ANALYSIS OF TEN VOWEL SOUNDS ACROSS GENDER AND REGIONAL / CULTURAL ACCENT

P Martland[‡†], S P Whiteside[‡], S W Beet[†] and L Baghai-Ravary[†]

[‡] Department of Speech Science,
[†] Department of Electronic and Electrical, Engineering,
University of Sheffield, UK.

ABSTRACT

This paper compares ten vowels sounds across gender and accent. Each formant for each vowel was analysed individually across data sets, but no comparison was drawn directly between the formant relationship within each vowel. The objective here was to examine the formants individually across gender and accent to establish a method for transforming vowel quality in a rule-based synthesis system and thus increase its range of voices. Further, it was hoped that it would make the comparison of English formant data across differing accents more simple.

Three sets of American English data were utilised in the analysis, and compared against two British English accents - Received Pronunciation (RP) and a General Northern accent (GN). Initial findings suggest that the relative positions of certain vowel formants are particularly static across gender, least variation being found with the second formant frequency. When accent was considered, a greater degree of variation occurred, this being predominantly found with mid-open and mid-closed vowel classes.

1. INTRODUCTION

The need for more choice in voice qualities is one of the major issues that has been addressed in speech synthesis in recent years [1, 2], especially when considering voice output communication aids (VOCAs) and the increasing needs of users of such devices. More emphasis has been placed on the research and production of more natural sounding male, female and child voices, made possible by the introduction of more powerful and flexible speech synthesisers and research tools [3]. However, when modelling a voice on more than one speaker, problems arise due to inter-speaker variability, reflecting the differing speaking styles of people which can occur even within controlled homogeneous dialectal sample sets. It therefore appears that a voice should only be modelled on a single speaker.

However, as the need for more synthetic voices incorporating extralinguistic and paralinguistic properties increases, the amount of analysis required also becomes greater. Further, for rule based synthesiser systems problems occur when trying to use data extracted, via acoustic analysis, from different speakers to model different extralinguistic or paralinguistic properties. This strategy may necessitate an overhaul of the rules in general to accommodate the parametric differences (e.g. segment durations, formant values, vowel turning points, etc.) between the speakers utilised in the modelling process. Thus, if methods could be found to reliably predict formant positions within the confines of a single modelled speaker's vowel space, then production of more voice qualities may be simplified.

2. COMPARING FORMANTS ACROSS GENDER

2.1 Data

Overall, five sets of data were analysed. Three of these sets were of American English, the data being provided by Peterson and Barney [4], Zahorian and Jagharghi [5], and Hillenbrand *et al.*[6]. Three sets of American English data were utilised as they provided three different levels of accental control. The Peterson and Barney data is much less controlled than the others, covering a wide range of accents; the Zahorian and Jagharghi data was recorded with speakers from four American regions (Southern, Mid-Atlantic, Northern, and New England) - half of which were from the Southern region. In contrast, 87% of the speakers recorded by Hillenbrand *et al.* were from Michigan's lower Peninsula, so the accent is quite well controlled. The British English RP data was taken from Gimson [7] while typical values for the British English GN accent was obtained by Whiteside [8].

The ten DARPAbet vowels /iy/, /ih/, /eh/, /er/, /ae/, /aa/, /ao/, /uh/, /uw/, /ah/ were chosen for the analysis, since these were common to all the data sets - apart from the vowel /ah/ which is not represented in the British GN accent. It should also be noted here that the /er/ vowel for the American data is retroflexed.

2.2 Data Representation

For all sets of data each formant was ordered in terms of its frequency value. This gave a direct comparison in terms of individual formant frequency values (table 1). Listing each vowel formant in order of its frequency value was chosen here purely for its simplicity; using the formant values themselves did not provide any advantage. The more perceptually relevant non-linear scaling (such as the mel or bark) were not employed for similar reasons, plus under certain circumstances they have shown to be of little or no advantage over linear scaling techniques [9].

