Journal of the International Phonetic Association (1997) 27, 47–55.

The Formants of Monophthong Vowels in Standard Southern British English Pronunciation

DAVID DETERDING
National Institute of Education, Nanyang Technological University

The formants of the eleven monophthong vowels of Standard Southern British (SSB) pronunciation of English were measured for five male and five female BBC broadcasters whose speech was included in the MARSEC database. The measurements were made using linear-prediction-based formant tracks overlaid on digital spectrograms for an average of ten instances of each vowel for each speakers, These measurements were taken from connected speech, allowing comparison with previous formant values measured from citation words. I was found that the male vowels were significantly less peripheral in the measurements from connected speech than in measurements from citation words.

1. Introduction

Many of the standard formant values for English vowels have depended on citation words spoken specially for the purpose of obtaining the measurements. For example, Gimson and Ramsaran (1989:100) used measurements of vowels from a single speaker from an unpublished thesis by Wells; and Cruttenden (1994:96) quotes figures from Deterding (1990) which were based on citation words ([hVd] words such as ‘heed’, ‘hid’, ‘head’ …) read by eight male and eight female speakers.

Modern advances in technology have made measurements of the vowels of continuous speech both easier and more reliable. Two advances in particular have made this possible: the availability of standard speech corpora; and the development of improved formant tracking software.

The measurements that are provided in this study are from the MARSEC database (Roach, Knowles, Varadi and Arnfield, 1993), so they can easily be checked or developed further by other researchers. This database consists of broadcasts from the BBC, so the data represents a style of speech that may be familiar to many people throughout the world through listening to the BBC World Service. This style of speech might be regarded as a kind of reference speech, in the sense that it is used as a model for pronunciation in many parts of the world, though of course it may differ considerably from the sort of speech that would be uttered in ordinary conversation.

2. Data

The MARSEC database consists of a set of monologues, such as newsreading and commentary, broadcast by the BBC during the 1980’s. The corpus is available on a CD-ROM. In each directory on the disk, there are a set of files from a single recording. Although some of the directories may have contributions from a number of different speakers, as, for example, when a news broadcast includes a report from journalists on site, it is always possible to find a reasonable stretch of speech from the broadcaster whose voice is heard first in the first file in each directory.

The present study considers the vowels of ten speakers, five male and five female. They were taken from the directories indicated in Table 1.

Table 1. Location of the speakers in the MARSEC database.

Directory	Sex	Contents
ASIG	female	religious affairs broadcast
BSIG	male	newsreading
CSIG	male	economics lecture
DSIG	female	arts lecture - on Dada
ESIG	female	prayers and Bible reading
FSIG	female	financial and share analysis
GSIG	female	story reading
HSIG	male	poetry reading
JSIG	male	report from a sports meeting
KSIG	male	discussion on employment

The speaker from the start of directory ASIG are referred to as speaker A, from BSIG as speaker B, and so on.

All the speakers have what might be termed a Standard Southern British accent (similar to RP), though there is inevitably a little variation between them. This can affect voice quality, so that speaker E has a very breathy voice, and speaker F has frequent use of creaky phonation in the middle of some words; and it can also affect the quality of some vowels, so that speaker H has an old-fashioned less open /æ/ than others (close to [ɛ]); and speaker K has some traces of a Northern accent with a few instances of a fronted vowel instead of /ɑː/ in ‘pass’ and ‘chance’ (these words were ignored in measuring the /ɑː/ vowel). However, the accentual differences between the different speakers is small, and we can assume that “the accent of all the speakers is RP or close to it” (Roach *et al.*, 1993:48).

3. Measurements

The formant measurements were made using the CSL software from Kay running on a 486 PC. Clear instances of each vowel were identified by listening, and then digital spectrograms were derived, with overlaid linear-prediction-based formant tracks, using a pre-emphasis coefficient of 0.9. The speech in the MARSEC database is digitized at 16 kHz, and after following the advice of Ladefoged (1996:212) to try out different analyses and “see which gives the most interpretable results”, 16^{th}-order linear prediction was used for all the data. In fact, this was insufficient in some cases, and there was no clear first formant for some tokens of open vowels such as /æ/ and /ʌ/. It is possible that, for these cases, a higher order linear prediction filter would be more appropriate, perhaps an 18^{th} order to follow the rule of thumb proposed by

Ladefoged (1996:212) of one linear prediction coefficient for each kHz of the sample rate, plus an additional two; but it was decided to keep consistent settings for all the measurements. In cases where the first two formants of vowels could not be reliably measured, alternative tokens were selected.

