor interviews showed that female speakers repeatedly produced a wider pitch range than male speakers.

When we compare average female-male F0 differences from a range of languages, it soon becomes clear that even average fundamental frequency of the voice is in part learned. Traunmüller and Eriksson (1995) bring together average female-male F0 values from studies of a number of European and non-European languages. While some cross-linguistic differences may be attributable to differences in measurement and the material recorded, others are not. Most striking are the findings from Wù dialects of Chinese (Rose 1991), in which the average male and female F0 is 170 and 187 Hz, respectively. But also in studies of European languages marked differences in average F0 have been found. For English, Takefuta et al. (1972) measured male and female averages of 127 and 186 Hz, respectively, whereas in a study of 60 French speakers, Boë et al. (1975) found averages of 118 Hz for men and 207 Hz for women. Since it would be unreasonable to account for such large differences in terms of anatomical differences in the populations being investigated, part of the difference must be attributed to learned behaviours.

In summary, as with all the other aspects we shall examine, differences in vocal fold anatomy and physiology account for only some of the differences that have been observed in male and female voices.

Articulation

VOWELS

Apart from differences in voice pitch and voice quality arising from differences in the size and shape of the vocal folds, there are also differences in the dimensions of the vocal tract above the glottis that have important consequences for the sound produced. Perhaps the single most important difference between males and females in this respect is vocal tract length. The average length of the adult female vocal tract, that is, the distance from the vocal folds to the lips, is on average 14–14.5 cm. The average male vocal tract is 17–18 cm. This difference chiefly arises from an

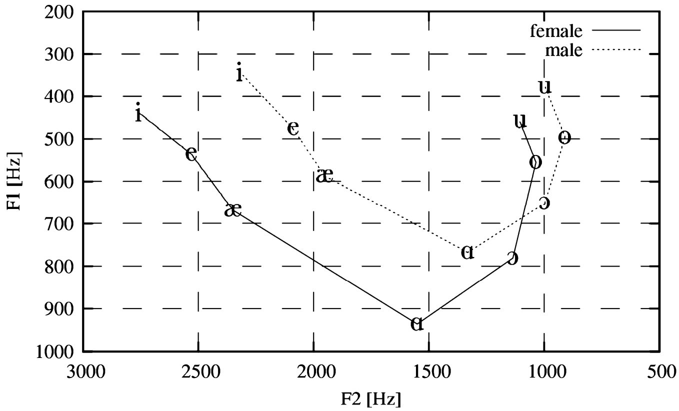

Fig. 2. Average male (dashed) and female (solid) acoustic vowel spaces from Hillenbrand et al. (1995).

increase in the length of the pharynx caused by the male larynx lowering during puberty. These differences in vocal tract length have consequences for certain aspects of sound production. The vocal tract modifies the tone being produced at the glottis. Due to the resonant frequencies (formants) of a particular vocal tract configuration, certain frequency components (overtones, harmonics) of the tone produced at the glottis are strengthened. We hear different vowel qualities, that is, an [i] or [e] or [o], depending on their different vocal tract configurations and hence on the different frequency components that are strengthened.

While the vocal tract length differences explain gross differences between the formant frequencies of male and female vowels, that is, female vowels have higher formant frequencies than males, it cannot explain the size of the acoustic differences between different male and female vowel qualities. This problem is illustrated in Figure 2. The diagram illustrates a common method of displaying acoustic vowel qualities. The acoustic quality of a vowel is mainly determined by the frequency of the first two formants, F1 and F2. In the diagram F2 of each vowel is plotted as a function of its F1 value. With the axes reversed (F1 values from top to bottom, F2 values from right to left), the acoustic vowel space becomes visually similar to that of the pseudo-articulatory vowel space used in descriptive phonetics. As we would predict from differences in vocal tract length, each male vowel quality has lower formant frequencies than its female congener. However, there is one further difference that is apparent from Figure 2: the female vowel space is larger than the male space, that is, the female vowel qualities stake out a larger acoustic area. The back rounded vowel qualities [u] and [o] are relatively close together; female [i] differs from male [i] mainly in the F2 dimension and the female open vowel [a] differs from the corresponding male vowel chiefly in the F1 dimension. In other words, the male and female qualities differ from each other depending on the vowel involved. This is interesting because if men and women were producing the same vowel qualities and only differed because they had different vocal tract lengths, it should be possible to map one set of vowels onto another set using a single constant. As we can see from Figure 2, this is not the case. The differences are non-uniform: it requires a different constant for each formant of each vowel. Although the values shown in Figure 2 are taken from American English, similar non-uniform differences in male and female vowel systems have been found in a several different languages.

Although non-uniform differences in male and female vowel articulation have attracted a good deal of attention over the last 50 years, the reasons for the differences are still far from clear. In part, this is because the differences are not attributable to a single factor. Furthermore, many of the reasons that have been proposed for the non-uniform differences have not successfully been proven to be true or false. We will try to assess each of these reasons critically.

ARTICULATORY DIMENSIONS

Fant (1966, 1975) analyses vowel data from a number of languages (Swedish, American English, Danish, Serbo-Croatian, Dutch and Estonian). He suggests that the non-uniformity is brought about in part by anatomical differences between male and female vocal tracts, male speakers having larger laryngeal cavities and a proportionally longer pharynx than female speakers (Chiba and Kajiyama 1941), which explains in particular the larger scale factors needed for open vowel categories.

