IS THERE A RELATIONSHIP BETWEEN ACOUSTIC VOWEL SPACE SIZE AND FUNDAMENTAL FREQUENCY?

Adrian P. Simpson

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

ABSTRACT

Differences in vowel formants in bass and tenor singers, as well as the perception of vowels synthesised at different fundamental frequencies predict that in a group of male or female speakers exhibiting a sufficiently wide range of average fundamental frequencies, there will be a positive correlation between the speakers' average f_0 and the dimensions of their acoustic vowel spaces. This hypothesis is tested using a group of 61 female German speakers producing the corner vowels in same-vowel sequences designed to elicit maximally peripheral vowels with intervocalic junctural glottalisation, in turn facilitating formant estimation.

Significant weak positive correlations are found between different measures of average speaker f_0 and acoustic vowel space dimensions suggesting that there is a relationship between speaker f_0 and acoustic vowel space size. However, the linear component of the relationship accounts for only 10 % of the variance suggesting a number of other factors are at work.

Keywords: sex-specific differences, articulatory-acoustic relations, sufficient contrast hypothesis

1. INTRODUCTION

Several reasons have been proposed to explain why the average female acoustic space is larger than the average male space and why the differences between individual vowel categories are not uniform. Accounts have been both biophysical [5, 6] as well as behavioural [1, 8, 9]. However, a perceptual reason has also been offered for the greater acoustic spacing between individual vowel categories [4, 7, 15]. It has been hypothesised that the size of a speaker's acoustic vowel space is a consequence of the speaker's fundamental frequency. The reasoning is as follows. The greater interharmonic spacing of a high f_0 leads to poorer definition of the spectral envelope of a vowel, which, in turn, can be compensated for by increasing the acoustic distance between individual vowel categories. As argued in [4], the different sizes of the male and female acoustic vowel spaces should not be seen directly as a sex-specific difference, but merely as a consequence of differences in average female and male f_0. Perceptual consequences of this hypothesis were tested with a series of listening experiments using vowel stimuli synthesised at different f_0 values [4]. In partial confirmation of this hypothesis, it was shown that it becomes increasingly difficult to distinguish between synthetic tokens of [ɪ] and [ʊ] as f_0 is increased. This would also seem to be supported by the finding that tenor singers have higher vowel formant frequencies than bass singers [3].

One possible consequence of what Diehl, et al. [4] have called the *sufficient contrast* hypothesis, and what also follows from the cited differences found between bass and tenor singers [3], is that we should predict within a group of male or female speakers with a sufficient range of average f_0 values that the size of the acoustic vowel space be positively correlated with the speakers' average f_0. In other words, a female speaker with a high average f_0 should have a larger acoustic vowel space than a female speaker with a lower average f_0.

In an attempt to test this prediction the vowel spaces and average fundamental frequency of 87 (17 male and 70 female) speakers were measured [16]. The findings of this study were inconclusive. A significant positive correlation between vowel space size and f_0 was found for female speakers, but only after the two most extreme outliers had been removed. However, for another parameter predicted to be positively correlated with f_0, the Euclidean distance between [eː] and [aː], no significant correlation was found in the female group. In the male group, no significant correlations were found between f_0 and vowel space dimensions.

It cannot be excluded that [16] did not come to more substantial conclusions due to problems with respect both to the data they analysed as well as the analysis itself.

This present study sets out to improve on some of the problems which were identified in [16] and restricts the analysis to female speakers since this seemed to be the group most likely to exhibit such a correlation. A data set was collected that was designed to elicit corner vowels /iː, aː, uː/ in a temporally privileged context in which they could be expected to have reached their most peripheral acoustic/articulatory values under normal reading conditions. At the same time, in an attempt to minimize the influence of differences in interharmonic spacing on estimating formant frequencies [13], the corner vowels were placed in a context designed to elicit glottalisation and thereby reduce interharmonic spacing independent of a speaker's average f_0.

2. DATA AND METHOD

Sixty-one female students, primarily from eastern central Germany, who were enrolled in experimental phonetics classes read twenty sentences including the three target sentences listed in (1).

