

Diphthongs in English

Bryan JENNER

(Received September 4, 1995)

Abstract

The English vowel system is very much larger and more complex than that of Japanese. In particular it contains several diphthongs: a phenomenon which is entirely lacking in Japanese. Moreover, the General American and Standard British varieties are normally claimed to possess different numbers of diphthongs, which makes the teaching and learning of English pronunciation extremely difficult in the Japanese context. Proposals are made for a simplification of the descriptive apparatus, by showing that, at an underlying level, British and American diphthong inventories have the same number of members, and that some surface diphthongs in the British variety are simply allophonic variants of a short vowel +[-r].

Key words : segment, diphthong, phonetic, phonological, allophone

Diphthongs in Human Languages

Diphthongs are a relatively rare phenomenon in human languages. Of the 317 languages surveyed in the UCLA Phonological Segment Inventory Database (UPSID, reported in Maddieson, 1984), only 23 were found to have phonological diphthongs. More striking still was the fact that out of a total of 2,549 vowel segments isolated in all of these languages, only 83 were clearly definable as diphthongal. Since more than a quarter of these are found in one of the Khoisan languages of Southern Africa, this makes the existence of any diphthongs at all in the remaining languages an extremely rare occurrence and worthy of special attention.

British English is frequently reported to have 8 diphthongs (cf. Gimson 1970) or even 9 (cf. Jones 1956), which makes it one of the richest sources in the world for the study of diphthongs. Some analysts, it must be said, draw a distinction between “true diphthongs” (of which British English is reported to have 3) and “vowel glides” (which account for the remaining 5, but perhaps also for one or two other “long vowels” which would not normally be defined as diphthongs).

American English, on the other hand, is normally reported to have 5 diphthongs, while Scottish English has only 3. Among other Indo-European languages, the West Germanic languages have at least 3, Slavic languages have 1 or 2 at the most, and of the Romance languages, only Portuguese seems to have diphthongs which in any sense resemble those of English and the other West Germanic languages.

There is, however, a certain amount of confusion in the literature as to what constitutes a diphthong and it would therefore be fruitful to examine a number of alternative definitions. In so doing we may also be able to shed some light on more general features shared by those languages which have relatively high numbers of diphthongs in their vowel inventories.

Some Definitions

Since diphthong is a term which is used by both phoneticians and phonologists it is not surprising that it has been used to cover a variety of phenomena and given a number of different (explicit or implicit) definitions.

Phonetic Definitions.

Phonetic definitions tend to refer to physically quantifiable articulatory gestures which are found in some segments but not in others. This type of definition also requires a further definition of the notion of segment in strictly phonetic rather than phonological terms.

Laver (1994) offers the following definition of segment :

.. a linear unit typically anchored in a short stretch of speech by a set of phonetic feature-values which are relatively unchanging. The segment is a construct of phonetic theory which relies here on a related concept of three different phases of articulation of any segment. (p. 112)

These three phases are subsequently defined as the medial phase, where the maximum degree of vocal tract constriction is achieved, and this is preceded by an onset phase and followed by an offset phase. This latter shows

the movement of the organs towards the medial phase of the next segment (and hence constitutes an overlapping phase with the onset phase of that next segment). (ibid.)

Elsewhere Laver applies this 3-phase analysis to monophthongs and diphthongs, suggesting that monophthongs show “a relatively unchanging quality through the medial phase”, whereas diphthongs are characterized by a “unidirectional change” in this phase of articulation. (p. 143)

Phonological definitions : Diphthongs and the Syllable

Against this very attractive physical type of definition we must set the problems posed by the need to provide a phonological definition of diphthongs which makes some reference to a perceptual concept of *syllable*. This has proved extremely elusive, and attempts to find a definition of a syllable based on auditory, acoustic or physiological phenomena have, over a period of many years, been inconclusive. (v. Ladefoged, 1982, p. 219 ff., for some discussion)

That is not to say that the syllable is not a physical reality : native speakers can normally count them without difficulty, and poetry in many languages (Japanese, French, Latin) is based on the native speaker's ability to perceive and produce an exact number of syllables in accordance with certain artistic conventions.