Difficulties in clear identification of both the first and second formants of all vowels are well known. Ladefoged (1967), using traditional analog spectrographic equipment, reported that separation of the first and second formant for back vowels was particularly difficult, even for the cardinal vowels of trained phoneticians. In contrast, for the computer-based measurements made in this study of the MARSEC data, the linear-prediction-based formant tracks generally achieved quite clear separation of the first two formants of back vowels; but, as mentioned above, it was more often the first formant of open vowels, such as /æ/ and /ʌ/, that caused problems. Nevertheless, it was possible to find reasonably consistent first and second formants for at least some tokens of the eleven vowels of all ten speakers.

Table 2. Average values of F_1, F_2 and F_3 in Hz.

	Male			Female		
	F_1	F_2	F_3	F_1	F_2	F_3
iː	280	2249	2765	303	2654	3203
ɪ	367	1757	2556	384	2174	2962
e	494	1650	2547	719	2063	2997
æ	690	1550	2463	1018	1799	2869
ʌ	644	1259	2551	914	1459	2831
ɑː	646	1155	2490	910	1316	2841
ɒ	558	1047	2481	751	1215	2790
ɔː	415	828	2619	389	888	2796
ʊ	379	1173	2445	410	1340	2697
uː	316	1191	2408	328	1437	2674
ɜː	478	1436	2488	606	1695	2839

Measurements of the first three formants were made for about ten tokens of each of the eleven monophthong vowels for each speaker. For most vowels of most speakers, there were many tokens that could be selected, and in such cases, vowels that occurred after the approximants /j/, /w/ and /r/ or before /l/ were avoided, as these approximants would have severe coarticulatory effects on the locations of the first three formants. However, for some vowels, particularly /ʊ/ and /uː/, it was not always possible to find enough tokens if these environments were avoided.

In no case were fewer than five tokens of a vowel measured, with the exception of /ʊ/ for two speakers: only two clear tokens of this vowel could be found for speaker A, and two for speaker E.

4. Results

The average values for F_1 and F_2 in Hz for the male and female speakers are shown in the Table 2. The average values for the individual speakers are shown in the Appendix.

Plots of the average male and female values are shown in Figures 1 and 2. The values have been converted to the auditory Bark scale, using the formula suggested by Zwicker and Terhardt (1980), where F is the frequency in Hertz and Z the frequency in Bark:

$$Z = 13 \arctan(0.00076F) + 3.5 \arctan(F/7500)^2$$

(The values of F_1 and F_2 in Bark are shown in Tables 4 and 5 below.)

Figures 1 and 2 show simple plots of F_1 against F_2. Many researchers prefer other kinds of plots to show the nature of vowels. For example, Ladefoged and Maddieson (1990) suggest that the difference between F_1 and F_2 gives a better representation of backness than F_2 alone. Let us consider such a scale briefly.