(1)
Sie fuhren letzte Woche zur IAA nach Frankfurt.
'They went to the IAA in Frankfurt last week.'
Sie fahren nächste Woche zur LUU nach Hannover.
'They're going to the LUU in Hannover next week.'
Wir wollen am Wochenende zur BII nach Hamburg.
'We want to go to the BII in Hamburg at the weekend.'

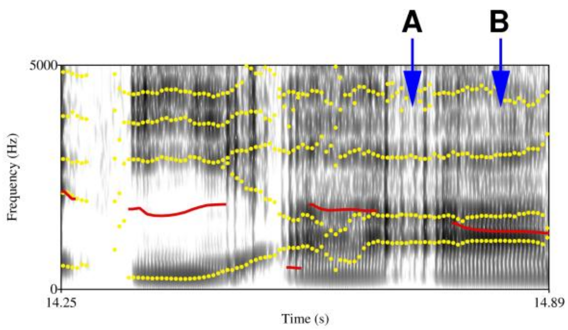
The target sentences were placed randomly throughout the list. The list was recorded three times for each speaker direct to a PC at a sampling rate of 22.05 kHz and 16-bit amplitude resolution using a condenser microphone (AKG C1000S) attached to a USB audio interface (M-AUDIO Fast Track Pro).

The underlined abbreviations in (1) were designed to elicit accented sequences of the corner vowels /iː, aː, uː/. German speakers are generally familiar with the abbreviation of the international car trade fair *IAA* so it was assumed that this abbreviation pattern would be produced analogously in the other same vowel sequences. And this was indeed the case.

Syllable-initial stressed vowels are generally produced with a degree of glottalisation in German [10, 14] so the same-vowel sequences were likely to exhibit a range of departures from a speaker's regular voicing: complete glottal closure, creaky phonation or, minimally, a reduction in f_0 and voicing amplitude [14]. As had been expected, the vast majority of speakers produced creaky phonation at the juncture between the two vowels, or a glottal stop which was often preceded and followed by a short period of creak.

For each of the 549 vowel tokens (= 61 speakers x 3 vowels x 3 repetitions), formant frequencies were estimated during the junctural stretch between the two same-vowel tokens by visual inspection of spectrograms and manually logging values identified by *praat*'s formant tracker [2]. A local estimation of the speaker's f_0 for each of the three vowel qualities was obtained by measuring f_0 in the middle of the second vowel in the same-vowel sequence, i.e. during the vowel portion once a speaker had returned to regular voicing. A token of *IAA* showing the formant tracks (dotted line) and f_0 curve (solid line) is shown in Figure 1. The arrow at A indicates a typical point during the glottalised junctural stretch during the same-vowel sequence where the formant measurement and that at B, the point at which f_0 was measured once regular voicing was made.

Figure 1: Figure 1 Spectrogram of IAA showing formant tracks (dotted line) and f_0 curve (solid line). Formant measurements were made at A, f_0 at B.



Apart from reducing the problems that arise from estimating formant values when f_0 is high and harmonics are spaced widely apart [13], glottalisation also reduced estimation problems in speakers who were otherwise using strong breathy voice. The vocalic stretch preceding A in Figure 1 is also an good example of the problems encountered by the formant tracker during a stretch of strong breathy phonation.

A global mean of each speaker's fundamental frequency was also calculated using the sixty sentences each speaker produced.

Means of formant and f_0 values of the three tokens of each vowel per speaker were calculated and converted following [20]. The difference between F1 and f_0 provides an acoustic measure of vowel openness [18] and the difference between F2 and F1 reflects the degree of frontness/backness [11]. These converted formants were then used to calculate two measures of vowel space dimension for each speaker: (a) the area of the vowel triangle in the converted F2 x F1 space, and (b) the difference between the converted F1 of /a:/ and /i:/, reflecting the vertical dimension of vowel space. The latter often represents one of the biggest non-uniform differences between male and female vowel spaces.

