ACOUSTIC PHONETICS IN SOCIOLINGUISTICS AND DIALECTOLOGY: THE VARIATION OF ENGLISH VOWELS IN SPONTANEOUS SPEECH

Reijo Aulanko and Terttu Nevalainen
Department of Phonetics and Department of English, University of Helsinki, Finland

ABSTRACT

The paper examines the potential of acoustic phonetics in dialectology. Using Wells's standard lexical sets as our frame of reference, we discuss the methodological issues that arise in the reconstruction of a Somerset speaker's vowel system from spontaneous speech. They include vowel-internal variation, lexical type and token variation, the effects of the immediate phonemic context, and linear as opposed to logarithmic scaling of the results.

ACOUSTIC DIALECTOLOGY?

Over the last decades, acoustic measurements have been gaining ground in urban sociolinguistics. They are typically used in vowel studies to show individual speakers' vowel systems within a framework that readily lends itself to further interindividual comparisons and generalizations [1]. We think that this methodology could also be profitably applied to more traditional dialectology.

The focus of our study [2] is methodological: we explore the various options that acoustic analysis opens up to a sociolinguist and dialectologist, as well as some of the problems connected with the approach. In this paper we show how the results of the vowel analysis may differ depending on a) the exact temporal location of the measurement point, b) the number of measurements per lexeme category, and c) the phonemic context of the sound studied.

DATA: EAST SOMERSET VOWELS

The material of our study consists of the stressed monophthongs of a rural West Country speaker. Our topic and data were first suggested to us by the late Professor Ossi Ihalainen, who made recordings of Somerset folk speech in the 1970s. Our study is based on a 45-minute spontaneous interview with a 69-year-old farmer from Wedmore, East Somerset, in his home in 1976. Although its quality is not ideal, the tape recording is fully intelligible and thus lends itself to acoustic analysis.

We identified the possible vowel contrasts (not necessarily all phonemic) in our data on the basis of Wells's [3] standard lexical sets. They consist of twenty-four matching pairs for vowels in strong syllables in standard British (RP) and American English (GA). More comprehensive, these sets seemed preferable to using either RP, 'Middle English', or GA alone as a baseline for the analysis of dialectal speech. A valuable earlier source of comparison was provided by the Wedmore data from 1956 in the *Survey of English Dialects* [4].

The lexical sets examined in this study are: 1. KIT, 2. DRESS, 3. TRAP, 4. LOT, 5. STRUT, 6. FOOT, 7. BATH, 8. CLOTH, 9. NURSE, 10. FLEECE, 12. PALM, 13. THOUGHT, 15. GOOSE, 21. START, 22. NORTH, and 23. FORCE.

ACOUSTIC ANALYSIS

The data were acoustically analysed with the Intelligent Speech Analyser (ISA) system designed by R. Toivonen [5], implemented in a Macintosh Quadra 700 computer. The ISA is an interactive speech analysis system which enables the analyst to make measurements from several simultaneous displays (FFT and LPC spectra, sound spectrograms etc.) and to monitor the digitized signal auditorily. The system also makes it possible to show the measured values graphically in different ways; e.g. the frequency scales in formant (F1/F2) charts can be linear or psychoacoustic (Bark scales).

In this study each vowel target was first located by means of a wide-band spectrogram and an intensity curve and by auditory monitoring of the signal. An LPC-based automatic formant analysis was used to suggest values for F1 and F2, which were accepted only if they conformed to the visual appearance of the formant structure in the spectrogram and the FFT spectrum. Usually only one pair of F1/F2 values was recorded for each vowel token, i.e. its possible diphthongization was not traced in this analysis (except for Fig. 1 below).

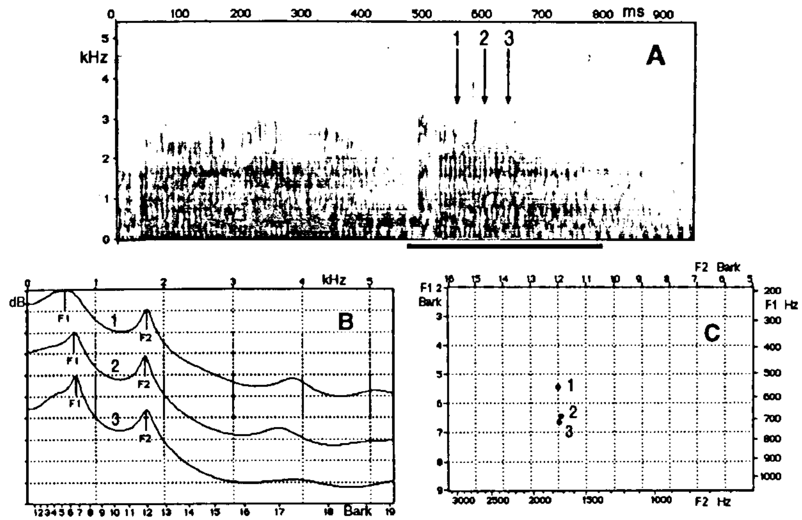

Figure 1. The effect of vowel-internal variation. The spectrogram in **A** *shows the target word in context "... cider did get, ...", the three LPC spectra in* **B** *were calculated from the time points indicated in* **A**, *and the F1/F2 plot in* **C** *shows the changes in the formant values of the target vowel.*