Nordström (1977) also considers the longer male pharynx to be the main factor causing the non-uniform differences. Goldstein (1980) brings together growth data from a wide range of sources that allow modelling of both vocal tract growth and gender-specific differences. In similar vein, Traunmüller (1984) considers the intriguing possibility that non-uniformity might arise because male speakers continue to use the motor commands they were using before puberty even after the pharynx has lengthened and the larynx has enlarged after puberty. Using model calculations, Nordström, Goldstein and Traunmüller are successful in producing non-uniform differences for different vowel categories, but all three researchers interpret their results as accounting only partially for observed male-female differences.

With regard to the relatively small differences between male and female back rounded vowels [o] and [u], Fant (1966: 25) suggests this may be due to female speakers making compensatory articulations in order to get closer to the male qualities. In simple acoustic terms, it is possible to think of the resonant properties of a back rounded vowel as being similar to two Helmholtz resonators linked together. In a vowel like [a], we can think of the vocal tract comprising two tubes linked to one another, one representing the pharynx, the other the oral cavity. The first resonant frequencies of these two tubes, which determine the first two formants of the vowel, depend crucially on the length of the tubes, and therefore on the overall length of the vocal tract. In a Helmholtz resonator, however, it is not the length but rather two other factors that determine the resonant frequencies: (a) the volume, and (b) the length and diameter of the neck of resonator. If the length and the diameter of the neck are reduced, then the first resonant frequency is reduced. In articulatory terms, Fant suggests that females make tighter and longer strictures between the lips as well as between the back of the tongue and the soft palate, causing a reduction in both F1 and F2 of [o] and [u], bringing them closer to male values.

ARTICULATORY SPEED

In considering the differences between average male and female vowel systems, it is easy to forget that the average measurements derive from individual measurements of individual vowels being produced in the ever changing flow of speech in which the articulators (tongue, lips, velum, etc.) are constantly moving. Simpson (2001, 2002) investigates the possibility that differences in the average articulatory dimensions of male and females might have consequences for the average size of the acoustic vowel space, in particular predicting a larger female acoustic vowel space.

Consider Figure 3. This diagram shows superimposed male and female vocal tract outlines for the vowels [i] and [ɑ]. The 'x' and '+' mark a central point of the tongue for each of male and female vowel positions. The diagrams are modified from Goldstein (1980), which in turn are based on a collection of growth data taken from a number of sources. From a careful examination of the schematic outlines of the male and female vowels in Figure 3 two things are apparent. First, we can see that the male (dashed lines) vocal tract is longer than the female vocal tract. Second, and more importantly for the discussion here, the distance that the male ('+') tongue has to travel to get from [i] to [ɑ] is approximately 11% more than it is for the female tongue ('x').

From these dimensional differences, we may hypothesise several consequences for articulatory and acoustic differences between male and female speech, depending on the initial assumptions we make. Most importantly for the discussion here is the possibility that if males and females were to produce their utterances in approximately the same time frame, that is, the vowels and consonants have the same duration, a female moving her articulators at the same speed as a male should be able to reach more extreme articulatory and, of course, acoustic targets. From

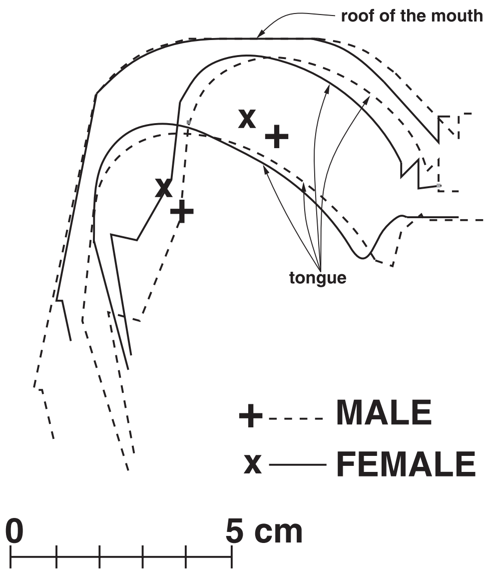

Fig. 3. Schematic representation of tongue positions for the vowels [i] (tongue close to hard palate) and [ɑ] (open tongue position) for a female ('x') and a male ('+') speaker.

this, we could predict that the female acoustic vowel space might be larger because females can reach more extreme vowel positions in the same time. In particular, this would predict the relatively more open position of female open vowels that has been repeatedly observed in languages.

Although these hypothesised differences are intuitively appealing, the examination of authentic articulatory data presents us, as always, with a more complex picture. Simpson (2001, 2002) uses articulatory data from 26 female and 22 male speakers in the University of Wisconsin X-ray Microbeam Speech Production Database[2] (Westbury 1994) to examine male and female differences in the tongue movements in the diphthong in 'light' and in the vowel sequence in 'th*ey a*ll'. In the diphthong, the tongue moves from an open to a close vowel position, in the vowel sequence, the tongue begins high and front in the mouth and moves back. Simpson found that male speakers did travel greater articulatory distances and did so at higher speeds. However, despite this, the acoustic product of the greater articulatory movements were smaller. So, for example, although the female speakers travelled a shorter articulatory distance in the diphthong [aɪ] in 'light', the acoustic space (change in F1 and F2) traversed was greater. Furthermore, and we shall return to this below, the female diphthong and vowel sequence were both found to be longer in duration than those of male speakers, something we might not expect given the shorter distance to be travelled.