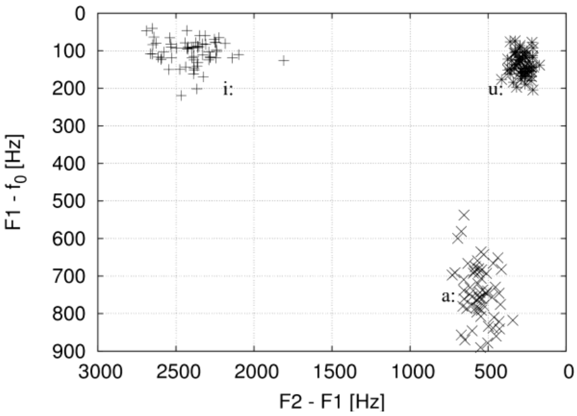
3. RESULTS

Global averages of speakers' fundamental frequency span 85 Hz, ranging from 171 to 256 Hz with an average of 203 Hz for the group. Furthermore, the distribution of mean f_0 values does not diverge significantly from a normal distribution (W = 0.9733, p = 0.2024). We will return to possible consequences of this in the discussion.

Local f_0 values, measured during the regular voicing portion of the second vowel in the same-vowel tokens also exhibited expected patterns. Group averages of 211 Hz for /i:/ and 216 Hz for /u:/ are insignificantly different, but are both significantly greater than the group average of 195 Hz for /a:/.

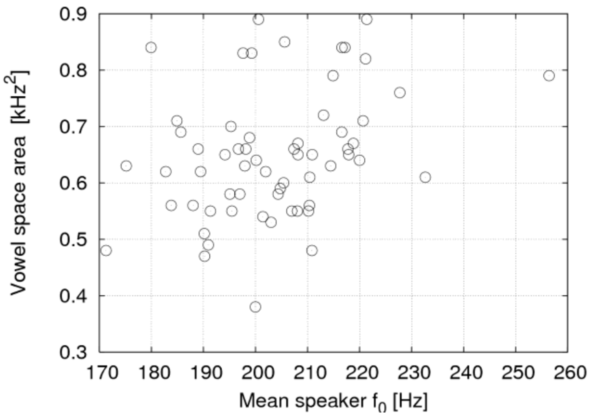
Figure 2 contains the plot of the transformed F1 (F1- f_0) and F2 (F2-F1) values. Due to the proximity of F1 and F2 for both /a:/ and /u:/, both vowels occupy the back of the vowel space. Values for /u:/ are tightly clustered. This is undoubtedly due to this vowel not only attaining maximum closeness and backness in this utterance context, but also having tight lip-rounding with protrusion [6]. By contrast, the other two vowels exhibit considerable interspeaker variation, with F1 varying most for /a:/, F2 for /i:/. There may be a number of reasons for this. Differences in F2 /i:/ could result from relatively small differences in the size of front oral cavity. The variation in F1 for /a:/ could result from at least two factors. First, it could arise from greater articulatory openness for some speakers. However, we must also expect a degree of intrinsic nasalisation in an open vowel of this type, independent of the fact that a nasal consonant is at the onset of the following word, which also gave rise to a degree of anticipatory nasalisation for some speakers. One of the acoustic correlates of this nasality, regardless of its provenance would be a lowered F1 [12, 17].

Figure 2: Speaker means of F1- f_0 as a function of F2-F1.



In order to investigate possible relationships between vowel space dimensions and a speaker's fundamental frequency, Pearson product moment coefficients were calculated using two different means of f_0 (global f_0 and the mean of local vowel f_0 values) and the two measures of vowel space dimension (vowel space area and F1 difference between /a:/ and /i:/). Figure 3 allows us to visually inspect one such relationship, in this case between global f_0 mean and vowel space area. Although Figure 3 does not necessarily make a relationship directly apparent, all correlations were positive and significant (df = 59, p < 0.05) with r ranging from a minimum of 0.3237 for the correlation between mean global f_0 and F1 difference between /i:/ and /a:/, and 0.3830 for the correlation between local f_0 mean and F1 differences. Correlations between mean f_0 and vowel space area lie between these extremes.

Figure 3: Vowel space as a function of global f_0 mean.