RESULTS

Vowel-internal variation

As only one LPC spectrum is normally determined for each monophthong, we may ask how stable the vowel is acoustically. To study its internal stability, several measurements can be made from one vowel (but care should be taken to exclude formant transitions, usually < 50 ms, due to neighbouring consonants). Fig. 1 shows changes in F1/F2 values caused by making the measurements early or late in the vowel, in addition to the normal mid-point measurement. As can be seen, in this case the effect is more pronounced in F1 than in F2 (160 Hz v. 30 Hz), indicating a slight lowering during the vowel in the word *get*. Cases like this seem to support the standard practice of using the vowel mid-point (2 in Fig. 1) as a value representative of the whole monophthong.

For the rest of the measurements, a total of 511 instances, we located the steady state of the vowel, a point where its quality would stay reasonably invariant for at least 30 ms in the middle region of the target vowel.

Lexeme token variation

Vowels may also differ in different tokens of the same word. Each dot in Fig. 2 shows the F1/F2 values from the middle points of four instances of the target word *get* pronounced in varying sentence stress conditions. The range of variation is 120 Hz for F1 and 70 Hz F2. The variation between tokens is thus approximately as large as the vowel-internal variation discussed above.

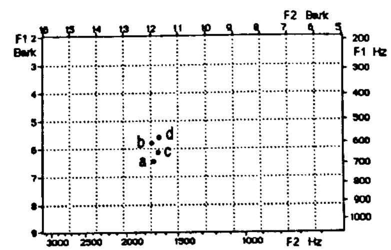

Figure 2. The distribution of four measurements representing different tokens of the word get *from the contexts*
(a) ...the cider did get,...;
(b) ...you could get 'em for...;
(c) ...you could get 'em,...;
(d) ...you could get 'em to what...

Lexeme type variation

Lexeme types may also give rise to variation. Fig. 3 shows the range of 35 F1/F2 values measured from the middle of the target vowel in the following lexemes which all represent the lexical set DRESS: *bed*, *dead* (2), *elm* (3), *fell* (2), *get* (3), *head* (2), *help*, *left* (2), *let* (3), *neck* (3), *net* (3), *plenty* (2), *press* (3), *sell* (3), and *went* (2). The number of lexeme tokens is shown in brackets.

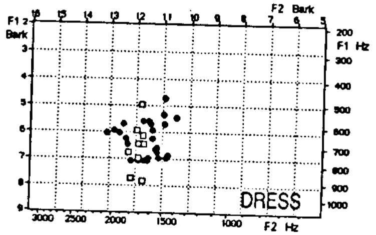

Figure 3. The distribution of 35 individual measurements representing 15 different word types from the DRESS *set. The 9 instances of the context* [C(C)_t#] *are shown by white squares, the others by black circles. (Adapted from* [2].*)*

As might be expected, the phonemic context of the vowel has an effect on its acoustic realization even within one and the same lexical set. In Fig. 3 most of the centralized tokens of the DRESS vowel (lowest F2) occur before /l/ in *elm*, *fell*, *help* and *sell*, and following /w/ in *went*. Fronted tokens (highest F2) are found in *head*, where the vowel lengthens before a voiced stop.

One way to minimize the effects of lexeme type variation is to introduce a fixed frame for all vowels. We tested this common practice with our material by selecting a context shared by the majority of our lexical sets, [C(C)_t#]. The white squares in Fig. 3 indicate the dispersion of values in this fixed context, while the black circles show all the other instances. As can be seen, the fixed context clearly reduces F2 variation in this case, whereas F1 variation is not reduced at all.

Another, more laborious approach is to take the vowel of the lexical set to be the mean value of all its lexical tokens. Fig. 4 shows that in the DRESS set there is no great difference between the means of the fixed context and the whole material.

Mean values in fixed contexts

When reconstructing vowel systems, contextual variation is usually minimized by calculating mean values for the data. Fig. 4A shows the average values obtained for eleven of our sixteen sets (i.e. 1, 2, 3, 4, 5, 6, 10, 13, 15, 21, and 22) in the fixed context [C(C)_t#]. For the sake of comparison, Fig. 4B gives the mean values for all the lexeme types representing each of the sixteen sets.

