VOWEL SHIFT AND LONG-TERM AVERAGE SPECTRA IN THE SURVEY OF VANCOUVER ENGLISH

JOHN H. ESLING

Department of Linguistics
University of Victoria
Victoria, B.C. V8W 2Y2 Canada

ABSTRACT

To investigate the relationship between long-term (voice setting) and short-term (segmental) components of accent in social varieties of Vancouver English, formant analysis of digitally sampled vowels and long-term average spectral (LTAS) analysis from context-controlled readings are compared. Four contrasting patterns of vowel formant frequency shift result for the four survey groups divided by socio-economic index. LTAS peaks for UWC and UMC subjects are significantly differentiated, paralleling consistent vowel system differences between these groups. Comparisons with articulatorily performed models permit tentative identification of supralaryngeal settings corresponding to each acoustic pattern. An explanation is offered of the potential effect of long-term configuration on the measurement of individual vowel formants.

SAMPLING AND SPEECH ANALYSIS

The objective of this research is to determine whether socio-economic divisions of an urban linguistic community can be distinguished on the basis of voice setting shifts as well as in terms of differences in individual vowels. Sociolinguistic data for acoustic analysis are drawn from the Survey of Vancouver English carried out by Gregg et al. at the University of British Columbia [1] and archived at the University of Victoria, which includes tape-recorded interviews with 240 native speakers of Canadian English. Subjects chosen for investigation are 32 female and 32 male natives of Greater Vancouver, from the youngest of the three age divisions (16-35) in the survey. Female and male subjects are divided into four socio-economic groups of 8 subjects each on the basis of social index scores established in the original survey using the Blishen & McRoberts [2] occupation scale and other social indicators. Group 1 represents low social index scores (Lower Working Class), and group 4 represents high social index scores (Upper Middle Class).

To compare vowel clusters across the four groups, vocalic nuclei are computed for two tokens of each of ten vowel phonemes for each speaker, from identical environments of the same text in reading style. Using ILS speech processing algorithms to determine formant frequencies, speech samples digitized at 10K samples per second are analyzed using 12-pole autoregressive linear predictive coding [3]. The analysis results in 12 reflection coefficients (K's) per frame (200 points/frame; 50 frames/sec). The K's are converted to filter coefficients (A's) to represent the vocal tract's filtering effects, and the filter response of the A's in each frame is calculated and displayed in a spectral array showing up to five resonant peaks (formants) in the 0-5000Hz range. The peaks' centre frequencies are calculated based on a -3dB shoulder and listed. Target vowels are isolated from remaining speech data auditorily, and mean F1,F2 frequencies are calculated and filed by group for statistical processing and plotting. Follow-up vowel measurements and data collection are now performed more expediently on the Micro Speech Lab package developed in the Centre for Speech Technology Research at the University of Victoria on the IBM-PC microcomputer.

For LTAS analysis, a 45sec sample of continuous speech for each speaker, from the same text used for vowel measurements, is digitized with a PDP-11 time-series data-capturing program. One long-term spectrum is computed for each voice, using a main-frame program accepting only voiced frames while excluding voiceless and low-energy frames. Power spectra of non-overlapping 20msec windows at 50Hz resolution and pre-emphasis factor 1 are integrated to obtain final LTAS.

STATISTICAL ANALYSIS

Statistics are performed on log-mean normalized F1,F2 data for approximately 600 female and 600 male vowels, respectively [4]. To compute distance between group vowel clusters, principal component analysis and canonical discriminant analysis are applied to the four female and four male groups, with the Mahalanobis distance calculated between each group. This yields a probability relating collections of vowels to each other, first as complete vocalic inventories by social group, then as individual vowel phoneme clusters by group.

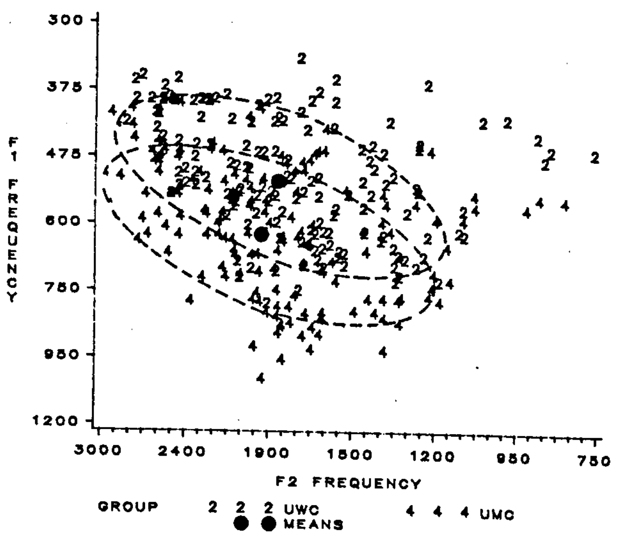
A generalized squared distance measure is used to classify F1,F2 coordinates, as unknown test values, into one of the four social groups as known reference cells. Vocalic inventories of the four male groups are also compared with equivalent vowels from texts performed by the author as models representing contrasting articulatory settings. In this case, test values are assigned to known reference models to yield numbers of vowels from each group that associate most closely with each model [5].

