TOWARDS A PHONETIC DEFINITION OF DIPHTHONGS

Emily Gorman[1], Sam Kirkham[1] and Patrycja Strycharczuk[2]
Lancaster University[1], University of Manchester[2]

A distinction between monophthongs and diphthongs is widely assumed, and it forms a part of a dominant methodological approach to representing and comparing vowels. While this distinction is generally thought to have systematic articulatory, acoustic and perceptual correlates, monophthongs and diphthongs are rarely explicitly defined. In this study, we explore whether an objective definition can be plausibly formulated, based on unsupervised clustering of phonetic exponents of vowel dynamics in two accents of British English.

We analysed combined acoustic and midsagittal EMA data from 16 speakers of two accents of British English: Standard Southern British English (SSBE; N=8, 3 females), and West-Yorkshire English (WYE, N=8, 5 females). The two accents are known to differ considerably in their dynamic properties: WYE shows monophthongisation of several vowels that are diphthongal in SSBE. The speakers produced a full set of English vowels in monosyllabic words, preceded by a non-lingual consonant /p, b, f, v/ and followed by a coronal stop (5 repetitions). We extracted three types of measures from each token: acoustic duration of the vowel, Euclidean distance between normalised f1 and f2 measured at 20 and 80% into the vowel, and tangential speed profiles of the Tongue Dorsum sensor. We focused on the TD sensor, since it captures most information relevant to tongue displacement data in monophthongs [1]. We further reduced the variance in the tangential speed profiles to two Principal Components, using Functional Data Analysis [3]. The two PCs, illustrated in Figure 1, captured 99.9 % of variance in the data.

We submitted by-dialect and by-vowel median value of each measure (duration, Euclidean distance, PC1 and PC2) to k-means clustering, systematically testing various combinations of input measures and clusters. Based on this procedure, we observe the following. Acoustic duration systematically sets lax vowels (e.g. KIT, STRUT) from other vowels, but it does not differentiate between ostensible monophthongs and diphthongs. Clustering based on only acoustic measures (duration and Euclidean distance), or only articulatory measures (PC1 and PC2) provides unexpected results, grouping some diphthongs with lax vowels and *vice versa*. A combination of Euclidean distance, PC1 and PC2 (Figure 2) provides a grouping closest to what we would expect based on dialect descriptions [2, 4, 5]. However, two groups of diphthongs emerge in this case. SSBE PRICE and CHOICE and WYE CHOICE are distinct from a cluster of SSBE FACE, MOUTH,GOAT and WYE PRICE, MOUTH and FACE. SQUARE patterns with monophthongs in both dialects, as does WYE GOAT.

We propose that CHOICE and SSBE PRICE are prototypical diphthongs. They have low PC1 and PC2 values, which correspond to a rise-and-fall velocity profile, consistent with a presence of two vowel targets (see Figure 1). Acoustically, this is reflected in increased Euclidean distance. In contrast, monophthongs tend to have high PC1 and PC2 values, which correspond to a U-shaped velocity profile, suggesting a single vowel target. Accordingly, Euclidean distance values in monophthongs are relatively lower. Vowels such as MOUTH, SSBE FACE, SSBE GOAT and WYE PRICE form an intermediate category: they overlap in PC values with some monophthongs, but typically have greater Euclidean distance, which we tentatively interpret as increased acoustic displacement, but only one discernible articulatory target. Interestingly, while individual vowels can have intermediate values along all three measures, plausible clusters emerge nonetheless at their intersection. This suggests that dynamic properties of vowels can be categorically distinguished, but that such distinctions are inherently multi-dimensional.

