SEX-SPECIFIC DIFFERENCES IN f_0 AND VOWEL SPACE

Adrian P. Simpson & Christine Ericsdotter

Friedrich-Schiller-Universität Jena, Germany
adrian.simpson@uni-jena.de ericsdotter@gmx.de

ABSTRACT

It has been suggested that the larger area of the average female acoustic vowel space is a consequence of compensating for poorer harmonic sampling of the spectral envelope resulting from a higher f_0. This predicts that there should be variation in vowel space size within any group of males or females representing sufficient interindividual range of average f_0. Inspired by this, the present paper examines whether there is a correlation between a speaker's f_0 and the size of the speaker's F2×F1 vowel space.

A highly significant correlation between f_0 and vowel space size is found in the *female* group of a sample of 27 German students. However, no such correlation is found between f_0 and the Euclidean distance between same speaker tokens of /eː/ and /aː/.

Keywords: sufficient contrast, sex-specific vowel differences

1. INTRODUCTION

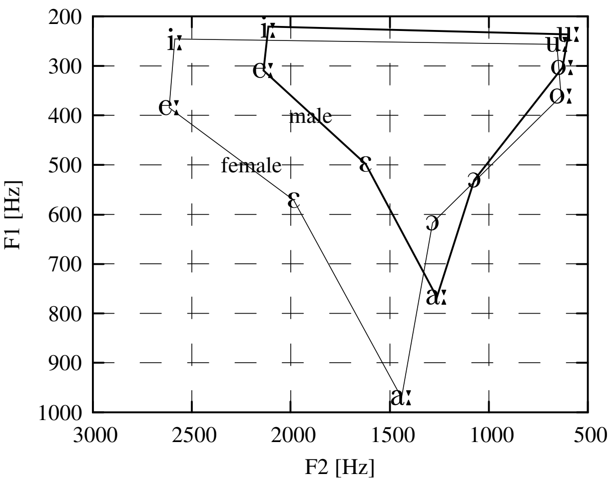
It is well established that acoustic differences between male and female vowel systems are the result of both biophysical and sociophonetic factors. However, there still remains a good deal of uncertainty and controversy about which differences are to be attributed to which set of factors, and there are areas of disagreement within the different biophysical reasons which have been offered.

Mean formant data plotted in Figure 1 from 27 female and 14 male German speakers in the present study illustrate two unresolved questions regarding vowel system differences which have been repeatedly observed in a number of languages:

(a) The female vowel system has a larger acoustic area.

(b) There are non-uniform acoustic differences between male and female vowels. In particular the acoustic differences between male and female tokens of the same vowel category increases with formant magnitude: the F1 difference increases with vowel openness while the F2 difference increases with vowel frontness. By contrast, differences between the male and female back vowels /oː/ and /uː/ are minimal.