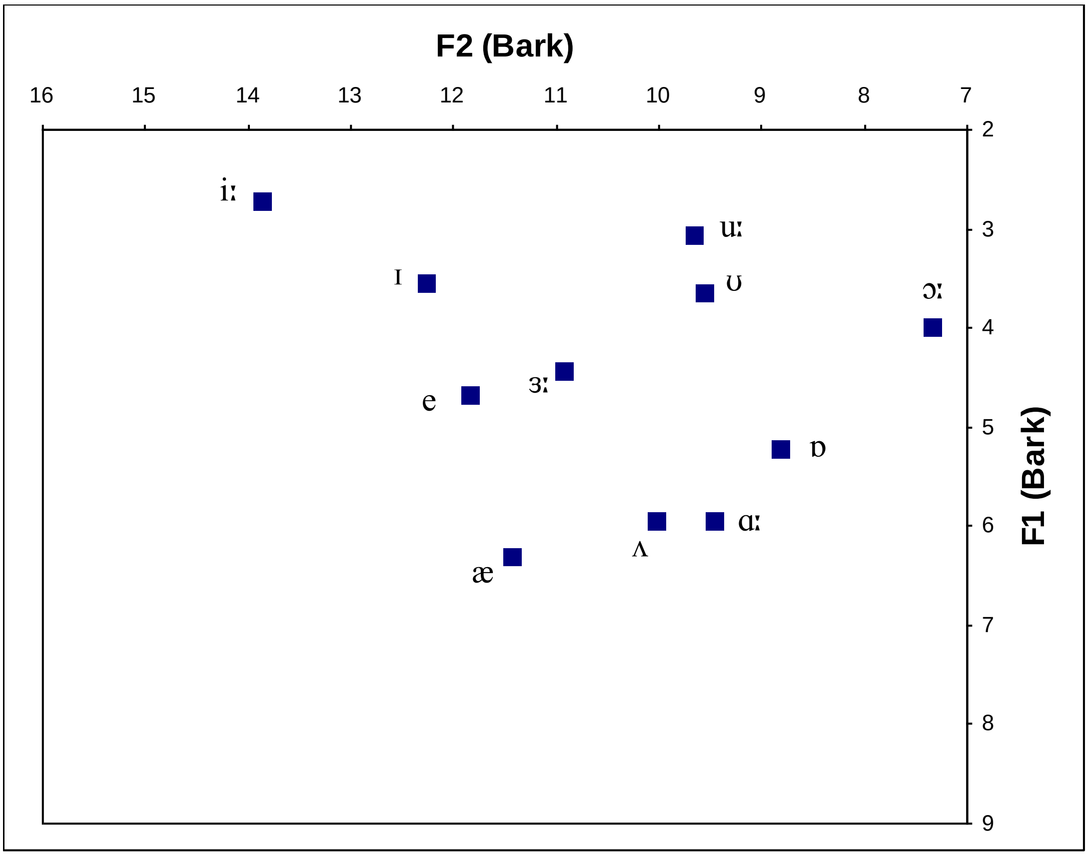

Figure 1. F_1/F_2 values for average male vowels.

The main effect of representing the front/back dimension in terms of F_2 - F_1 would be to normalize for speaker differences, particularly male-female differences in formant frequencies. Some researchers, such as Traunmüller (1981), suggest that, in addition to using F_2 - F_1 as a speaker-independent measure of vowel frontness, F_1 - F_0 can be used as a speaker-independent measure of vowel openness, as the fundamental frequency F_0 can serve to normalize the differences between male and female first formants. However, if F_1 - F_0 were really to provide a speaker-independent indication of vowel openness, then we would expect that, for the same vowel quality, a speaker should have a higher F_1 when speaking on a high pitch than when speaking on a low pitch; and the measurements of Ladefoged (1967) of phoneticians uttering the cardinal vowels on different pitches indicate that this kind of shift in F_1 does not occur. In fact, a speaker-independent measure of vowel quality is still elusive.