INTERACTION OF PITCH AND ARTICULATION

As described above, when we speak we use an airflow, generally from the lungs, to create one or more sound sources at the vocal folds or with the articulators above the glottis. The sound produced is then modified by particular configurations of the pharynx, tongue, velum and lips. So, for instance, in the case of a vowel, air passing between the vocal folds causes them to open and close in quick succession, creating a pattern of fluctuations in sound pressure that we hear as a tone with a particular pitch. The vowel quality that we hear being produced on a particular pitch is dependent on the position of the tongue in the mouth and the configuration of the lips. Different configurations of the tongue and lips cause different harmonics in the signal being produced at the glottis to be strengthened. So, for instance, if the tongue is high up in the front of the mouth and the lips are spread [i] is produced; if the tongue is raised up high in the back of the mouth and the lips are rounded [u] is produced.

From the point of view of producing sound, then, the sound source of voicing and the vowel configuration being articulated should be independent of one another. However, some researchers have suggested that the larger female acoustic vowel space may be related to the higher average pitch of the female voice (Goldstein 1980; Ryalls and Lieberman 1982; Diehl et al. 1996). The line of argument goes something like this. The vowel quality (e.g. [i] vs. [a]) is acoustically determined by the relative positions of the first two or three formants (F1–F3). As hearers we never hear the resonant frequencies (formants) themselves, instead we hear the harmonic components of the tone being produced at the glottis being strengthened as a result of the formant frequencies. The harmonic components of the glottis signal are simple multiples of the fundamental frequency. Therefore, the glottis signal of a typical male speaking at a fundamental frequency of 110 Hz will have harmonics at 220, 330, 440, 550, 660, 770, 880 Hz, etc. In contrast, the glottis signal of a typical female, speaking at a fundamental frequency of 200 Hz, will have harmonics at 400, 600, 800, 1000 Hz, etc. Depending on the vowel being produced, harmonic components will be strengthened when they are near a particular formant frequency. This relationship between harmonics and formant frequencies is illustrated in Figure 4. The magnetic resonance images are of a male (top) and a female (bottom) speaker articulating the vowel [i]. In (a), the harmonics of the glottal signal are represented as evenly spaced vertical lines. In regular, voicing the strength of the individual harmonics decreases as the frequency increases. The diagram in (b) represents the resonant frequencies of the vocal tract configured for [i], that is, the tongue is high up in the front of the mouth and the lips are spread. The diagram in (c) represents the combination of (a) and (b), that is, the relative strengths of the harmonics produced at the glottis have been modified by the formant (resonant) frequencies of the particular vowel.