F1	Zahorian		Hillenbrand		Peterson		Gimson		Whiteside	
	male	female	male	female	male	female	male	female	male	female
lowest	**iy**	**iy**	**iy**	**iy**	**iy**	**iy**	**iy**	**iy**	uw	iy
	uw	**uw**	**uw**	**uw**	**uw**	**uw**	**uw**	**uw**	iy	uw
	ih	**ih**	**ih**	**ih**	**ih**	**ih**	ih	uh	**uh**	**uh**
	uh	**uh**	**uh**	**uh**	**uh**	**uh**	uh	ao	**ih**	**ih**
	er	**er**	**er**	**er**	**er**	**er**	ao	ih	er	ao
	eh	**eh**	eh	ae	eh	ao	er	eh	ao	eh
	ah	**ah**	ae	eh	ao	eh	eh	er	eh	er
	ao	**ao**	**ah**	**ah**	**ah**	**ah**	**aa**	**aa**	**aa**	**aa**
	ae	**ae**	**ao**	**ao**	ae	aa	**ah**	**ah**	**ae**	**ae**
highest	**aa**	**aa**	**aa**	**aa**	aa	ae	**ae**	**ae**		

F2	Zahorian		Hillenbrand		Peterson		Gimson		Whiteside	
	male	female	male	female	male	female	male	female	male	female
lowest	**ao**	**ao**	**uw**	**uw**	**ao**	**ao**	**ao**	**ao**	uw	ao
	aa	uh	**ao**	**ao**	**uw**	**uw**	uh	aa	ao	uh
	uh	aa	**uh**	**uh**	**uh**	**uh**	aa	uh	**aa**	**aa**
	er	**er**	**ah**	**ah**	**aa**	**aa**	**uw**	**uw**	uh	uw
	ah	**ah**	**aa**	**aa**	**ah**	**ah**	**ah**	**ah**	**er**	**er**
	uw	**uw**	**er**	**er**	**er**	**er**	**er**	**er**	ae	ih
	eh	ae	**eh**	**eh**	**ae**	**ae**	**ae**	**ae**	**eh**	**eh**
	ae	eh	**ae**	**ae**	**eh**	**eh**	**eh**	**eh**	ih	ae
	ih	**ih**	**ih**	**ih**	**ih**	**ih**	**ih**	**ih**	**iy**	**iy**
highest	**iy**	**iy**	**iy**	**iy**	**iy**	**iy**	**iy**	**iy**		

F3	Zahorian		Hillenbrand		Peterson		Whiteside	
	male	female	male	female	male	female	male	female
lowest	**er**	**er**	**er**	**er**	**er**	**er**	er	ih
	uh	uw	**uw**	**uw**	**uw**	**uw**	uh	er
	uw	aa	uh	aa	**uh**	**uh**	ao	eh
	ah	uh	aa	ao	ah	ao	**uw**	**uw**
	ae	ao	ao	uh	ao	ah	ae	uh
	aa	ah	**ah**	**ah**	ae	aa	ih	iy
	eh	ae	**ae**	**ae**	aa	ae	eh	ao
	ao	eh	**eh**	**eh**	**eh**	**eh**	iy	aa
	ih	**ih**	**ih**	**ih**	**ih**	**ih**	aa	ae
highest	**iy**	**iy**	**iy**	**iy**	**iy**	**iy**		

Table 1. Each vowel formant is listed in ascending order of formant frequency for each gender. Those in bold text show a direct male/female correlation within their data set.

So that the data could be analysed statistically, a common baseline needed to be established. The variation in formant frequency for the same vowel sound was therefore overcome by making each of the individual vowel F1 formant frequencies proportional to the highest F1 formant frequency value. Thus, the formant in the highest position attained a value of 100%. The same procedure was repeated for the F2 and F3 formants (where possible). This also has the advantage that the F1 formant frequency is related to vowel height, while F2 has been related to backness. A relationship can thus be established between the averaged data analysed, and the relative positions of the vowels within the vowel plane. Further, this does not restrict the use of the comparative method to large data sets, and will allow comparison of single speaker data.