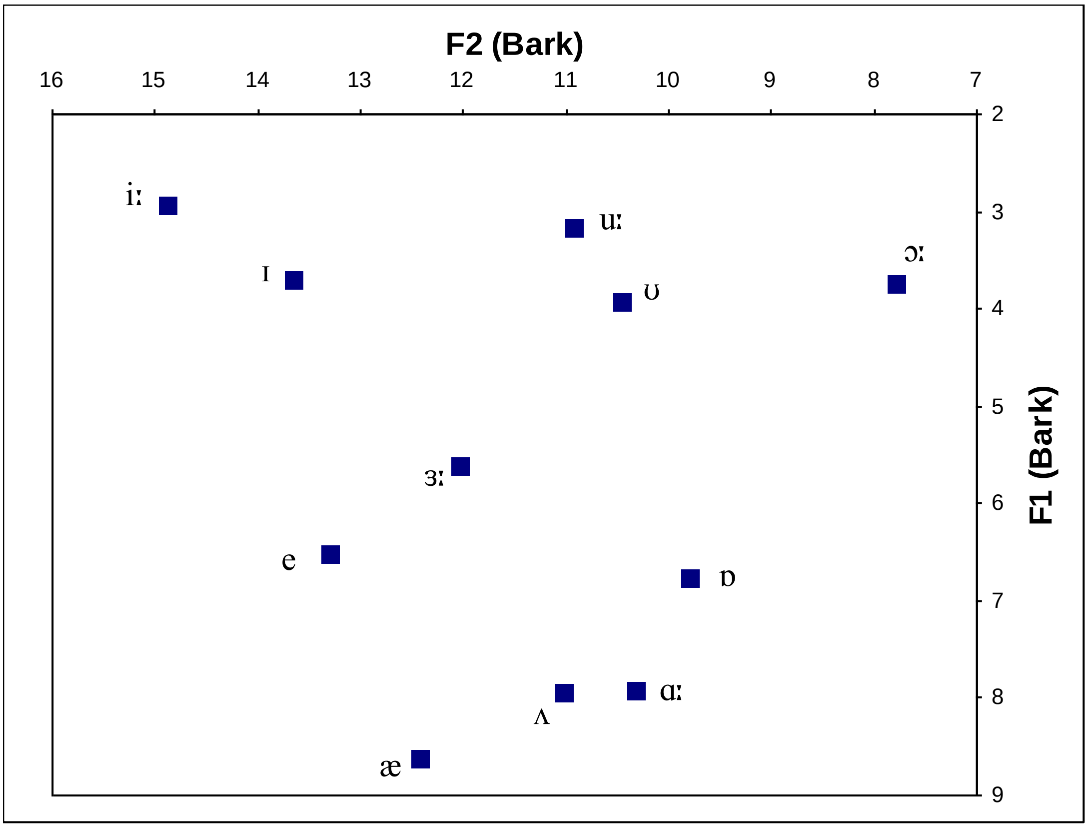

Figure 2. F_1/F_2 values for average female vowels.

Given that the best way to represent vowel quality is not certain, a simple plot of F_1 against F_2 is shown here. In considering Figures 1 and 2, one must remember that there is not necessarily an absolute link between vowel openness and F_1 or between vowel frontness and F_2. For example, Kent and Read (1992:93) stress that a single vowel quality can be associated with more than one formant pattern.

5. Comparison with Previous Data

We can now look at these measurements of the MARSEC vowels and compare them with previous measurements of citation forms, to try to determine the effect of taking vowels from connected speech.

Table 3. Average male values of F_1 and F_2 in Hz for connected speech (from the MARSEC database) compared with citation forms (from Deterding (1990)).

	connected		citation	
	F_1	F_2	F_1	F_2
iː	280	2249	275	2221
ɪ	367	1757	382	1958
e	494	1650	560	1797
æ	690	1550	732	1527
ʌ	644	1259	695	1224
ɑː	646	1155	687	1077
ɒ	558	1047	593	866
ɔː	415	828	453	642
ʊ	379	1173	414	1051
uː	316	1191	302	1131
ɜː	478	1436	513	1377

Table 3 allows a comparison of the average male vowels from MARSEC against the citation forms from Deterding (1990), and Table 4 shows the same comparison for female vowels. Only the first two formants are shown, as measurements of the third formant are not available from the earlier data.

Table 4. Average female values of F_1 and F_2 in Hz for connected speech (from the MARSEC database) compared with citation forms (from Deterding (1990)).

	connected		citation	
	F_1	F_2	F_1	F_2
iː	303	2654	319	2723
ɪ	384	2174	432	2296
e	719	2063	645	2287
æ	1018	1799	1011	1759
ʌ	914	1459	813	1422
ɑː	910	1316	779	1181
ɒ	751	1215	602	994
ɔː	389	888	431	799
ʊ	410	1340	414	1203
uː	328	1437	339	1396
ɜː	606	1695	650	1593

Table 5. Average male values of F_1 and F_2 in Bark and distances from the centroid for connected speech (from the MARSEC database) compared with citation forms (from Deterding (1990)).