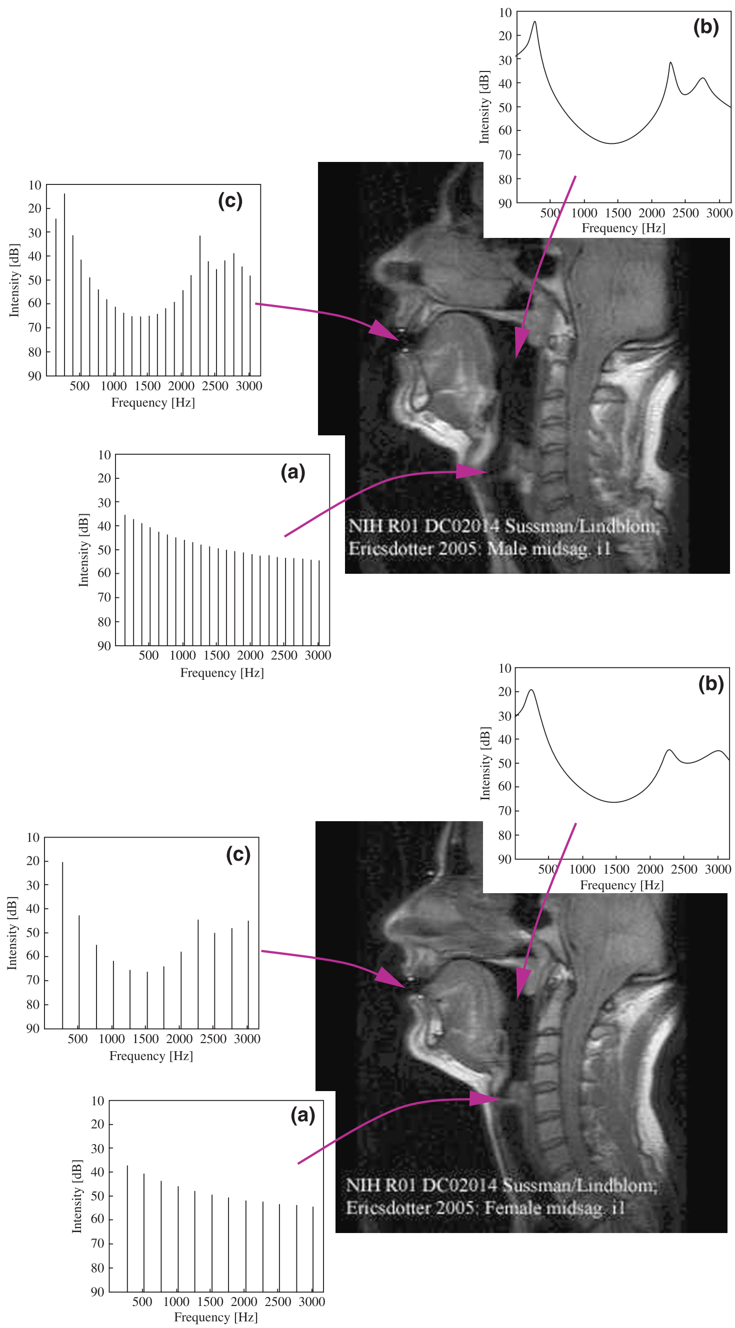

Fig. 4. Relationship between (a) source spectrum, (b) vocal tract filter, and (c) output spectrum for the vowel [i] for a male speaker (top) and a female speaker (bottom). Magnetic resonance images are taken from Ericsdotter (2005).

Two important differences must be noted between the speakers. First, the harmonic spacing in the male speaker is much tighter than it is for the female speaker, which is a direct consequence of the lower pitch of his voice. Second, and this is less apparent, the formant frequencies of the female [i] are slightly higher than those of the male, due in part at least to the shorter female vocal tract length. One explanation researchers (Goldstein 1980; Ryalls and Lieberman 1982; Diehl et al. 1996) have offered for female vowels being more widely spaced acoustically is that the wider harmonic spacing of the female voice does a poorer job of defining vowel qualities and spacing them further apart acoustically compensates for this.

Diehl et al. (1996) tested this hypothesis from a perceptual point of view. They synthesised the vowels [ɪ] and [ʊ] at different fundamental frequencies, starting with the equivalent of a low male pitch of 90 Hz and in 30 Hz steps going up to a pitch of 360 Hz, which is more in the range of a young child. The stimuli were randomised and played to listeners who were asked for each stimulus whether they heard [ɪ] or [ʊ]. The results of a series of related experiments suggest that a listener's ability to identify a particular vowel quality decreases as the fundamental frequency increases, in turn increasing the spacing of the individual harmonics. Interestingly, the hypothesis Diehl et al. (1996) are testing is that vowel space differences are not primarily gender-specific, but rather are first and foremost related to the fundamental frequency of the speaker's voice. It just happens that the average female voice has a higher fundamental frequency than the average male and so the female vowel space is larger.

The relationship between voice pitch and acoustic vowel space size is still poorly understood. The results of Diehl et al.'s experiments still leave a number of questions unanswered. They only examined the perception of a pair of vowels and they are aware that other factors may have influenced their results. A more recent study (Simpson and Ericsdotter 2007) tested the hypothesis that if vowel space size does correlate with fundamental frequency then we should find variation in acoustic vowel space size within a group of male or female speakers as a function of the average fundamental frequency of the individual speakers. The results of this study are also inconclusive. So, while a weak correlation was found in one part of the data, another part showed no such correlation and merely indicated that further better controlled data elicitation was required.

SOCIOPHONETIC

One factor that is undoubtedly involved in all of the differences we have described is sociophonetic, that is, male and female speakers have learnt to speak in ways appropriate to their gender. In her comparison of the vowel systems of a number of different languages (three varieties of English, Dutch, Swedish and French), Henton (1995) is able to show that there are differences in the size of the differences between male and female vowel systems. In other words, although expected differences were found in vowel space size between male and female speakers, the size of the differences varied from language to language. In a later study, Johnson (2006) brings together average male and female formant data of the vowels [i], [e], [a], [o], [u] from seventeen languages. He shows that there is only a very weak correlation between cross-language female–male differences and differences in average height (Tolonen et al. 2000), which in turn have been shown to correlate with differences in vocal tract length (Cherng et al. 2002). This finding is perhaps unsurprising since studies have repeatedly failed to find anything but very weak correlations between speaker height and acoustic measures within same-sex groups (e.g. Greisbach 1999; González 2004). A large body of sociolinguistic research has repeatedly shown that gender is one of the most important factors that must be considered when trying to account for phonetic variation found within a speech community (e.g. Eckert 1989; Labov 1990). Most importantly for our discussion here is that, at the time we take a snapshot of the sounds being produced by the men and women in the same speech community, we must expect part of the male–female differences we observe to be a direct result of men and women talking differently.[3]

Voice Onset Time

In many languages, including English, certain consonant distinctions are made using differences in voice onset time (henceforth, VOT). This refers to the time lag between the release of a plosive and the onset of vocal fold vibration. If vocal fold vibration begins after plosive release, as it does in voiceless plosives, VOT has a positive value. In contrast, we find a negative VOT in voiced plosives, where vocal fold vibration begins before the plosive is released. In word and utterance initial position, the English plosives /p–b, t–d, k–ɡ/ are often all unvoiced, the main difference residing in /p, t, k/ having a greater positive VOT than /b, d, ɡ/. Apart from bringing about meaningful distinctions in languages, VOT has also been shown to vary in populations as a function of age, hormonal variation and gender. Although results vary from study to study, one of the most consistent findings is that if there are significant male–female differences, it is the women who in general are found to have longer VOTs. Swartz (1992) found that men had shorter VOTs than women in the plosive pair /t, d/. Significantly higher VOTs for voiceless female plosives /p, t, k/ were also found in a number of further studies (Ryalls et al. 1997; Koenig 2000; Robb et al. 2005).

