MULTI-SPEAKER FORMANT DATA ON THE AUSTRALIAN ENGLISH VOWELS: A TRIBUTE TO J.R.L. BERNARD'S (1967) PIONEERING RESEARCH

Frantz Clermont

School of Computer Science,
University College, University of New South Wales,
Australian Defence Force Academy,
Canberra, ACT 2601, AUSTRALIA

ABSTRACT - In this paper we endeavour to reinforce the importance of Bernard's (1967) pioneering study of the Australian English vowels, by attempting a more complete account of the extent and the diversity of his *multi-speaker*, spectrographic estimates of the first three formant frequencies (F1, F2 and F3). Through a careful restoration of these parameters we show that, similarly to Peterson and Barney's (1952) seminal study of the American English vowels, Bernard's work has yielded an invaluable contribution to that body of data, which are still to date difficult to acquire automatically from the speech signal, but which continue to provide theoretically-robust probes into the articulatory and acoustic processes of speech communication.

INTRODUCTION

The motivation for this endeavour arose in our current acoustic studies of Australian English, for which we sought to acquire some comprehensive measurements of the first three formant-frequencies (F1, F2 and F3) of spoken vowels. We are interested in this acoustic space because it affords very low dimensionality, but more importantly because it lends itself more directly to phonetic and articulatory interpretations of the speech signal. Besides, there is some recent evidence (Furui and Akagi, 1985; Kuwabara and Takagi, 1991; Kitamura and Akagi, 1994; Mella, 1994; Mokhtari and Clermont, 1994; 1996), which clearly substantiates a long-standing speculation (Lewis and Tuthill, 1940; Peterson and Barney, 1952; Stevens, 1971) that the high spectral regions of the monophthongal vowels (spanning ranges of F2 and F3) contain relatively more speaker-specific information. In the same vein, some promising observations have been reported on the Australian English diphthongs (Mitchell and Delbridge, 1965; Burgess, 1969; Sampath, 1995), which collectively indicate that the upper formant space of these more complex vowels could provide important insights into dialect-induced variability.

While the formant frequencies represent a desirable parameter set for reasons such as those briefly exposed above, its estimation from the speech signal is fraught with persistent difficulties which cannot yet be completely overcome by currently available techniques. This is exemplified by the scarcity of comprehensive formant data since Peterson and Barney's spectrographic study of the American English vowels. However, our search for such data on Australian speech prompted our renewed interest in Bernard's (1967) remarkable spectrographic study of the vowels of that language, which were recorded in /hVd/ context by a total of 170 adult male, native speakers of three idiolects referred to as Broad, General and Cultivated. In comparison with Peterson and Barney's study which has played a seminal role in acoustic-phonetic research, Bernard's analogous endeavour in Australia is not well-known but no less significant. Indeed, the formant data collected by Bernard not only embody comparable phonetic complexity and speaker dimensionality, but also include an auditory assessment of idiolectal variations amongst the spoken vowels of Australian English.

Unfortunately, however, Bernard's published legacy to this aspect of speech research has to date been a primarily speaker-averaged representation of his formant estimates. Although this type of data reduction may well have facilitated the process of building an acoustic-phonetic framework for Australian speech, it would still be interesting and useful to quantify and to interpret vowel-speaker interactions in the formant space available from Bernard's original data. Toward this end, therefore, we undertook the preliminary task of restoring those Australian English data of nearly three decades ago, details and results of which are offered here for the "steady-state" nuclei of the monophthongs.

PARTICULARS OF BERNARD'S (1967) STUDY OF THE AUSTRALIAN ENGLISH VOWELS

Retaining the conceptual bases of, but not numerically rendering his complete spectrographic study of 1967, are indeed Bernard's 1970 and 1989 publications which have contributed (per-vowel) formant estimates averaged across two separate recordings as well as across speakers of each idiolectal group. As a consequence, there has hitherto been a lack of knowledge about inter-speaker distributions in the formant space estimated by Bernard. Our approach will then be to: (1) first recall particulars of Bernard's investigation of 1967 for the sake of completeness; and (2) present the results of our restoration of his multi-speaker data.