Figure 1: Mean vowel spaces from 27 female and 14 male German speakers

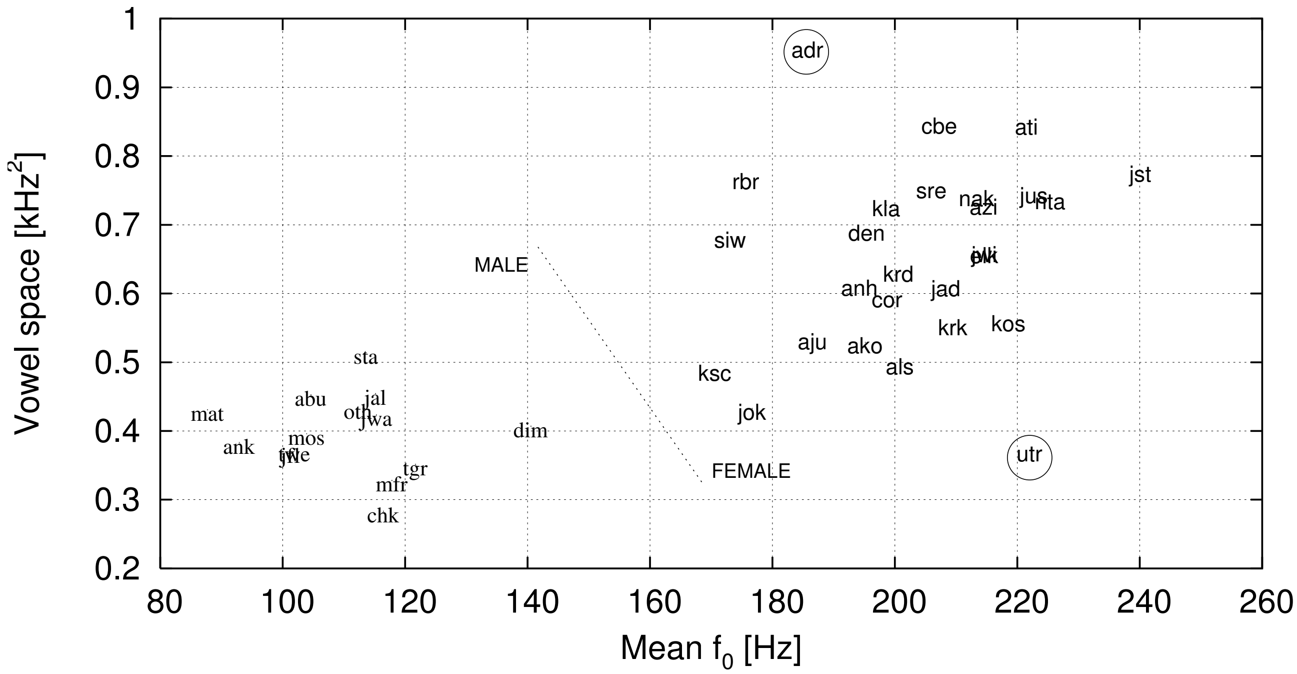


Average differences in male and female vocal tract geometry have gone some way to accounting for non-uniformity but have failed to resolve the magnitude of the similar differences in vowel space size which have been found [3, 4, 5, 10, 16]. So, for instance, Fant [5] has suggested that female speakers lower the formant values of the back vowels [oː] and [uː] thus bringing them closer to male values by using tighter and longer dorso-velar and labial strictures, exploiting the double Helmholtz resonator-like properties of these particular vowel categories.

The sociophonetic aspect in vowel space magnitude differences becomes clear from cross-linguistic comparison. Although the larger female vowel space has been repeatedly found in data from different languages, the size of the difference is by no means constant [8].

Another reason which has been proposed for these differences in vowel space size (and non-uniformity) is that female speakers space vowel qualities further apart acoustically in order to compensate for the poorer harmonic sampling of the spectral envelope caused by a higher fundamental frequency [6, 11, 2]. More specifically, Diehl et al. [2] suggest that the differences we can observe in Figure 1 might not be gender-based at all, but simply a consequence of the f_0 differences between two speaker groups. As part

Figure 2: Vowel space area plotted as a function of average f_0 for the 27 female (right) and 14 male speakers (left). Female speakers with the largest and smallest vowel space areas have been circled (see text).



of their evidence they also cite differences in vowel space size which has been observed in bass and tenor singers [1].

Although the finding that a sufficiently high f_0 can impair perception of vowel quality in synthetic stimuli, there is as yet no evidence that speech in its normal communicative environment suffers from the male-female-child range of f_0 with which it is produced. Indeed Diehl et al.'s claim that the 3% pitch change used in their stimuli is not to be found in the majority of vowels in connected speech would seem to be confounded by the fact that normal speech is characterised by a constantly changing spectral envelope, even if there is little f_0 movement.

In this paper we test possible consequence of what Diehl et al. have called the *sufficient contrast* hypothesis. If the larger size of the female acoustic vowel space is indeed merely a consequence of a compensation being made for a more poorly sampled spectral envelope, then we should expect differences in vowel space dimensions to be present in a sufficiently large sample of male or female speakers who exhibit a sufficiently large range of average f_0 differences. In other words, a female speaker with a low f_0 should exhibit a smaller vowel space than a female with a high f_0.

2. METHOD

Recordings of 17 male and 70 female speakers had been made over a course of four years as part of an introductory course in acoustic phonetics. In order to be able to analyse a small sample of their own speech, students were required to record a number of word lists as well as ten short sentences which together covered many aspects of the German vowel and consonant systems. Two parts of these recordings were used in this study:

1. A word list involving speakers producing long vowels in isolated words with the following structure: *hVːb(ə)*, e.g. *Hub*, *habe*.
2. The ten short sentences taken from the Berlin sentence set [14, 15], e.g. *Über die Felder weht ein Wind*.

All recordings were made in a sound-treated room direct to PC using CSL, digitizing at a sampling rate of 16 kHz and an amplitude resolution of 16 bits.

Formant frequency estimation was carried out on tokens of the long vowels /iː, eː, aː, oː, uː/ in the words *Hieb*, *heb*, *hab hob*, *Hub*, *Hiebe*, *hebe*, *habe*, *Hobel*, *Hube* as well as in the two short vowels /ɛ, ɔ/ from the words *Frühlingswetter* and *Sonne* contained in the sentence material. It is recognized that mixing the long vowels from words spoken in isolation with tokens of two short vowels spoken in a sentential context is far from ideal, but we wanted to have two vowel space defining points between the close-mid /eː, oː/ and the open /aː/. The relatively central position of /ɛ/ and /ɔ/ shown in Figure 1 is undoubtedly in part due to these contextual differences.