Some studies have also examined VOT differences in boys and girls. Koenig (2000) compared VOT in the American English of a group of 5-year-olds with that of a group of adults. She found no significant differences in average VOT for the two groups, but children were found to be significantly more variable in their VOT suggesting that they have yet to acquire the necessary laryngeal control. In contrast, in two studies involving children speaking British English, VOT differences were found across different age groups. Whiteside and Marshall (2001) examined differences in VOT for the English plosive pairs /p-b, t-d, k-g/ in a group of 15 girls and 15 boys split evenly across three age groups (7, 9 and 11). The results of this investigation are quite complex, but the clearest and most interesting finding was that girls at the age of eleven made the largest VOT contrast between the plosives in a pair, that is, between /p/ and /b/ or between /t/ and /d/. In a subsequent larger study (Whiteside et al. 2004b), which included two further groups expanding the age range to younger (5) and older (13) children, the results of the earlier study were largely replicated. In particular, the VOTs of the girls from the oldest group displayed significantly higher VOTs than the boys.

VOT has also been used to investigate the effects of changes in hormone (oestrogen, progesterone) levels during the menstrual cycle. Although the effects of hormone levels during the female cycle have been studied extensively in other domains (e.g. Hampson 1990), they have as yet received little attention in speech. Two related studies (Whiteside et al. 2004a; Wadnerkar et al. 2006) analysed female VOT at two different times during the menstrual cycle, once at the beginning when levels of the female hormones oestrogen and progesterone are at their lowest level and then 18–25 days into the cycle when oestrogen and progesterone levels are at their highest. Two interesting findings come out of these studies. As expected from the findings of other studies, females had longer VOTs than males. More interestingly, at the high hormone levels, females exhibited significantly longer VOTs for the fortis plosives /p, t, k/ and shorter VOTs for the lenis plosives /b, d, g/ than they did at lower levels. In other words, females significantly increased the size of the VOT difference between fortis and lenis plosives when levels of hormone are at their highest, thus enhancing speech clarity.

How are gender-related differences in VOT to be explained? Beginning with the hormone-related variability, Wadnerkar et al. (2006) are not sure whether changes in peripheral tissue systems or in the central nervous system might be responsible. Differences in vocal fold morphology and behaviour have been suggested to account for gender-related differences. Robb et al. (2005) suggest that the posterior opening of the vocal folds that has been observed during female voicing (Bless and Abbs 1983) might be responsible for increased air pressure during the closure phases of plosives causing voicing to begin later (greater VOT) in following vowels. Both Koenig (2000) and Whiteside et al. (2004b) suggest that it might be relative differences in the stiffness of male and female vocal folds that give rise to an earlier onset of voicing in laxer male vocal folds than in stiffer female vocal folds. However, neither of these explanations is sufficient to account for the significant changes found during the menstrual cycle, since females exhibit both shorter as well as longer VOT than male speakers.

Duration and Reduction

In an acoustic investigation of German vowels from a sample of 29 male and 25 female speakers, Simpson (1998) found consistent and systematic differences in the durations of male and female vowels. Although the data were drawn from both spontaneous and read speech, the duration of male vowels was shorter than female vowels in every category, with an average of about 11%.

A survey of other studies shows that this trend is restricted neither to German, nor to Indo-European languages. In a large-scale acoustic investigation of American English vowels (Hillenbrand et al. 1995), female speakers exhibited longer vowel duration consistent across all categories. In Quebecois French (Martin 1995; 1998a,b; 2001), female vowel durations were found to be longer for certain front vowel categories, although no significant differences were found between back vowel categories [o, ɔ, ɔ̃] and male [ø] was found to be longer. In an investigation of acoustic vowel systems in 20 speakers from Jamaican Creole-dominant and Jamaican English-dominant communities, Wassink (1999) found that female speakers from both groups exhibited greater durational differences between long and short vowel pairs. Female speakers of the non-Indo-European language Creek (Johnson and Martin 2001) were also found to produce greater durational differences between long and short vowel categories. In a study aimed specifically at investigating sex-specific duration patterns in Swedish, Ericsdotter and Ericsson (2001) found that female speakers in general produced greater vowel durations than males, but the most consistent pattern was that female speakers produced greater durational differences between stressed and unstressed tokens of vowels in the same monosyllabic words.

In a few studies, utterance duration has also been explored. Here the findings are less conclusive. In studies of American English (Byrd 1992), Northern British English (Whiteside 1996) as well as German (Simpson 1998) females were found to have longer utterance durations than males. In another study of American English and Swedish data (Simpson and Ericsdotter 2003), however, no significant differences were found between female and male utterance durations.

Besides duration, both Byrd (1992) and Whiteside (1996) examine patterns of reduction. Byrd (1992) found not only that male speakers had a higher speaking rate than females, but that significantly more vowels were reduced to the central vowel [ə]. Whiteside (1996) found similar patterns in her Northern British English data. So, for example, male speakers used a reduced [ə] form of the copula 'are' more often than the female speakers did and were also more likely to elide it completely.