2.3 Data Analysis

For each data set, the male and female data was arranged so that the order of the vowels were identical. For this, the male data was used as a reference and the female data altered accordingly. The difference between the male and female data was then calculated. This represented the degree of error in transposing the female values to male values. The mean and standard deviations were then calculated for this difference over all the vowels for each formant in the data sets. This gave a clearer indication of the relationship between the male and female data sets in terms of vowel space, as opposed to formant frequencies.

3. COMPARING FORMANTS ACROSS ACCENT

The formants were compared and analysed in a similar manner across accent. Here, only the first two formants were examined, since the Gimson data lacked values for F3. The male data was compared first, followed by the female data. Each was analysed separately, so the differences inherent within each gender group could be observed.

For examination with regard to accent, the mean and standard deviation were calculated for each vowel over the five data sets for both males and females. The mean and standard deviations across all the vowels were also found.

4. RESULTS

4.1 Comparison Across Gender

For each of the five data sets, the following results were obtained for the difference between the male and female formant frequency positions across all vowels (Table 2).

FORMANT	MEAN	STD	VARIANCE	
F1	0.9	3.36	11.29	**Zahorian & Jagharghi**
F2	3.6	1.62	2.64	
F3	-0.9	1.92	3.69	
F1	1.1	4.3	1.43	**Hillenbrand et al.**
F2	1.4	1.43	2.04	
F3	-3.6	1.92	79.64	
F1	1.8	5.23	27.36	**Peterson & Barney**
F2	1.3	2.61	6.81	
F3	-5	8.92	79.64	
F1	10.7	5.66	32.01	**Gimson**
F2	2	2.45	6	
F1	14.7	8.32	69.28	**Whiteside**
F2	9.6	9.31	86.96	
F3	0.22	7	49.06	

Table 2. Comparison of the error incurred in translating the female formant positions to match the male. The results show the errors across all vowels for each formant in terms of the percentage F1, F2, and F3 vowel space.

4.2 Comparison Across Accent

For the comparison across accent, the male and female data was analysed separately. The results of the analysis are given in table 3 for the first formant and second formant frequencies respectively. Unlike the analysis across gender, the results give the mean and standard deviation of each vowel as well as across all vowels. In a similar manner to the cross gender analysis, the female data was transposed in terms of the male vowel formant position.

These results are also presented graphically in figures 1 to 4.

	Female F1			Male F1		
VOWEL	MEAN	STD	VAR	MEAN	STD	VAR
iy	35.8	5.95	35.36	40.4	4.22	17.84
uw	39.8	6.18	38.16	43.2	2.99	8.96
ih	47.6	3.83	14.64	54.8	2.14	4.56
uh	48.4	7.79	60.64	58.6	1.85	3.44
er	55.4	6.5	42.24	65.2	4.07	16.56
eh	65.6	7	49.04	74.6	1.62	2.64
ao	66.8	15.93	253.76	76.6	8.87	78.64
ah	82.25	3.34	11.19	86	6.04	36.5
ae	93	11.24	126.4	91	8.58	73.6
aa	89	13.31	177.2	94	9.3	86.4
ALL	62.9	8.18	66.98	68.2	2.38	5.66

	Female F2			Male F2		
VOWEL	MEAN	STD	VAR	MEAN	STD	VAR
ao	37	5.06	25.6	40.6	6.97	48.64
uw	45.8	8.08	65.36	48.6	8.45	71.44
aa	48.2	4.79	22.96	51.4	3.98	15.84
uh	45.2	3.06	9.36	48.8	4.07	16.56
ah	52.5	2.18	4.75	54.5	2.6	6.75
er	58.4	2.58	6.64	60.8	2.93	8.56
ae	73.4	6.59	43.44	76.4	4.96	24.64
eh	76.2	7.73	59.76	83.6	4.5	20.24
ih	80.2	9.95	98.96	87	1.55	2.4
iy	100	0	0	100	0	0
ALL	66.3	4.1	16.81	69	2.84	8.05

Table 3. Indicating the mean and standard deviations of vowel formant position. The very strong position of the vowel /iy/ as the highest position in the F2 vowel plane can be clearly observed.