Speech material and speaker set

Bernard's study of Australian speech was based on the vocalic nuclei of /hVd/-syllables. The syllable set used comprises 11 nominal monophthongs as defined in "heed", "hid", "head", "had", "hard", "hud", "hod", "hoard", "hood", "who'd" & "herd", and 8 diphthongs as defined in "hade", "hide", "hoyed", "howd", "hoad", "heered", "haired" & "hoored". All syllables were recorded from flash cards by 170 adult male, Australian-born and normal speakers, once in stressed sentence-final position and once in isolated citation; and once in a sustained phonation in the case of the monophthongs. Apparently, only 100 speakers or so were able to produce adequate tokens in the sustained recording style.

Idiolectal classification

Bernard's 1967 doctoral thesis together with his 1970 and 1989 articles reinforce a point of view initially put forward by Mitchell and Delbridge, that "Australian English is phonologically one dialect only, but certain phonetic features, most notably, but not exclusively, the vocalic allophones, show ranges of continuous variation across the population". At one end of this continuum are perceived the so-called Broad pronunciations which embody what Bernard described as "local features" of Australian speech; at the other end there can be found Cultivated pronunciations which are the least coloured by local features. Between these two extremes, it has then become common to refer to the General pronunciations. This idiolectal classification of Australian speech was duly considered by Bernard who proceeded as follows.

First, every speaker was classified as either Broad, General or Cultivated from auditory impressions obtained by Bernard during an enrolment interview, which was typically initiated by prompting the speaker to talk about his most recent weekend activities. Following the interview, every speaker recorded the /hVd/-syllables of interest in the more controlled manner described earlier; and each vocalic nucleus was automatically given the same idiolectal label as that assigned globally to the speaker who produced it. In this approach, the speech material used was evidently produced spontaneously and almost continuously, and may therefore be assumed to have facilitated Bernard's auditory judgement of phonetic and prosodic gestures peculiar to individual speakers.

It is relevant to add that Bernard also approached this tripartite classification by restricting his auditory impressions to six (6) sounds (as in "heed", "who'd", "hade", "hide", "howd" and "hoad"), which Mitchell and Delbridge claimed to carry diagnostic cues of idiolectal differences in spoken Australian English. Although Bernard's data files do include his results of this approach, his main published data and our current restoration procedure are concerned only with the more global auditory impressions.

F-pattern (F1, F2 and F3) estimation in vowel nuclei

The vocalic nucleus of every /hVd/-syllable was segmented by visual inspection of a corresponding spectrogram, on which the centre frequencies of the first three formants were *manually* and *carefully* measured by Bernard at the onset, target and offset sections of the nucleus. In the case of a true monophthong (i.e., one which acoustically embodies the phonetic colour of the single vowel intended), only one target was defined in the most stationary section where F1 and F2 trajectories were both horizontal or where either of these trajectories reversed direction. By contrast, a two-target status was given to both diphthongal and diphthong-like (most notably, those in "heed" and "who'd") nuclei. However, in order to facilitate this preliminary stage of our study of Bernard's data, only the F-patterns he estimated at the second target were taken as representative of the monophthongs realised with diphthongal quality.

Table 1. Statistical data (in Hertz) on the F-patterns (averaged across contexts A & B) of the Australian English vowels spoken by 14 adult male speakers of the "BROAD" group.