What are the possible reasons for these differences in duration and reduction (if we may assume them to be related)? As with vowel space size differences, a sociophonetic reason is the most likely. The larger female vowel space, the longer duration of stressed vowels, the greater durational distinction between stressed and unstressed vowels and less reduction of vowel qualities to [ə] can all be treated as phonetic correlates of speaking clearly. Alternatively, opposite patterns in male speakers can be seen as a tendency to speak less clearly. But what are possible reasons for female to speak more or males to speak less clearly? There is no obvious reason why male speakers should speak less clearly, but it has been suggested that female clarity might be related to the role of primary care-givers that they more often fulfil. In this role one of the main functions in the initial years of a child's life at least is to be one of the main sources of language input, which in turn could be an important reason for speaking more clearly (Labov 1990).

Differences between the Speech of Girls and Boys

Differences between the speech of boys and girls are particularly intriguing. Up until now, we have concentrated on differences between adult males and females, mentioning child language only in relation to VOT. However, a number of the differences that have been described can already be found in the speech of girls and boys. This is interesting because prior to the onset of puberty the vocal folds and vocal tracts of boys and girls exhibit only minor differences. In other words, from the point of view of their speech organs, preadolescent girls and boys can be considered to be identical.

Studies that have looked at different acoustic parameters (fundamental frequency, formant values) in groups of preadolescent children have found certain differences. Busby and Plant (1995) investigated the speech of 20 girls and 20 boys split across four age groups (5, 7, 9, and 11) reading a selection of words embedded in a short sentence. Although, they found an expected decrease of voice pitch as the children's age increased, no difference was found between the genders. For the formants, the most important finding was that girls had higher F1 and F2 values for the majority of the vowel categories analysed. In a large study involving 436 children aged between 5 and 17 years, Lee et al. (1999) found similar patterns with respect to fundamental frequency and formants. Significant differences between the fundamental frequency of boys and girls were found from 12 years on, that is, with the onset of puberty. With regard to formants, the data also reveal higher female values for F1 and F2 across the majority of vowel qualities. Interestingly, though, Lee et al. (1999) found no significant gender-specific durational differences (see also Whiteside 2001).

Now, although analysts have found significant acoustic differences between boys and girls, the differences have not been found in the pitch of the voice, the parameter that we seem to rely on quite heavily when distinguishing adult males from females. So, are listeners able to identify a boy or a girl from a voice sample alone? Günzburger et al. (1987) had 11 boys and six girls aged between 7 and 8 produce isolated vowel sounds and read short sentences. When attending to the isolated vowel stimuli, listeners were only able to identify boys at a level slightly better than chance, but were not able do this with the stimuli produced by the girls. Only when they were listening to the sentence material did listeners perform considerably better, correctly identifying a child's sex in more than 70% of cases. Interestingly, an examination of the typical acoustic parameters (fundamental frequency, formant frequencies) found no significant differences in the data set between boys and girls. Following the failure to find any significant differences in basic acoustic parameters, a group of visually handicapped listeners were asked to make a series of voice quality judgements on the voices of the three boys and three girls who received the best identification ratings. Listeners were asked to judge each stimulus with one value from a series of opposites ('clear-dull', 'soft-loud' and 'precise-careless'). In contrast to the acoustic parameters analysed, voice quality judgements made a number of distinctions between the girls and boys. So, while the girls were judged to be clear, soft, shrill, high pitch, melodious and precise, boys were significantly found to be dull, loud, deep, low pitch, monotonous and careless. Unfortunately, interesting though these findings are, it is unclear to what extent these voice quality judgements are affected by listeners first identifying the child's gender, and subsequently attributing the child with stereotypical qualities that might not be present in the stimuli they are listening to.

Conclusion

As we have seen, if we randomly pick out a group of male and female speakers of a language, we can expect to find several differences in their speech. Although it is relatively easy to describe these differences, it is a much harder task to explain exactly why the differences are the way they are. We know that we must attribute some differences to biophysical consequences of differences in anatomy and physiology and others to differences in learned behaviours. What is less clear is where the dividing line between the two is always to be drawn.

Short Biography

Adrian Simpson is a phonetician. He began his academic career teaching English phonetics at the University of Tübingen (Germany) before spending 9 years as a lecturer in phonetics at the Institute of Phonetics and Digital Speech Processing, Kiel (Germany). Since 2001 he has been professor of speech at the Institute of German Linguistics, Jena (Germany).

His main areas of research are the phonetics and phonology of spontaneous speech, the phonetics and phonology of German and the phonetics of gender. He has also authored and co-authored papers on the phonetics of gemination in Malayalam, as well as in the fields of clinical phonetics and speech synthesis. He holds a BA in Language and a PhD both from the University of York (UK). He is co-editor of the Journal of the International Phonetic Association.

Notes

* Correspondence address: Adrian P. Simpson, Institut für Germanistische Sprachwissenschaft, Friedrich-Schiller-Universität Jena, 07737 Jena, Germany. E-mail: adrian.simpson@uni-jena.de.

[1] Creaky voice is brought about by closing the vocal folds tightly at the back and making them short and lax. Vibration in this configuration is low frequency, generally below 80 Hz.
[2] Articulatory movements are recorded using x-ray microbeams that track the positions of pellets placed on the mid-line surface of the tongue, lips, jaw and nose. The reader should consult Westbury (1994) for further details of the recording procedure.
[3] Jannedy and Hay (2006) contains a number of recent papers exploring different aspects of describing, explaining and modelling sociophonetic variation.

Works Cited

Bless, D. M., and J. Abbs. 1983. Vocal fold physiology: Contemporary research and clinical issues. San Diego, CA: College-Hill Press.

