Introduction

The vowels in the words deeds and food have been phonologically analyzed as monophthongs. However, formant stability in these vowels varies across dialects in the British Isles.

Standard Southern British English (sse) is known to exhibit rather diphthongized realizations of deeds and food whereas Scottish Highlands English (shl) has true monophthongs (see Fig. 1 and Fig. 2).

Goal:

Can the degree of monophthong diphthongization alone constitute a reliable cue to dialect classification?

Comparison of classification based on formant trajectories alone with classification performed by a trained phonetician

Figure 1: Broad-band spectrogram (300 Hz filters; 4000 Hz displayed) of the vowel in deeds spoken by a male speaker from the Scottish Highlands.

Figure 2: Broad-band spectrogram (300 Hz filters; 4000 Hz displayed) of the vowel in deeds spoken by a male speaker of Standard Southern British English.

Perceptual experiment

Identification task: an expert native English phonetician was asked to listen to the vocalic portion of diphthongized and non-diphthongized realizations of deeds and food. He was told whether the stimulus was uttered by a speaker of the Scottish Highlands or a speaker of some other British English dialect (see) known to diphthongize these vowels to a certain extent.

Three types of stimuli:

- Natural vowels (nat): the vowels of deeds and food were segmented (boundary placed at the inception of the formant structure and at the end of it), and extracted, and normalized for amplitude
- Neutralized duration (ND): mean duration of the vowel in deed was divided by the vowel in food (normalized over the vowel tokens from the two dialects)
- Neutralized and speeded vowel (syn): vowel re-synthesized at 80% speed; monophthongized

Acoustic measurements

Three measures of diphthongization

(Central frequency for the first three formants in Bark measured with the Praat program)

1. ΔF: difference between the monophthong and diphthong in F1 and F2 at 20% of duration

2. Δf: standard deviation of 9 values extracted from each formant: every 10% of the duration

3. Δ: 9 values extracted from each formant (same as above)

\[ \Delta = \frac{1}{9} \sum_{i=1}^{9} |F_i - \bar{F}| \]

Where

- \( F_i \): frequency of a given formant (F1, F2 or F3) at point i (1/10th duration between measurements) of total vowel duration

Classification: linear discriminant analysis

Comparison of classification based on formant trajectories alone with classification performed by a trained phonetician

Results

Table showing accuracy of classification for different measures and conditions.

Conclusions

In the perceptual experiment, the phonetician managed to correctly classify above chance level only the diphthongized vowel, specifically in the natural condition. Although diphthongization must have helped, it seems that duration was the main cue used in this task. The poor results for food suggest a ceiling effect perhaps due to lack of within-dialect homogeneity.

The best classification score with linear discriminant analysis for deeds using formants is achieved with ΔF and F1 trajectory as the most relevant dimension. However, diphthongization constitutes the most reliable cue. As for food, ΔF is the best metric, with F2 showing the highest correlation with the discriminant function. Duration here only plays a marginal role.