In LTAS evaluation, the same procedure is used to compute probabilities and distance relating spectra in the four female and four male groups, although statistics operate on unnormalized data. Male LTAS are compared with LTAS of the articulatory models using generalized squared distance to identify clustering patterns and to relate LTAS shift to vowel formant shift.

VOWEL FORMANT ANALYSIS

For female subjects, the complete vocalic inventories of all four social groups are significantly differentiated ($p<0.001$), and a majority of individually compared vowel phoneme clusters are also separated across socio-economic group. The acoustic characteristics of each group's vowels match the four corners of the two-dimensional vowel space: Group 1 (high F1,low F2); Group 2 (low F1,low F2); Group 3 (low F1,high F2); Group 4 (high F1,high F2). The most coherent and best differentiated groups are groups 2 (Upper Working Class) and 4 (Upper Middle Class), illustrated in figure 1. Linguistic contexts are identical; only speakers vary by group affiliation.

FIGURE 1.
SURVEY OF VANCOUVER ENGLISH, FEMALE UWC AND UMC. IDENTICAL VOWELS OF SOCIO-ECONOMIC GROUPS 2 AND 4.



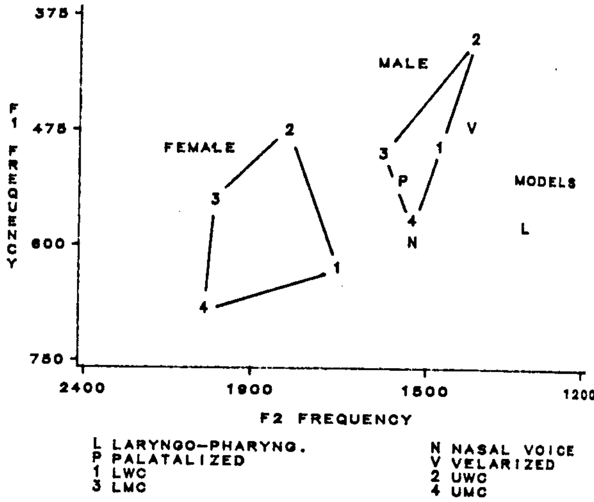
Male vowel cluster values follow the pattern of female vowels except that differentiation between groups 1 and 3 is marginal for speaker-normalized vowels, and not significant using unnormalized data. All other pairings show significant separation ($p<0.001$). As with female groups, male UWC is furthest separated from other male groups particularly UMC. Figure 2 illustrates normalized means of the four socio-economic groups by sex, and also vocalic means of four comparable model settings.

In the analysis of individual vowel phoneme clusters by group, 77% of all possible pairings for the ten vowels are significantly differentiated for female speakers across the four survey groups ($p<0.05$), and 43% of all pairings remain separated at the $p<0.001$ level. Social groups 2 and 4 are successfully differentiated for all ten vowels individually ($p<0.01$). For groups 1 and 3, which are most difficult to differentiate, only four of the ten vowels show no separation. This supports the distinctions reported for the complete vowel systems of these groups. The rank order of most significantly separated vowels across groups for female speakers, /ʊ/ /e/ /ɛ/ /ʌ/ /i/ /o/ /æ/ /u/ /ɪ/ /ɒ/, suggests no obvious principles, except that mid, front to central vowels tend to be better differentiated than peripheral, especially open vowels.

Individual vowels for male speakers demonstrate less separation than female speakers' vowels across the four groups. At the $p<0.05$ level of significance, 62% of all possible pairings for male vowels are differentiated, while only 27% separate at the $p<0.001$ level. The analysis of individual vowels positively separates male groups 2 and 4, where all vowels differentiate significantly ($p<0.001$) except /i/, but is not successful in separating the individual vowels of groups 1 and 3. The rank order of socially best differentiated vowels for male speakers is: /ɛ/ /ɪ/ /ʌ/ /ɒ/ /æ/ /u/ /ʊ/ /e/ /o/ /i/. The Spearman rank order correlation coefficient relating male and female rank orders (rho=-.24) indicates that the two lists do not correlate, suggesting that those vowels which function as salient social markers for female speakers are not the same vowels that function as principal social markers for male speakers in the same social classes.