Boë, L-J., M. Contini, and H. Rakotofiringa. 1975. Étude statistique de la fréquence laryngienne. Phonetica 32.1–23.

Busby, P. A., and G. L. Plant. 1995. Formant frequency values of vowels produced by preadolescent boys and girls. Journal of the Acoustical Society of America 97.2603–2607.

Byrd, D. 1992. Preliminary results on speaker-dependent variation in the TIMIT database. Journal of the Acoustical Society of America 92.593–596.

Cherng, C.-H., C.-S. Wong, C.-H. Hsu, and S.-T. Ho. 2002. Air length in adults: estimation of the optimal endotracheal tube length for orotracheal intubation. Journal of Clinical Anesthesia 14.271–274.

Chiba, T., and M. Kajiyama. 1941. The Vowel – Its Nature and Structure. Tokyo, Japan: Tokyo-Kaiseikan.

Diehl, R. L., B. Lindblom, K. A. Hoemeke, and R. P. Fahey. 1996. On explaining certain male-female differences in the phonetic realization of vowel categories. Journal of Phonetics 24.187–208.

Eckert, P. 1989. The whole woman: Sex and gender differences in variation. Language Variation and Change 1.245–267.

Ericsdotter, C. 2005. Articulatory-acoustic relationships in Swedish vowel sounds. PhD thesis. Stockholm University.

Ericsdotter, C., and A. M. Ericsson. 2001. Gender differences in vowel duration in read Swedish: Preliminary results. In Proc. Fonetik 2001, XIVth Swedish Phonetics Conference. Working Papers of the Department of Linguistics, Lund University 49.34–37.

Fant, G. 1966. A note on vocal tract size factors and non-uniform F-pattern scalings. STL-QPSR 4.22–30.

———. 1975. Non-uniform vowel normalization. STL-QPSR 2–3.1–19.

Goldstein, U. 1980. An articulatory model for the vocal tracts of growing children. PhD. thesis, MIT, MA.

González, J. 2004. Formant frequencies and body size of speaker: a weak relationship in adult humans. Journal of Phonetics 32.277–287.

Greisbach, R. 1999. 'Estimation of speaker height from formant frequencies.' Forensic Linguistics 6.265–277.

Günzburger, D., A. Bresser, and M. ter Keurs. 1987. Voice identification of prepubertal boys and girls by normally sighted and visually handicapped subjects. Language and Speech 30.47–58.

Haan, J. 2002. Speaking of questions. An exploration of Dutch question intonation. Utrecht, The Netherlands: LOT.

Haan, J., and V. J. van Heuven 1999. Male vs. female pitch range in Dutch questions. In Proc. XIVth ICPhS, San Francisco, 1581–1584.

Hampson, E. 1990. Variations in sex-related cognitive abilities across the menstrual cycle. Brain and Cognition 14.26–43.

Hardcastle, W., and N. Hewlett (eds.) 2006. Coarticulation: Theory, Data, and Techniques. Cambridge, UK: Cambridge University Press.

Henton, C. G. 1989. Fact and fiction in the description of female and male pitch. Language and Communication 9.299–311.

———. 1992. The abnormality of male speech. New Departures in Linguistics, ed. by G. Wolf, 27–59. New York, NY: Garland.

———. 1995. Cross-language variation in the vowels of female and male speakers. In Proc. XIIIth ICPhS volume 4, Stockholm, 420–423.

Henton, C. G., and R. A. W. Bladon. 1985. Breathiness in normal female speech: Inefficiency versus desirability. Language and Communication 5.221–227.

———. 1988. Creak as a sociophonetic marker. Language, Speech and Mind: Studies in Honour of Victoria A. Fromkin, ed. by L. M. Hyman and C. N. Li, 3–29. London: Routledge.

Herbst, L. 1964. Untersuchungen zur Indifferenzlage der Sprechstimme. Studien zur Problematik des physiologischen Hauptsprechtonbereichs. PhD thesis, Halle, Germany: Universität Halle-Wittenberg.

———. 1969. Die Umfänge der physiologischen Hauptsprechtonbereiche von Frauen und Männern. Zeitschrift für Phonetik und Sprachliche Kommunikation 22.426–438.

Hillenbrand, J., L. A. Getty, M. J. Clark, and K. Wheeler. 1995. Acoustic characteristics of American English vowels. Journal of the Acoustical Society of America 97.3099–3111.

Jannedy, S., and J. Hay (eds.) 2006. Modelling sociophonetic variation. Journal of Phonetics 34(4). doi:10.1016/j.wocn.2006.08.001

Johnson, K. 2006. Resonance in an exemplar-based lexicon: the emergence of social identity and phonology. Journal of Phonetics 34.485–499.

Johnson, K., and J. Martin. 2001. Acoustic vowel reduction in Creek: effects of distinctive length and position in the word. Phonetica 58.81–102.

Klatt, D. H., and L. C. Klatt. 1990. Analysis, synthesis, and perception of voice quality variations among female and male talkers. Journal of the Acoustical Society of America 87.820–857.

Koenig, L. 2000. Laryngeal factors in voiceless consonant production in men, women, and 5-year-olds. Journal of Speech, Language and Hearing Research 43.1211–1228.

Labov, W. 1990. The intersection of sex and social class in the course of linguistic change. Language Variation and Change 2.205–254.