	S[JB]: mean(std)			[JB] - S[JB]		
Vowel-ID	F1	F2	F3	F1	F2	F3
{heed}	329 (40)	2130 (111)	2620 (134)	**-4**	**75**	**105**
{hid}	364 (33)	2149 (154)	2655 (129)	**1**	**76**	**105**
{head}	466 (35)	1998 (169)	2609 (157)	**-11**	**62**	**86**
{had}	600 (68)	1858 (130)	2597 (137)	**25**	**37**	**33**
{hard}	728 (56)	1383 (74)	2492 (129)	**17**	**2**	**3**
{hud}	725 (95)	1395 (97)	2525 (132)	**15**	**20**	**0**
{hod}	646 (71)	1096 (68)	2480 (219)	**-16**	**-21**	**-40**
{hoard}	450 (39)	911 (77)	2459 (154)	**0**	**-46**	**-49**
{hood}	410 (32)	967 (76)	2409 (90)	**-5**	**-37**	**-24**
{who'd}	367 (50)	1627 (106)	2360 (97)	**-2**	**33**	**45**
{herd}	486 (29)	1517 (96)	2508 (105)	**--6**	**43**	**27**

Table 2. Statistical data (in Hertz) on the F-patterns (averaged across contexts A & B) of the Australian English vowels spoken by 11 adult male speakers of the "GENERAL" group.

	S[JB]: mean (std)			[JB] - S[JB]		
Vowel-ID	F1	F2	F3	F1	F2	F3
{heed}	286 (18)	2282 (143)	2738 (99)	**24**	**-22**	**32**
{hid}	359 (33)	2197 (124)	2686 (81)	**6**	**23**	**74**
{head}	457 (30)	2024 (90)	2613 (92)	**-2**	**36**	**37**
{had}	638 (74)	1830 (114)	2557 (116)	**-3**	**65**	**33**
{hard}	724 (57)	1344 (83)	2434 (115)	**16**	**16**	**31**
{hud}	737 (89)	1393 (83)	2402 (120)	**3**	**17**	**98**
{hod}	633 (48)	1052 (49)	2395 (118)	**-8**	**8**	**15**
{hoard}	441 (17)	817 (55)	2470 (183)	**-1**	**8**	**-70**
{hood}	400 (21)	880 (71)	2388 (152)	**0**	**30**	**2**
{who'd}	330 (43)	1558 (124)	2313 (153)	**20**	**52**	**52**
{herd}	482 (17)	1474 (66)	2539 (102)	**-7**	**41**	**-4**

Table 3. Statistical data (in Hertz) on the F-patterns (averaged across contexts A & B) of the Australian English vowels spoken by 11 adult male speakers of the "CULTIVATED" group.

	S[JB]: mean(std)			[JB] - S[JB]		
Vowel-ID	F1	F2	F3	F1	F2	F3
{heed}	291 (24)	2253 (104)	2875 (171)	**-1**	**67**	**-15**
{hid}	361 (36)	2114 (126)	2711 (127)	**9**	**76**	**14**
{head}	468 (35)	1976 (87)	2584 (104)	**-3**	**14**	**6**
{had}	648 (43)	1800 (85)	2581 (93)	**17**	**30**	**4**
{hard}	743 (43)	1290 (85)	2543 (115)	**-13**	**0**	**-28**
{hud}	749 (56)	1338 (84)	2560 (141)	**-9**	**22**	**-20**
{hod}	612 (52)	1038 (75)	2471 (125)	**28**	**12**	**-41**
{hoard}	447 (38)	831 (71)	2367 (212)	**-7**	**-11**	**-27**
{hood}	402 (36)	912 (107)	2344 (91)	**-7**	**-22**	**-54**
{who'd}	335 (38)	1556 (98)	2275 (140)	**0**	**-41**	**-5**
{herd}	484 (32)	1436 (94)	2509 (118)	**-9**	**24**	**-39**