One possible interpretation of the male order is that /i/ functions as a pivotal vowel, virtually identical in all groups, and that peripheral tense vowels /e/ and /o/ remain more or less the same across groups, while the majority of shifting occurs on open or mid-open vowels. Greatest differentiation appears in the area of /ɪ/ /ɛ/ /ʌ/ /æ/ /ɒ/, where a decrease in F1,F2 accompanies raising and backing for group 2, and an increase in F1,F2 accompanies fronting with nasalization for group 4.

FIGURE 2.
FEMALE AND MALE NORMALIZED GROUP VOWEL MEANS, VOCALIC MEANS OF FOUR VOICE SETTING MODELS.



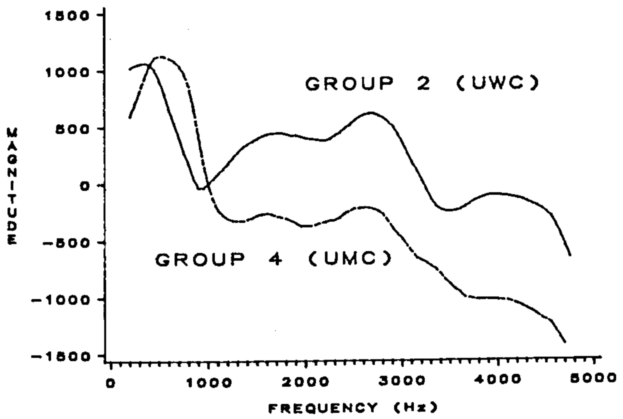
LONG-TERM AVERAGE SPECTRAL ANALYSIS

For LTAS analysis, 45-60sec of the 64 subjects' voices are low-pass filtered at 5KHz and digitized at 10K samples/sec with high-shaping to accentuate frequency information and remove DC. Digitized data are processed in 200 sample point frames through a Hamming window and FFT routine to obtain 20msec power spectral arrays. After unvoiced and silent frames are removed, a swept filter adjusted according to expected harmonic spacing produces smoothed spectra accumulated in a single array to represent the average vocal tract response of the utterance.

For articulatory identification, LTAS of three 40sec phonetic texts performed by the author using controlled voice settings described by Laver [6] and Esling [7] are analyzed: close rounding (CLR), close jaw (CLJ), dentalization (DEN), retroflexion (RET), palatalization (PAL), uvularization (UVU), velarization (VEL), laryngo-pharyngalization (LAR), nasalization (NAS), faucal constriction (FAU), raised larynx (RLX) and lowered larynx (LLX). Root-mean-squared distance measures indicate that each text resembles more closely other texts with the same voice setting than it does identical texts with different settings. Speaker recognition research corroborates that samples of this length are relatively text-independent [8].

The first two formants of four settings, LAR, VEL, PAL, NAS, parallel F1,F2 plots of survey data (see figure 2). The first two dominant LTAS peaks (P1,P2) of these models also correspond to F1,F2 in their relative acoustic orientation, but with P1,P2 systematically lower in frequency than F1,F2. Superimposing laryngo-pharyngalization on a given text increases P1 and decreases P2, which conforms with acoustic predictions for extreme tongue retraction [6]; velarization produces an approximation of P1 and P2 as for an [u]-quality vowel; palatalization results in a systematic shift in mean spectral peaks as for an [i]-quality vowel; and a nasal setting results in higher-frequency P1, with attenuation in the magnitude of P1 relative to P2. In evaluating LTAS data for Vancouver survey groups, it is expected that group 1 will demonstrate high P1,low P2; that group 2 will demonstrate low P1,P2; that group 3 will demonstrate low P1,high P2; and that group 4 will demonstrate high P1,P2. The relative influence of each of the first four LTAS peaks in distinguishing the social divisions of the survey will also be determined.

FIGURE 3.
LTAS OF FEMALE SOCIAL GROUPS 2(UWC) AND 4(UMC).



For female groups, LTAS data significantly differentiate social group 1 from group 2 and group 2 from group 4 ($p<0.01$) as in figure 3, while other relationships show no significant separation. Spectra are set to zero magnitude at 1000Hz for comparability and to minimize the effect of amplitude variation. Female LTAS data corroborate socio-economic distributions of vowel formant data in that groups 2 and 4 are separated by both measures. Due to the presence of voiced obstruents in LTAS, frequencies are predictably lower than for vowel nuclei. Relative F1,F2 orientations are preserved primarily in P1 values and not in P2, as much of the difference between groups is therefore present in third and fourth LTAS peaks.