	connected			citation		
	F_1	F_2	distance	F_1	F_2	distance
iː	2.73	13.85	3.83	2.68	13.77	4.19
ɪ	3.54	12.26	2.04	3.68	12.97	3.02
e	4.68	11.84	1.39	5.25	12.40	2.31
æ	6.31	11.42	2.03	6.63	11.32	2.20
ʌ	5.94	10.02	1.50	6.35	9.83	1.61
ɑː	5.96	9.45	1.77	6.28	8.99	1.90
ɒ	5.23	8.81	1.81	5.53	7.61	2.64
ɔː	3.98	7.34	3.16	4.32	5.93	4.24
ʊ	3.65	9.55	1.25	3.97	8.83	1.54
uː	3.07	9.65	1.66	2.94	9.31	2.01
ɜː	4.54	10.91	(0.44)	4.85	10.62	(0.49)
ave	4.51	10.46	2.04	4.77	10.14	2.57

Table 6. Average female values of F_1 and F_2 in Bark and distances from the centroid for connected speech (from the MARSEC database) compared with citation forms (from Deterding (1990)).

	connected			citation		
	F_1	F_2	distance	F_1	F_2	distance
iː	2.95	14.87	4.26	3.10	15.03	4.44
ɪ	3.70	13.64	2.82	4.14	13.98	3.03
e	6.53	13.30	2.06	5.95	13.96	2.81
æ	8.62	12.41	3.22	8.58	12.26	3.39
ʌ	7.94	11.01	2.45	7.24	10.84	1.91
ɑː	7.92	10.32	2.65	6.99	9.60	2.29
ɒ	6.78	9.78	2.11	5.60	8.47	2.75
ɔː	3.75	7.77	4.14	4.13	7.13	4.26
ʊ	3.94	10.44	1.92	3.97	9.72	2.04
uː	3.18	10.91	2.43	3.29	10.72	2.13
ɜː	5.63	12.02	(0.53)	5.99	11.60	(0.74)
ave	5.54	11.50	2.81	5.36	11.21	2.90

One might expect the citation vowels to be more peripheral than the vowels from connected speech, partly because of the effects of coarticulation with neighbouring consonants, and particularly because one would expect fluent speakers to economize somewhat in their vocal effort in connected speech (Lindblom, 1983). In order to estimate how peripheral the vowels are, we can calculate the distance in Bark (using a simple Euclidean distance) of each vowel (except the central vowel /ɜː/) from the centroid of all vowels (which is calculated as the average value of F_1 and F_2). For this purpose, all the values of Tables 3 and 4 are repeated in Tables 5 and 6, with the values converted to Bark.

The lowest right-hand figures give average distances from the centroid of 2.04 and 2.57 Bark for male connected and citation speech, and 2.81 and 2.90 Bark for female connected and citation speech. Though these figures suggest that the citation speech may be more peripheral, the difference is only statistically significant (using a correlated samples t-test) for the male speech ($t=4.29$, $df=9$, $p<0.01$) and not for the female speech ($t=0.77$, $df=9$, $p>0.05$).

In comparing the data for connected speech against citation words, one should be careful, as the data are for different speakers under different conditions. We have no way of knowing how the BBC speakers might have produced citation words.

6. Conclusion

Some new measurements of the vowels of Standard Southern British English pronunciation have been presented. As these vowels are taken from reasonably natural connected speech, they represent somewhat different data from the more common citation forms, so they may be a little less artificial than tokens derived from specially articulated citation speech. It is hoped that these measurements can serve as a reference for other researchers, and, since these data come from a standard database, it is also hoped that others will be able easily to monitor their accuracy, build upon them in further studies.

References

CRUTTENDEN, A. (1994). *Gimson's Pronunciation of English*. Fifth edition of GIMSON, A.C. (1962). *An introduction to the pronunciation of English*. London: Edward Arnold.

DETERDING, D. (1990). Speaker Normalization for Automatic Speech Recognition. Ph.D. Thesis, Cambridge University.

GIMSON, A.C. & RAMSARAN, S. (1989). *An Introduction to the Pronunciation of English* (4th Edition). London: Edward Arnold.

KENT, R.D. & READ, C. (1992). *The Acoustic Analysis of Speech*. San Diego: Singular Publishing Group.