4. DISCUSSION

The findings of this study lend support to the hypothesis that there is a relationship between average speaker f_0 and vowel space size [4]. However, despite their significance, the correlations found are nevertheless weak, and the linear component accounts for only around 10 % of the variance, suggesting a number of other factors must be considered. It has been suggested that females speak more clearly under lab conditions than males [19] exhibiting, among other things, larger vowel spaces as a correlate of clarity, but we must also assume that within a group of females or males, individual speakers will make greater or smaller accommodations to the recording situation. Nasality's effect on F1, especially on the acoustic openness of /aː/, must also be considered. Although the vowel tokens were recorded in contexts designed to elicit maximally peripheral values, intrinsic nasality, as well as the possibility of early anticipatory nasality from the nasal consonant at the onset of the word *nach* are likely to be reducing F1 considerably [12, 17].

Finally, although there was a good range of average speaker f_0 spanning 85 Hz, the mean speaker f_0 values in this sample do not differ significantly from a normal distribution. At present, more females speakers are being actively recruited who have both higher and lower pitched voices in order to get a more evenly distributed sample.

5. REFERENCES

[1] Bladon, R., Henton, G., Pickering, J. 1983. Towards an auditory theory of speaker normalization. *Language and Communication* 4, 59-69.

[2] Boersma, P., Weenink, D. 2011. Praat: doing phonetics by computer. Version 5.2.16, retrieved 20.2.11 from *http://www.praat.org*.

[3] Cleveland, T.F. 1977. Acoustic properties of voice timbre types and their influence on voice classification. *J. Acoust. Soc. Am.* 61, 1622-1629.

[4] 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.

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

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

[7] Goldstein, U., 1980. *An Articulatory Model for the Vocal Tracts of Growing Children*. Ph.D. thesis, M.I.T., Massachusetts.

[8] Henton, C.G. 1992. The abnormality of male speech. In Wolf, G. (ed.), *New Departures in Linguistics*. New York: Garland, 27-59.

[9] Henton, C.G. 1995. Cross-language variation in the vowels of female and male speakers. *Proc. 13th ICPhS* Stockholm, 420-423.

[10] Kohler, K.J. 1994. Glottal stops and glottalization in German. Data and theory of connected speech processes. *Phonetica* 51, 38-51.

[11] Ladefoged, P., Maddieson, I. 1990. Vowels of the world's languages. *J. Phonetics* 18, 93-122.

[12] Maeda, S. 1993. Acoustics of vowel nasalization and articulatory shifts in French nasal vowels. In Huffman, M.K., Krakow, R.A. (eds.), *Nasals, Nasalization, and the Velum*. San Diego: Academic Press, 147-167.

[13] Maurer, D., Cook, N., Landis, T., d'Heureuse, C. 1991. Are measured differences between the formants of men, women and children due to f_0 differences? *J. Int. Phonetic Assoc.* 21, 66-79.

[14] Rodgers, J.E.J. 2000. The phonatory correlates of juncture in German. *Proceedings of the 5th Seminar on Speech Production: Models and Data*, Kloster Seeon, 289-292.

[15] Ryalls, J.H., Lieberman, P. 1982. Fundamental frequency and vowel perception. *J. Acoust. Soc. Am.* 72, 1631-1634.

[16] Simpson, A.P., Ericsdotter, C. 2007. Sex-specific differences in f_0 and vowel space. *Proc. 16th ICPhS* Saarbrücken, 933-936.

[17] Stevens, K.N. 1998. *Acoustic Phonetics*. M.I.T. Press, Massachusetts.

[18] Traunmüller, H. 1981. Perceptual dimension of openness in vowels. *J. Acoust. Soc. Am.* 69, 1465-1475.

[19] Whiteside, S.P. 1996. Temporal-based acoustic-phonetic patterns in read speech: Some evidence for speaker sex differences. *J. Int. Phonetic Assoc.* 26, 23-40.

[20] Whiteside, S.P. 2001. Sex-specific fundamental and formant frequency patterns in a cross-sectional study. *J. Acoust. Soc. Am.*110, 464-478.