Table 1. Female vowel formant and LTAS means.

	F1 , F2 (Hz)	P1 , P2 (Hz)
Group 1(LWC):	631 , 1702	450 , 1600
Group 2(UWC):	477 , 1813	350 , 1725
Group 3(LMC):	552 , 2006	400 , 1600
Group 4(UMC):	683 , 2039	550 , 1600

Male LTAS results are also successful in significantly differentiating group 2 from group 4 and group 3 from group 4 ($p<0.05$). Other relationships again are not significant. The relationship between F1,F2 values and LTAS P1,P2 values is clearer for male groups than for female groups. Both F1,F2 and P1,P2 for male group 2 are low, resembling the predicted pattern of velarization, while F1,F2 and P1,P2 for group 4 increase, coinciding with the shift predicted for nasalization. P1,P2 are systematically lower than F1,F2, confirming that LTAS data include voiced speech information which has the effect of lowering average frequencies.

INTERPRETATION OF RESULTS

An articulatory interpretation of the acoustic differentiation of vowels across the social scale of Vancouver English is proposed which associates LWC vowel clusters with tongue backing and lowering (laryngo-pharyngalization); UWC with tongue backing and raising (velarization); LMC with tongue fronting and raising (palatalization); and UMC with tongue fronting and nasal voice setting. To quantify these associations, male survey data are compared with equivalent vowel systems of four articulatorily modelled settings which are included in the male normalization routine. The generalized squared distance algorithm takes the four models as reference cells and forces tokens from survey data into one of the four cells. Internally, there is considerable misclassification of vowel tokens among the four settings, and the majority of survey values cluster with the velarized model. However, classification of survey data differentiates significantly in the case of groups 2 (UWC) and 4 (UMC) and the VEL and NAS models as tabulated below.

Table 2. Assignments of male vowels by group to model setting vowel sets (rounded %).

	LAR	VEL	PAL	NAS	n
Group 1(LWC):	13%	68%	14%	5%	139
Group 2(UWC):	3%	97%	0%	0%	145
Group 3(LMC):	14%	67%	12%	8%	153
Group 4(UMC):	19%	56%	10%	15%	145
Totals:	12%	72%	9%	7%	582

These distributions reflect the same articulatory pattern as female vowel clusters. Individual vowel phonemes classify primarily into VEL from group 2, and into NAS from group 4. Chi-squared tests indicate that there is significant evidence for an association between groups 2 and 4 and the four reference models LAR, VEL, PAL, NAS (3 d.f., $p<0.001$) and, furthermore, that the two groups are significantly differentiated on the basis of assignment into VEL, NAS (1 d.f., $p<0.001$). Broader interpretations of these results depend on variables such as performance conditions of the models and limitations of using only two formants. Nevertheless, they permit identification of the relative susceptibility of vowels to the shift from UWC to UMC quality, reflected in the acoustic shift from low to high F1,F2 values.

LTAS data support conclusions reached on vowel formant evidence. Tukey's test for variable effect is applied to the four models, LAR, VEL, PAL, NAS, to assess the relative influence of each LTAS peak. The result indicates that P1 is a better predictor of VEL or NAS than is P2. P3 is also a successful variable in separating VEL and NAS settings, and in separating fronting from backing. P4 does not distinguish PAL from VEL or NAS, but does separate it from LAR, as does P2. This suggests that P3 adds information

to P1, and that P4 adds to P2, when LTAS data are used in addition to F1,F2 to distinguish voices.
Statistical comparisons of male LTAS data with the 12 models indicate that the models as a set are significantly differentiated from the four survey groups ($p<0.05$). The generalized squared distance function indicates high internal coherence for each survey group, and yields similar associations to those previously discovered by vowel formant analysis, namely the association of tongue-retracted settings UVU, VEL with groups 1 and 2 (LWC/UWC) and of NAS, PAL with group 4 (UMC), shown in table 3.

Table 3. Distance between voice setting models and male Vancouver social groups in %.