Following manual labelling of the relevant vocalic portions for each speaker, formant estimation was carried out automatically at the centre of each vocalic portion using two root-solving algorithms [12, 13] with default settings. Formant measurement was considered to be reliable when the sum of the two estimates was less than 10 % the squared difference between the estimates.

The average f_0 for a speaker was obtained by calculating the mean f_0 from the ten sentences.

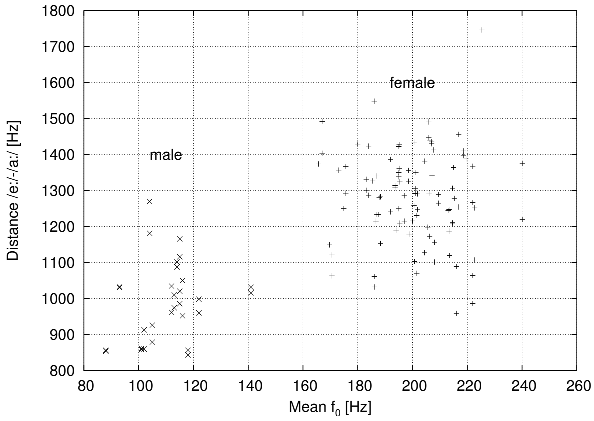
3. RESULTS

Interindividual average f_0 ranged from a minimum of 163 Hz to a maximum of 240 Hz in the female group ($n = 70$), and a minimum of 88 Hz to a maximum of 141 Hz in the male group ($n = 17$).

Out of the 17 male and 70 female speakers, reliable measurements for the complete set of vowels defining the vowel space ([iː, eː, ɛ, aː, ɔ, oː, uː]) were obtained for 14 male and 27 female subjects. Mean male and female vowel spaces calculated from these speakers are displayed in Figure 1. As mentioned in the introduction, Figure 1 confirms that this speaker group exhibits the same differences between male and female vowel system that has been described in other languages, e.g. Korean [2].

Figure 2 plots the area of the individual vowel spaces as a function of the average f_0 for each speaker. As we would expect from Figure 1 the female areas are in general larger than the male areas, although there is a certain amount of overlap with the smallest female and largest male areas. Correlations between average f_0 and vowel space area exhibit interesting patterns. First, as can be seen, there is no significant correlation between mean f_0 and vowel space area in the male group ($r = -0.09$). In particular, the male speaker with the highest average f_0 (dim) has a vowel space of 0.4 kHz^2 which is almost identical with the mean male area (0.41 kHz^2). In the female group, the situation is more complex. Taken as a group there is no significant correlation between female f_0 and area ($n = 27, df = 25, r = 0.2, p > 0.05$). However, if the two speakers with the largest (adr) and the smallest (utr) areas (circled in Figure 2) are removed, we do find a highly significant positive correlation ($n = 25, df = 23, r = 0.49, p < 0.01$), indicating in contrast to the male group a relationship between f_0 and area.

Figure 3: Euclidean distance between same speaker tokens of the vowels /eː/ and /aː/ plotted as a function of a speaker's average f_0. Male values to the left, female to the right.



Since the complete vowel space could only reliably be determined for less than half of the female sample using the method described, the correlation between f_0 and another parameter directly related to vowel space size but requiring fewer formant estimates was sought. Since both F2 of front non-open vowel qualities and F1 of open vowels are chiefly responsible for the dimensional differences between male and female vowel spaces, the Euclidean distance between same speaker tokens of the vowels /eː/ and /aː/ in the F1×F2 space was chosen. Speakers produced two tokens of each of these vowels, once in a monosyllabic (*hab*) and once in a disyllabic (*habe*) environment. In all, 126 (96 female, 30 male) reliable formant estimates pairing off these vowels at least once per speaker were found. Figure 3 plots the Euclidean distance between same speaker tokens of /eː/ and /aː/ in the F1×F2 space as a function of a speaker's average f_0. In contrast to the significant correlation found between female vowel space area and mean f_0, correlations for both the male and the female are insignificant (male: $r = 0.22$; female: $r = -0.05$).

4. DISCUSSION

The main aim of this study was to find out whether the size of individual vowel spaces within a group of male and female speakers is positively correlated with interspeaker differences in average f_0. If speakers compensate for a poorer density of harmonic sampling of the spectral envelope by spacing their vowel categories further apart acoustically (and auditorily), then a larger vowel space should be one consequence.