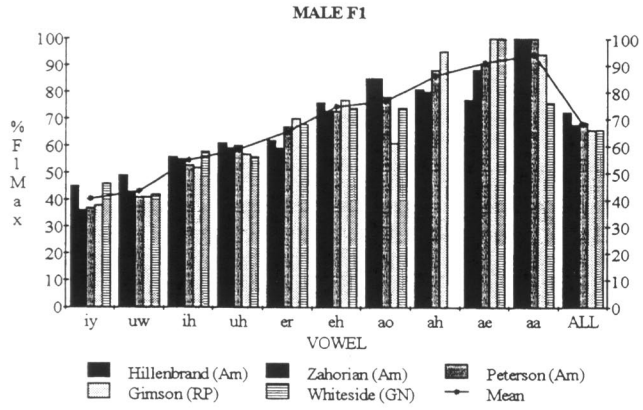

Figure 1. Variation in the male first formant as a function of the highest F1 position.

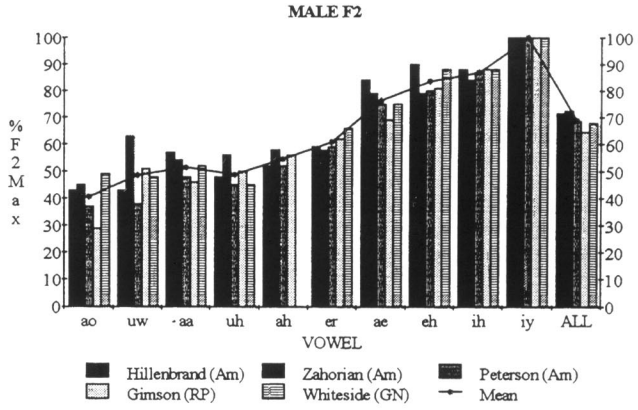

Figure 2. Variation in the male second formant as a function of the highest F2 position.

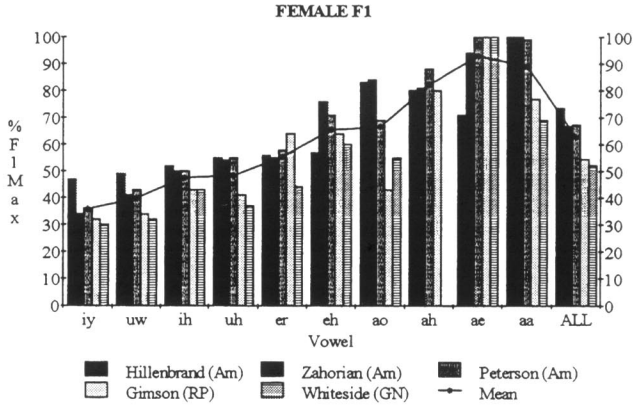

Figure 3. Variation in the female first formant as a function of the highest F1 position.

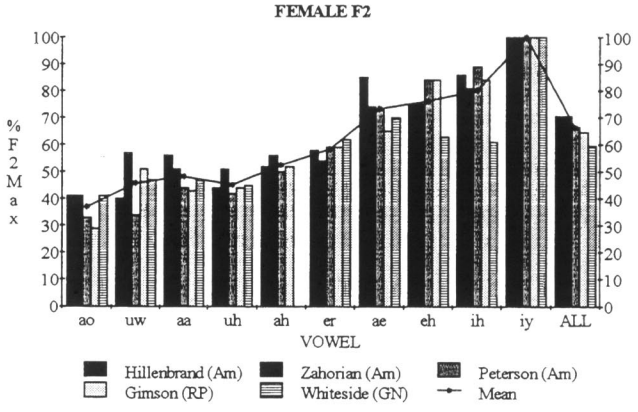

Figure 4. Variation in the female second formant as a function of the highest F2 position.

5. DISCUSSION

Table 2 shows the error incurred in translating the female formant values to match the corresponding male positions in each data set. This difference between them is thus an indication of how well each set of data correlates in terms of position across gender. The low standard deviation and variance factors found across the data sets is indicative of a closely correlated F2 formant position, in terms of vowel space for males and females. The only exception to this is Whiteside's British English General Northern data, where a comparatively large standard deviation and variance can be observed.