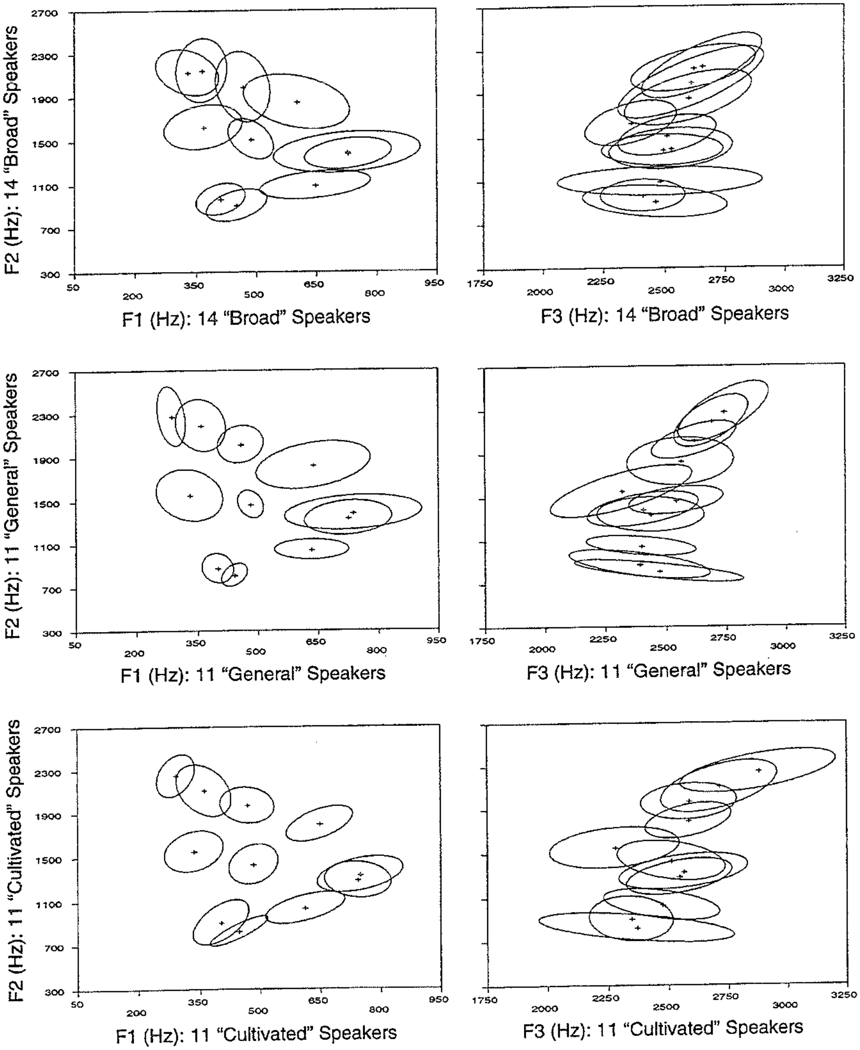

Figure 1. The *left and right graphs* illustrate, respectively, F1-F2 and F3-F2 steady-state distributions of the 11 Australian English monophthongal vowels placed in /hVd/ context. The *top, middle and bottom graphs* concern, respectively, our subset S[JB] of Broad (14 adult male), General (11 adult male) and Cultivated (11 adult male) speakers retained from Bernard's (1967) original data. Every vowel cluster is represented by a 2-sigma ellipse the centroid of which is denoted by a "+" sign. The multi-speaker data enclosed in each cluster represent averages across two separate recordings - namely context A (stressed sentence-final position) and context B (isolated citation).

RESTORATION OF BERNARD'S (1967) MULTI-SPEAKER FORMANT DATA

This task was evidently undertaken by first obtaining Bernard's original formant data, a copy of which was prepared by Dr Robert Mannell of Macquarie University. These data exist in 19 separate computer ASCII files, which clearly concern the 11 monophthongs and the 8 diphthongs referred to earlier. Therefore, each of these files contains Bernard's measurements for one of the 19 vocalic nuclei produced by all speakers in sentence context (context A), in isolated citation (context B) and in sustained phonation for the monophthongs (context C). However, only the 11 monophthongs recorded in contexts A & B were considered in our restoration procedure which consisted of: (1) decomposing the corresponding data files on a per-speaker and a per-context basis; and (2) then validating each speaker's data using certain criteria described below.