Despite having a relatively large sample of speakers, and in particular, of female subjects ($n = 70$) with average f_0 ranging from ca. 160 to 240 Hz, the results of this study have been inconclusive. For a subset of the female speakers a highly significant correlation was found between average f_0 and F1×F2 vowel space size, albeit after the speakers with the largest and the smallest acoustic vowel spaces had been excluded. Taken on its own this finding would seem to substantiate the prediction made in the Introduction. However, questions and uncertainties remain. If the male data in Figure 2 are as reliable as the female data, why is no such correlation present in the male group? One possible reason is that compensation only becomes necessary once harmonic spacing reaches a particular level. Below that, i.e. in a normal sample of male speakers, no compensation is required. Uncertainty remains because no significant correlation was found between f_0 and the intercategorial distance between same speaker tokens of /eː/ and /aː/, which could reasonably have been expected, given the highly significant correlation found in the vowel space data.

It is worth remembering here that speaker size differences within the two sex groups can safely be excluded from the outset, i.e. we do not expect a larger speaker to have lower f_0 and formant frequencies and therefore a smaller vowel space. Previous studies, carried out mainly within a forensic framework have found only weak or marginal correlations between size and acoustic parameters [9, 7].

Given the questions and uncertainties that remain, it is perhaps more reasonable to see the present study as exploratory, making use of a fortuitous collection of data from a relatively large speaker sample. Acquisition of speakers and of speech material tailored to addressing the questions raised by this study are underway. Despite having a range of interindividual mean f_0 for both sexes, both female and male speakers with high average f_0 are poorly represented in the present sample. Further male and female speakers are being acquired with average f_0 which lie in the upper distributional extremes of each group. It is hoped that a sufficient sample of male speakers with an average f_0 between 140 and 180 Hz will be able to find out when space size compensation begin. Likewise, although the highest female mean f_0 is 240 Hz, only six females represent the 220–240 Hz group at present. The new speech material which is being collected attempts to alleviate the formant estimation problem which is always present with high f_0. Regardless of their sex or mean f_0, many German speakers regularly produce low frequency, yet relatively strong, creak at vowel onset, i.e. signal portions with close harmonic spacing and good sampling of the spectral envelope. The use of abbreviations containing vowel sequences (e.g. *I.A.A.*) placed in a sentential context, e.g. *Sie fahren am Wochenende zur I.A.A. nach Frankfurt* reliably elicits such creaky stretches.

5. REFERENCES

[1] Cleveland, T. F. 1977. Acoustic properties of voice timbre types and their influence on voice classification. *Journal of the Acoustical Society of America* 1622–1629.

[2] Diehl, R. L., Lindblom, B., Hoemeke, K. A., Fahey, R. P. 1996. On explaining certain male-female differences in the phonetic realization of vowel categories. *Journal of Phonetics* 24, 187–208.

[3] Fant, G. 1960. *Acoustic Theory of Speech Production*. The Hague: Mouton.

[4] Fant, G. 1966. A note on vocal tract size factors and non-uniform F-pattern scalings. *STL-QPSR* 4, 22–30.

[5] Fant, G. 1975. Non-uniform vowel normalization. *STL-QPSR* 2–3, 1–19.

[6] Goldstein, U. 1980. *An articulatory model for the vocal tracts of growing children*. PhD thesis M.I.T. Massachusetts.

[7] Greisbach, R. 1999. Estimation of speaker height from formant frequencies. *Forensic Linguistics* 6, 265–277.

[8] Henton, C. G. 1995. Cross-language variation in the vowels of female and male speakers. *Proc. XIIIth ICPhS* volume 4 Stockholm. 420–423.

[9] Künzel, H. J. 1989. How well does average fundamental frequency correlate with speaker height and weight? *Phonetica* 46, 117–125.

[10] Nordström, P.-E. 1977. Female and infant vocal tracts simulated from male area functions. *Journal of Phonetics* 5, 81–92.

[11] Ryalls, J. H., Lieberman, P. 1982. Fundamental frequency and vowel perception. *Journal of the Acoustical Society of America* 72, 1631–1634.

[12] Scheffers, M. T. M., Simpson, A. P. 1995. LACS: Label assisted copy synthesis. *Proc. XIIIth ICPhS* volume 2 Stockholm. 346–349.

[13] Sjölander, K., Beskow, J. 2000. WaveSurfer – an open source speech tool. *Proc. International Conference on Spoken Language Processing* Beijing. 464–467.

[14] Sotschek, J. 1976. Methoden zur Messung der Sprachgüte I: Verfahren zur Bestimmung der Satz- und Wortverständlichkeit. *Der Fernmeldeingenieur* 30(10), 1–31.

[15] Sotschek, J. 1976. Methoden zur Messung der Sprachgüte II: Verfahren zur Bestimmung der Satz- und Wortverständlichkeit. *Der Fernmeldeingenieur* 30(12), 1–33.

[16] Traunmüller, H. 1984. Articulatory and perceptual factors controlling the age- and sex-conditioned variability in formant frequencies of vowels. *Speech Communication* 3, 49–61.