In fact, this data shows that the F2 (figure 4) parameters for the female front vowels /ae/ ,/eh/, and /ih/ in the British English GN accent are far lower than their male counterparts (figure 2), the error between the male and female data being as great as 30% for /eh/, and /ih/. In general, the female F2 values for the back vowels are also less than when compared to the male, although to a very much smaller degree than the front vowels. The F1 formant positions for both the British English RP, and the General Northern data follow a similar trend to the F2 data of the General Northern accent, where, on average, the female formant position tends to be lower.

This may imply that British English women articulate slightly differently than men within the same accental grouping, for example with more lip-spreading and tongue backing. However, this difference could also have been dependent upon the degree of control when recording subjects for analysis. A wide range of speaking styles would possibly still contribute and fall within the heading of the British English General Northern accent.

Comparing the data across accent, the differences in standard deviation for individual vowels shows clearly larger values for both F1 and F2 parameters when compared to the male data. Examination of figures 1 to 4 does show a far greater deviation from the mean value curves for the British English Female data, and this includes both accents. As a result, the average standard deviations over all vowels are substantially higher than those obtained with the male data. The greatest average deviation for the female F1 is thus clearly a result of the influence of the female speakers of both British accents considered. The overall highest deviation of 15.93 for the vowel /ao/ is clearly justifiable, as can be seen in figure 3.

6. CONCLUSIONS

Certain vowels appear to remain fairly static in their relative positions within a speaker's vowel space across gender, as seen in table 1. This is particularly so for the second formant frequency, which therefore may act as a good gender identifier. However, large translation errors are possible, as observed with the General Northern accent data, and these deviations increase as different accents are considered. However, for scaling between different accents in terms of synthetic speech, the data shows that this is plausible based upon the vowel space of a single speaker. On the other hand, it does also suggest that for this to be effective, all the vowels must be scaled in relative proportions to the vowel space of the speaker being modelled upon. A differing accent in terms of vowel quality may not be fully achieved by altering a single vowel. Further, due to the scaling in terms of vowel space rather than formant frequency, it would not be expected to scale across gender.

In retrospect, a more non-linear scaling for the data representation and results may have provided a more clear realisation of vowel position, within the vowel space continuum. It is hoped to test whether the methodology employed here to examine formants has been useful in the near future.

7. ACKNOWLEDGEMENTS

PM is supported by an EPSRC CASE award from Barnsley District General Hospital NHS Trust. LB-R is employed on the EC-TIDE project, "Voice, Attitude, and Emotions Speech Syntheses" (VAESS). The authors are grateful to both the Barnsley District General Hospital and the VAESS project for their assistance and co-operation.

8. REFERENCES

1. Karlsson, I. "Female voices in speech synthesis", *Journal of Phonetics, Vol. 19, 1991*, p 113.

2. Karlsson, I. "Modelling voice variations in female speech synthesis", *Speech Communication, Vol. 11, 1992*, pp 491-495.

3. Carlson, R., Granström, B., Karlsson, I. "Experiments with voice modelling in speech synthesis", *Speech Communication, Vol. 10, 1991*, pp. 481-489.

4. Peterson, G. E., Barney, H. L., "Control methods used in the study of vowels", *JASA, Vol. 32(6), 1952*, pp 693-703.

5. Zahorian, S. A., Jagharghi, A. J. "Spectral shape versus formants as acoustic correlates for vowels", *JASA, Vol. 94(4), 1993*, pp 1966-1982.

6. Hillenbrand, J., Getty, L. A., Clark, M. J., Wheeler, K. "Acoustic characteristics of American English Vowels", *JASA, Vol. 97(5), Pt. 1, 1995*, pp 1-13.

7. Gimson, A. C., *An Introduction to the Pronunciation of English*, Edward Arnold Publishers, London, 1989.

8. Whiteside, S. P., *Towards Improved Synthesis of Women's Speech: British General Northern Accent,* Ph.D. thesis, University of Leeds, 1992.

9. Hillenbrand, J., Gayvert, R. T. "Vowel classification based upon Fundamental Frequency and Formant Frequencies", *Journal of Speech and Hearing Research, Vol. 36, 1993*, pp 694-700.