Ladefoged, P. 1996. Elements of Acoustic Phonetics. Chicago, IL: The University of Chicago Press.

Lee, S., A. Potamianos, and S. Narayanan, S. 1999. Acoustics of children's speech: Developmental changes of temporal and spectral parameters. Journal of the Acoustical Society of America 105.1455–1468.

Manuel, S. Y. 1990. The role of contrast in limiting vowel-to-vowel coarticulation in different languages. Journal of the Acoustical Society of America 88.1286–1298.

———. 2006. Cross-language studies: relating language-particular coarticulation patterns to other language-particular facts. Coarticulation: Theory, Data and Techniques, ed. by W. J. Hardcastle and N. Hewlett, 179–198. Cambridge, MA: Cambridge University Press.

Martin, P. 1995. The opposition between /ɛ/ and /ɛː/ in current Quebec French. La Linguistique 31.33–45.

———. 1998a. Dynamique vocalique en français du Québec. La Linguistique 34.67–76.

———. 1998b. Les voyelles d'aperture moyenne en français du Québec. Cahiers de l'ILSL 11.215–242.

———. 2001. Les voyelles nasales en français du Québec. La Linguistique 37.49–70.
Nordström, P.-E. 1977. Female and infant vocal tracts simulated from male area functions. Journal of Phonetics 5.81–92.
Robb, M., H. Gilbert, and J. Lerman. 2005. Influence of gender and environmental setting on voice onset time. Folia Phoniatrica et Logopaedica 57.125–133.
Rose, P. 1991. How effective are long term mean and standard deviation as normalisation parameters for tonal fundamental frequency? Speech Communication 10.229–247.
Ryalls, J. H., and P. Lieberman. 1982. Fundamental frequency and vowel perception. Journal of the Acoustical Society of America 72.1631–1634.
Ryalls, J. H., A. Zipprer, and P. Baldauff. 1997. A preliminary investigation of the effects of gender and race on voice onset time. Journal of Speech, Language and Hearing Research 40.642–645.
Simpson, A. P. 1998. Phonetische Datenbanken des Deutschen in der empirischen Sprachforschung und der phonologischen Theoriebildung. Arbeitsberichte des Instituts für Phonetik und digitale Sprachverarbeitung der Universität Kiel (AIPUK) 33.
———. 2001. Dynamic consequences of differences in male and female vocal tract dimensions. Journal of the Acoustical Society of America 109.2153–2164.
———. 2002. Gender-specific articulatory-acoustic relations in vowel sequences. Journal of Phonetics 30.417–435.
Simpson, A. P., and C. Ericsdotter. 2003. Sex-specific durational differences in English and Swedish. In Proc. XVth ICPhS, Barcelona, 1113–1116.
———. 2007. Sex-specific differences in f0 and vowel space. In: Proc. XVIth ICPhS. Saarbrücken.
Swartz, B. L. 1992. Gender difference in voice onset time. Perceptual and Motor Skills 75.983–992.
Takefuta, Y., E. G. Jancosek, and M. Brunt. 1972. A statistical analysis of melody curves in the intonation of American English. In Proc. VIIth ICPhS, Montreal 1971, 1035–1039.
Titze, I. R. 1989. Physiologic and acoustic differences between male and female voices. Journal of the Acoustical Society of America 85.1699–1707.
Traunmüller, H. 1984. Articulatory and perceptual factors controlling the age- and sex-conditioned variability in formant frequencies of vowels. Speech Communication 3.49–61.
Traunmüller, H., and A. Eriksson. 1995. The frequency range of the voice fundamental in the speech of male and female adults. Unpublished manuscript. Stockholm <http://www.ling.su.se/staff/hartmut/f0_m&f.pdf>
Tolonen, H., K. Kuulasmaa, and E. Ruokokoski. 2000. MONICA Population Survey Data Book. World Health Organization <http://www.ktl.fi/publications/monica/surveydb/title.htm>
Wadnerkar, M. B., P. E. Cowell, and S. P. Whiteside. 2006. Speech across the menstrual cycle: A replication and extension study. Neuroscience Letters 408.21–24.
Wassink, A. B. 1999. A sociophonetic analysis of Jamaican vowels. PhD thesis, Michigan: University of Michigan.
Westbury, J. R. 1994. X-ray Microbeam Speech Production Database User's Handbook, Version 1.0. Madison WI. <http://www.medsch.wisc.edu/~milenkvc/pdf/ubdbman.pdf>
Whiteside, S. P. 1996. Temporal-based acoustic-phonetic patterns in read speech: Some evidence for speaker sex differences. Journal of the International Phonetic Association 26.23–40.
———. 2001. Sex-specific fundamental and formant frequency patterns in a cross-sectional study. Journal of the Acoustical Society of America 110.464–478.
Whiteside, S. P., and J. Marshall. 2001. Developmental trends in Voice Onset Time: some evidence for sex differences. Phonetica 58.196–210.
Whiteside, S. P., A. Hanson, and P. E. Cowell. 2004a. Hormones and temporal components of speech: sex differences and effects of menstrual cyclicity on speech. Neuroscience Letters 367.44–47.
Whiteside, S. P., L. Henry, and R. Dobbin. 2004b. Sex differences in voice onset time: A developmental study of phonetic context effects in British English. Journal of the Acoustical Society of America 116.1179–1183.