	1(LWC)	2(UWC)	3(LMC)	4(UMC)
UVU	0.50	0.38	0.02	0.09
VEL	0.51	0.23	0.00	0.25
LAR	0.05	0.06	0.83	0.07
LLX	0.09	0.02	0.85	0.05
FAU	0.03	0.02	0.95	0.00
DEN	0.22	0.17	0.32	0.29
CLR	0.31	0.16	0.02	0.51
CLJ	0.32	0.09	0.01	0.58
LLX	0.09	0.19	0.12	0.59
RET	0.20	0.10	0.02	0.68
PAL	0.17	0.06	0.01	0.77
NAS	0.02	0.01	0.00	0.97

Bearing in mind the significant separation of groups 2 and 4, that groups 1 and 3 are not distinguished except for certain vowels, and that group 3 LTAS are more coherent than group 1 LTAS, assignments to group 3 (e.g., LAR) must be treated circumspectly. Assignment of VEL and UVU to both groups 1 and 2, on the other hand, provides supporting evidence to the vowel formant procedure that these groups occupy a different acoustic space from group 4 (if not from group 3) with its closer association to NAS and PAL. Despite the single-speaker limitations of the performed model approach, the associations suggested here are a positive indication that sociolinguistically obtained dialect survey groups can be analyzed, differentiated and tentatively classified using both vowel formant analysis and LTAS analysis techniques.

EFFECTS ON FORMANT MEASUREMENT

There is evidence in this study that long-term settings may influence formant frequency measurement, contributing to why vocalic data values are often difficult to measure. Monsen & Engebretson [9], comparing spectrographic with linear prediction techniques of formant analysis, find that "for fundamental frequencies between 100 and 300Hz, both methods are accurate to within approximately ±60Hz for both first and second formants." They also observe that formant frequencies can be obscured by masking from the fundamental or by broadening of bandwidths.

> It may be easier or harder to accurately recover the resonances of the vocal tract in the vowel sound wave depending on objective factors such as the fundamental frequency, the degree of nasalization of the vowel, or the position of the articulators.

The ILS peak-picking routine used here is observed to encounter masking problems of just this sort. Group 1 vowels produce greatest loss of second formant, resulting in a smaller number of tokens that are acceptable for inclusion, and (perhaps not incidentally) in wider deviation of the tokens that remain. Group 2 is the easiest group to measure, with all formant peaks and bandwidths clearly distinguishable, and has correspondingly the most coherent set of formant values. Group 3 is also not difficult to measure, but group 4 begins to demonstrate the appearance of an intermediate peak and widening bandwidths in all vowels for the largest number of speakers both male and female. This secondary, usually higher amplitude peak overlaps in bandwidth with peak 1, and has therefore been averaged into the computation of F1 since it is distinctly not associated with F2. This phenomenon occurs only rarely in other groups and when it does the voice demonstrates pronounced nasality. It seems likely, therefore, that a generalized low back position of the articulators in group 1, evident in the F1,F2 values of retained vowels, causes a decreasing F2 peak to merge with an increasing F1 peak for many tokens. The fronted and nasalized setting of group 4, implied by the damped but increased values of F1 due to the combined calculation, and the slightly higher values of F2, would not be apparent if these somewhat spectrally confusing tokens had to be eliminated. In this way, the results of this study help to isolate those contributions of vocal tract resonance that are of longer-term duration than individual vowels, and also help to identify how contrasting articulatory configurations affect otherwise identical vowels.

REFERENCES

[1] R. Gregg et al., An urban dialect survey of the English spoken in Vancouver, in H. Warkentyne (Ed.), Papers from the Fourth International Conference on Methods in Dialectology (pp. 41-65), University of Victoria, 1981.

[2] B. Blishen, H. McRoberts, A revised socioeconomic index for occupations in Canada, Canadian Review of Sociology and Anthropology, 13, 71-79, 1976.

[3] J. Markel, A. Gray, Linear prediction of speech, Springer-Verlag, 1976.

[4] D. Hindle, Approaches to vowel normalization in the study of natural speech, in D. Sankoff (Ed.), Linguistic variation: Models and methods (pp. 161-171), Academic Press, 1978.

[5] J. Esling, C. Dickson, Acoustical procedures for articulatory setting analysis in accent, in H. Warkentyne (Ed.), Papers from the Fifth International Conference on Methods in Dialectology (pp. 155-170), University of Victoria, 1985.

[6] J. Laver, The phonetic description of voice quality, Cambridge University Press, 1980.

[7] J. Esling, The identification of features of voice quality in social groups, JIPA, 8, 18-23, 1978.

[8] C. Dickson, An investigation of theories and parameters pertaining to speaker recognition, University of Victoria, 1980.

[9] R. Monsen, M. Engebretson, The accuracy of formant frequency measurements: A comparison of spectrographic analysis and linear prediction, J. Speech and Hearing Research, 26, 89-97, 1983.

Acknowledgements. Thanks are due to Prof. R.J. Gregg, Dr. S.K. Wong, and J. Clayards. The cooperation of B. Craig Dickson and the Centre for Speech Technology Research, and grants from the President's Committee on Research, University of Victoria, are gratefully acknowledged. The present investigation was supported by grant #410-85-0431 from the Social Sciences and Humanities Research Council of Canada.