LADEFOGED, P. (1967). *Three Areas of Experimental Phonetics*. Oxford: Oxford University Press.

LADEFOGED, P. (1996). *Elements of Acoustic Phonetics* (Second Edition). Chicago: University of Chicago Press.

LADEFOGED, P. & MADDIESON, I. (1990). Vowels of the world's languages. *Journal of Phonetics*. 18, 93-122.

LINDBLOM, B. (1983). Economy of human gesture. In *The Production of Speech* (ed. P. F. MacNeilage). New York: Springer-Verlag.

ROACH, P., KNOWLES, G., VARADI, T. & ARNFIELD, S. (1993). MARSEC: A machine-readable spoken English corpus. *Journal of the International Phonetic Association*, 23, 47-54.

TRAUNMÜLLER, H. (1981). Perceptual dimension of openness in vowels. *Journal of the Acoustical Society of America*, 69(5), 1465-1475.

ZWICKER, E. & TERHARDT, E. (1980). Analytical expression for critical-band rate and critical bandwidth as a function of frequency. *Journal of the Acoustical Society of America*, 68(5), 1523-1525.

Appendix

Table A1. Average formant values for each of the vowels of each of the male speakers.

	B			C			H			J			K		
	F1	F2	F3	F1	F2	F3	F1	F2	F3	F1	F2	F3	F1	F2	F3
iː	281	2016	2337	276	2218	3090	280	2600	3128	302	2008	2517	261	2402	2752
ɪ	335	1430	2198	396	1659	2592	367	1987	2887	395	1670	2450	344	2041	2653
e	490	1397	2127	509	1520	2590	444	1923	2902	512	1587	2544	515	1823	2573
æ	661	1328	2139	546	1542	2306	579	1769	2790	790	1558	2559	872	1555	2522
ʌ	635	1237	2186	537	1219	2383	687	1382	2833	704	1204	2553	659	1251	2798
ɑː	694	1202	2183	540	1108	2195	625	1165	2738	649	1117	2524	720	1185	2811
ɒ	611	1113	2111	482	1042	2200	609	1125	2753	558	1000	2574	530	956	2769
ɔː	419	906	2157	397	709	2627	448	925	2802	425	835	2657	388	764	2854
ʊ	370	1195	2055	378	1323	2332	391	1136	2642	387	1268	2391	368	945	2804
uː	321	1247	2149	298	1373	2234	327	1123	2659	343	1343	2404	291	870	2593
ɜː	472	1265	2183	507	1397	2482	523	1468	2748	462	1398	2523	425	1651	2506

Table A2. Average formant values for each of the vowels of each of the female speakers.

	A			D			E			F			G		
	F1	F2	F3	F1	F2	F3	F1	F2	F3	F1	F2	F3	F1	F2	F3
iː	304	2664	3248	284	2694	3315	300	2582	3234	321	2606	3161	306	2725	3055
ɪ	365	2157	2953	387	2215	2960	410	2070	3032	392	2147	2887	364	2279	2977
e	853	2054	3056	620	2157	2968	634	1926	2992	738	2065	2906	750	2114	3063
æ	1067	1690	2791	971	1892	2761	1045	1766	3121	972	1884	2744	1033	1761	2928
ʌ	1044	1495	2740	950	1512	2851	843	1464	2929	875	1489	2638	860	1335	2998
ɑː	1010	1304	2815	903	1305	2876	903	1393	2945	895	1327	2685	837	1250	2883
ɒ	761	1243	2661	765	1216	2791	680	1249	2869	823	1243	2651	727	1123	2980
ɔː	398	934	2669	373	849	2778	334	959	3027	427	876	2689	412	823	2817
ʊ	391	1798	2627	421	1361	2740	415	1234	2702	406	1199	2638	418	1109	2780
uː	333	1529	2657	319	1521	2627	328	1396	2746	343	1437	2683	316	1302	2657
ɜː	443	1762	2663	746	1627	2842	517	1676	2953	695	1705	2762	631	1704	2974

The individual values are available at:

http://videoweb.nie.edu.sg/phonetic/data/jipa-vowels/index.htm