Search for complete F-patterns and "valid" speakers

Our preliminary examination of Bernard's data indicated that there are certain formants, F3 in particular (Bernard, 1967: 99-101, 641), which he apparently could not estimate with sufficient certainty for certain vowels spoken by a number of speakers, and which he wisely left as unspecified with zero entries. Such data together with the speakers involved were, however, not to be considered as "valid" for the purpose of our study. Consequently, in order to achieve our goal of restoring not only F1 and F2 but also F3, it became necessary to search for a *complete* subset (i.e., non-zero F-patterns at target) of Bernard's original data, with little hope of being able to retain all 170 speakers. Indeed, our search results clearly point to the F3 of the vowel nuclei of "hoard" and "hood" (consistently in both contexts A & B) as the dominant culprits in limiting the number of our "valid" speakers to 14 Broad, 11 General and 11 Cultivated: 36 in all for the subset retained. It is worth noting that nearly all 170 speakers would form a "valid" set if one were interested only in the F1-F2 space.

Some descriptive statistics of the subset S[JB]

Similarly to Bernard's findings, the F-patterns retained through our validation search do not numerically differ very much from context A to context B, and thus were averaged to form the subset S[JB] used here for further observations.

The considerable reduction of speaker dimensionality from 170 to 36, for example, raises the question of possible displacement of the new means relative to those published in 1970 by Bernard [JB]. Per-vowel, signed differences were then calculated between the two sets of means ([JB]-S[JB]), and entered in the three rightmost columns of Tables 1 (Broad group), 2 (General group) & 3 (Cultivated group). The differences appear to be quite tolerable for the three idiolectal groups, with a common trend of larger discrepancies for F2 and F3 than for F1.

The per-vowel means and standard deviations (std) were also generated for the S[JB] data, and appear in the three central columns (for F1, F2 & F3 from left to right) of Tables 1, 2 & 3. Perhaps the most striking observation is that deviations are relatively smaller for F1, and that they tend to be the largest for F2 of front vowels and for F3 of the back and most of the front vowels - a result consistent with the growing evidence in the literature which lends relatively more status to these two formants for speaker characterisation.

Vowel and speaker distribution in formant space

The profile of standard deviations reported in Tables 1, 2 & 3, indeed, suggests that the overall inter-speaker spread is consistently larger in F3-F2 space. A more revealing illustration of this is offered in the right-hand side graphs of Figure 1, which highlight that inter-speaker distances are dominant in F3-F2 space while distances between vowel classes appear therein to be quite small. In sharp contrast, examination of the left-hand side graphs of Figure 1 leaves very little doubt that our 11 vowel classes are far better separated from one another in F1-F2 space - the predominantly phonetic space as it has become known since Potter and Peterson's (1948) study. It seems clear, however, that even in that space complete vowel discrimination would be hampered by speaker and idiolectal variations.

The types of problems illustrated above, which arise from phonetic-speaker interactions, are persistently challenging in the search for robust methods of speech or speaker recognition. Further advances will depend partly on gaining more fundamental insights into the determinants of acoustic-phonetic variability. In this regard, Bernard's (multi-speaker) formant data are still relevant and potentially enlightening.

CONCLUDING COMMENTS

The results of our endeavour to restore Bernard's multi-speaker data on spoken Australian English vowels are not at all displeasing. The 36-male dataset (14 Broad, 11 General and 11 Cultivated) retained from Bernard's pioneering spectrographic analyses, contains a *complete* set of the formant-frequencies F1, F2 and F3 which were manually estimated in the "steady-state" section of 11 monophthongs produced in /hVd/ settings, once in isolated and once in sentential context, yielding a total of 792 Australian English F-patterns. This is quite comparable to Peterson and Barney's well-known 33-men dataset, which includes spectrograhic estimates of 660 F-patterns also obtained in the "steady-state" section of 10 spoken American English vowels. It is important to further note that the 36-speaker dataset is relatively more comprehensive than its American English counterpart, which does not explicitly contain dialect-related information.