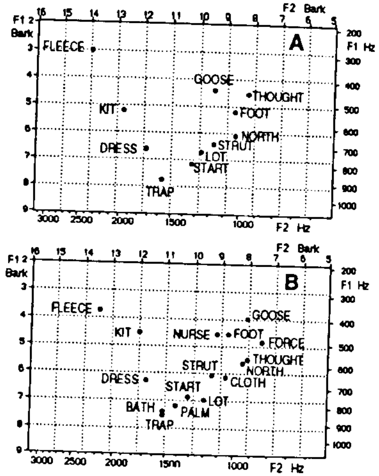

Figure 4. Average formant values measured from [C(C)_t#] *contexts (*A*) compared with the corresponding values from all available phonetic contexts (*B*). (*B *adapted from* [2].*)*

Fig. 4 indicates that the fixed frame yields a more regular distribution for the front vowels in FLEECE, KIT, DRESS and TRAP by increasing the distance between the close front vowel /i/ in FLEECE and the half-close /ɪ/ in KIT. The START, LOT and STRUT vowels are similarly kept separate in both figures, but the fixed frame suggests a more central realization for LOT.

The GOOSE and FOOT vowels are also distinguished in the two figures, but the value for GOOSE is more central in 4A. Finally, the main distributional difference between the two figures can be seen in the values for THOUGHT and NORTH. 4A clearly separates them, while 4B suggests little or no difference. The reason for this is that, as opposed to 26 cases in 4B, our material only contains two tokens for THOUGHT in the fixed context, *caught* and *thought*. Their mean in 4A also conceals a 100 Hz difference in their F1 values. Hence the distinction between the two vowels suggested by 4A may be more apparent than real.

Visual scaling of the results

How salient are the vowel differences in Fig. 4 perceptually? Various scales have been proposed for presenting vowel formant values [1, 6].

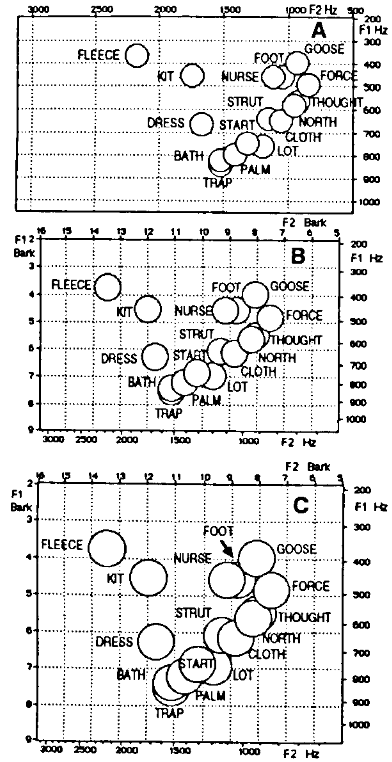

Figure 5. Average F1/F2 values for all lexical sets. **A**: *linear frequency scale (with an arbitrary circle size);* **B**: *Bark scale (with a one-Bark circle size; adapted from* [2]*);* **C**: *Bark scale with the F1 scale and the circle size expanded by 35% (see* [6]*).*

In this study we have used the Bark scale, which modifies the frequency scale according to the critical bands of human hearing. Bark-sized circles marking F1/F2 values visually simulate the psychoacoustic distances that separate any given pair of vowel qualities that are clearly distinguishable for the listener.

Fig. 5 shows the mean F1/F2 values of our lexical sets using three different frequency scales: linear, psychoacoustic (Bark), and a modified version of the latter with the F1 scale expanded by 35% (see [6]). When we set out to explore our speaker's phoneme system, the advantages of the psychoacoustic display are obvious. A comparison of different visual scaling methods in Fig. 5, however, shows that the relative differences between linear and logarithmic displays are in fact not very great.

FINAL REMARKS

Acoustic vowel studies offer dialectologists a neutral basis for comparing sound systems across speakers. The 'free' variation contained in spontaneous speech may also be an invaluable indicator of a sound change in progress.

Compared with the *SED*, our data, for instance, suggest changes in progress in the rural speech of East Somerset. The open front vowel /a/ is losing some of its functional load, while the rounded back vowels are undergoing both qualitative and distributional changes [2].

REFERENCES

[1] Labov, W. (1994), *Principles of Linguistic Change. Volume 1: Internal Factors*, Oxford: Blackwell.

[2] Nevalainen, T. & Aulanko, R. (in press) Stressed vowels in East Somerset: An acoustic-phonetic case study. In J. Klemola, M. Kytö & M. Rissanen, eds., *Studies in Memory of Ossi Ihalainen*, Frankfurt am Main: Peter Lang.

[3] Wells, J.C. (1982) *Accents of English. Vol. 1.* Cambridge: Cambridge University Press.

[4] Orton, H. & Wakelin, M.F. (1967–1968) *Survey of English Dialects, The Basic Material, Vol. IV, The Southern Counties, Parts I–III.* Leeds University.

[5] Toivonen, R. (1988) Intelligent Speech Analyser (ISA) järjestelmän käyttöohje. Tampere, 22.6.1988.

[6] Iivonen, A. (1992) Articulatory vowel gesture presented in a psychoacoustical F1/F2-space. In R. Aulanko & M. Lehtihalmes, eds., *Studies in Logopedics and Phonetics 3*, Publications of the Department of Phonetics, University of Helsinki, Series B, 4, 19–45.