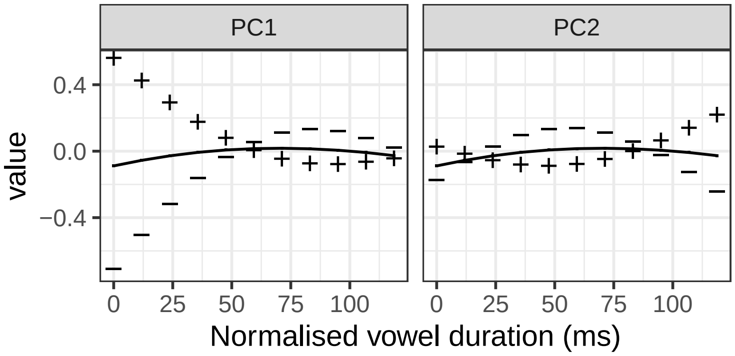

Figure 1. Principal Components 1 and 2

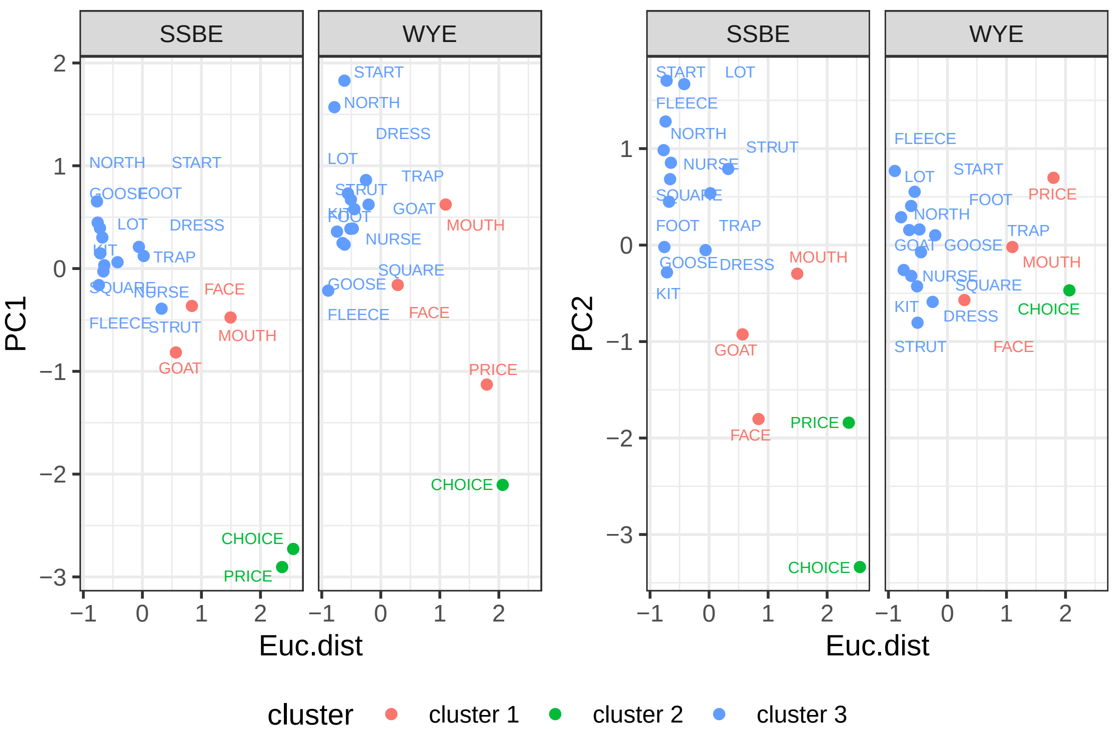

Figure 2. Clustering results

References [1] A. Blackwood Ximenes, J. A. Shaw, and C. Carignan. "A comparison of acoustic and articulatory methods for analyzing vowel differences across dialects: Data from American and Australian English". In: *The Journal of the Acoustical Society of America* 142 (2017). [2] K. Earnshaw and E. Gold. "Variation in the FACE vowel across West Yorkshire: Implications for forensic speaker comparisons". In: *Interspeech 2018: Speech Research for Emerging Markets in Multilingual Societies*. 2018. [3] M. Gubian, F. Torreira, and L. Boves. "Using Functional Data Analysis for investigating multidimensional dynamic phonetic contrasts". In: *Journal of Phonetics* 49 (2015). [4] J. Wells. *Accents of English 2: The British Isles*. Vol. 2. Cambridge: Cambridge University Press, 1982. [5] D. Williams and P. Escudero. "A cross-dialectal acoustic comparison of vowels in Northern and Southern British English". In: *The Journal of the Acoustical Society of America* 136 (2014).