On the one hand, this very attempt to restore thirty-year old estimates of vowel formants is perhaps one further indication that such data are still difficult to gather, owing mainly to the limited capabilities of currently available formant-tracking techniques. On the other hand, the account presented here of Bernard's measurements, does unveil an invaluable contribution to that body of *hard-to-get*, multi-speaker data which still hold a place of theoretical importance in the search for fundamental insights into speech production and recognition.

ACKNOWLEDGEMENTS

I am grateful to Dr Robert Mannell for kindly providing me with a copy of Bernard's formant data files. I thank Mr Kishor Kumar for his computer programming assistance.

REFERENCES

Bernard, J.R.L. (1967) "Some measurements of some sounds of Australian English", Doctoral Thesis, The University of Sydney.

Bernard, J.R.L. (1970) "Toward the acoustic specification of Australian English", Zeitschrift fur phonetik, Sprachwissenschaft und Kommunikationsforschung, vol. 23, 113-128.

Bernard, J.R.L. (1989) "Quantitative aspects of the sounds of Australian English", *Australian English: The language of a new society*", Collins & Blair (eds), University of Queensland Press, Queensland, Ch. 16, 187-204.

Burgess, N. (1969) "A spectrographic investigation of some diphthongal phonemes in Australian English", Language and Speech, vol. 12, 238-246.

Furui, S. & Akagi, M. (1985) "Perception of voice individuality and physical correlates", J. Acoust. Soc. Japan (English reprint), H-85-18.

Kitamura, T. & Akagi, M. (1994) "Speaker individualities in speech spectral envelopes", Proc. 3rd Int. Conf. on Spoken Language Processing, Yokohama, Japan, 1183-1186.

Kuwabara, H. & Takagi, T. (1991) "Acoustical parameters of voice individuality and voice-quality control by analysis-synthesis method", Speech Communication, vol. 10, 491-495.

Lewis, D. & Tuthill, C. (1940) "Resonant frequencies and damping constants of resonators involved in the production of sustained vowels "O" and "Ah" ", J. Acoust. Soc. Am., vol. 11, 451-456.

Mella, O. (1994) "Extraction of formants of oral vowels and critical analysis for speaker characterization", ESCA Workshop on Automatic Speaker Recog., Identif. & Verif., Suisse, 193-196.

Mitchell, A.G. & Delbridge, A. (1965) *The Speech of Australian Adolescents*, Angus & Robertson, Sydney.

Mokhtari, P. & Clermont, F. (1994) "Contributions of selected spectral regions to vowel classification accuracy", Proc. 3rd Int. Conf. on Spoken Language Processing, Yokohama, Japan, 1923-1926.

Mokhtari, P. & Clermont, F. (1996) "A methodology for investigating vowel-speaker interactions in the acoustic-phonetic domain", ASSTA-Award winning paper accepted for presentation at the 6th Australian Int. Conf. on Speech Science and Technology, Adelaide, Australia.

Peterson, G.E. & Barney, H.L. (1952) "Control methods used in a study of the vowels", J. Acoust. Soc. Am., vol. 24, 175-184.

Potter, R.K. & Peterson, G.E. (1948) "The representation of vowels and their movements", J. Acoust. Soc. Am., vol. 20, 528-535.

Sampath, S. (1995) "Dialect specific features of Australian English diphthongs in spontaneous speech", Proc. 4th European Conf. on Speech Comm. & Tech., Madrid, Spain, vol. 3, 1929-1932.

Stevens, K.N. (1971) "Sources of inter- and intra-speaker variability in the acoustical properties of speech sounds", Proc. 7th Int. Congr. of Phonetic Sciences, 206-232.