There are, however, “grey areas”, where even native speakers disagree about the number of syllables in a given stretch of speech : one such area involves words like *here* and *real*, or *hour* and

tower (all with non-rhotic pronunciation) where morphological considerations may cause disagreement as to whether the (phonetically identical) double vowels are perceived as belonging to one syllable or two. Many native speakers would, for example, allocate a single syllable to *here* and *hour*, but would be quite certain that *real* and *tower* were disyllabic, because of phonological changes which affect their vowels in derivative words.

If we can accept the syllable as a physical reality and also as a useful phonological construct, we may safely say that a “diphthong is a double articulation vowel which belongs to a single phonological syllable”. In Laver’s purely phonetic terms, it would be a vowel segment with a “unidirectional change in the medial phase”.

Diphthongs and the Vowel Inventory

A very interesting further dimension to the phonological type of definition has recently been added by Giegerich (1992), who observes that in British English all vowel phonemes may be grouped in contrasting pairs, with the exception the 3 three true diphthongs /aɪ/, /aʊ/ and /ɔɪ/. The pairs which he identifies are /i:/ - /ɪ/, /u:/ - /ʊ/, /e:/ - /e/, /ɑ:/ - /ɑ/ (= /æ/), /o:/ - /ʌ/, and /ɔ:/ - /ɒ/.

In American English the same system may be applied, with the exception that the /ɔ:/ - /ɒ/ contrast, as in *law-cough*, has been lost. A similar process is found in Scottish where the inventory is reduced by a symmetrical absence of long-short oppositions in /u:/ - /ʊ/, /ɑ:/ - /ɑ/, and also, as in American English, in /ɔ:/ - /ɒ/.

In both the British and American varieties Giegerich treats the so-called diphthongs /eɪ/ and /əʊ/ (or /oʊ/) as realisational variants of underlying long vowels which he represents as /e:/ and /o:/. The true diphthongs, on the other hand, are common to all varieties of English

On Phonemes and Allophones: an example from English

A phoneme is conventionally defined as a unique phonological segment in a given language capable of forming contrasts with all other segments of the same class in that language. An allophone is a contextual or idiosyncratic variant of a phoneme.

Applied to the English language - and probably to other “world languages” as well—these definitions are simplistic and troublesome, since they depend crucially on a further definition of the term *language*. For example, it is conventionally claimed that modern British English has 3 diphthongal phonemes (/ɪə/, eə/ and /ʊə/) which are not found in General American. In this rhotic variety of English the 3 phonemes in question are realised as /ɪ/, /e/ and /ʊ/ plus /-r/. In other words, they are identical to 3 short vowel phonemes found in other phonetic environments. If we wish to retain the definition of phoneme given above we would be forced to claim that British and American English are different languages - at least at the phonological level—since they do share the same number of phonemes.

It would be much simpler - and theoretically more powerful - if the two sets of sounds could be handled in the same way in both major varieties of English. We have already outlined Giegerich’s argument for treating the long - short pairs in this way for all varieties of English. We may now propose that in the British variety, like the American, the 3 sounds in question are not diphthongal

at all but consist of a simple short vowel + some variant of /-r/. The British English post-vocalic /-ə/ would simply be a variant of /-r/ in certain contexts. This would be only part of a process of looking for a more unified “archi-variety” of English of which British and American are particular manifestations.¹⁾

Before that could be accepted, however, it is necessary to examine how the process would work in other [V + -r] environments, such as /a:r/, /ɜ:r/, and /ɔ:r/, as in *car*, *nurse* and *horse* respectively. In these contexts American English has the normal rhoticity, while British English has a perceptually simple long vowel. Unlike the 3 segments considered earlier (/ɪə/, /eə/ and /ʊə/), however, which have an initially raised tongue position, these 3 have low or mid tongue height which renders the travel to a shwa position so slight as to be imperceptible. In other words, we might claim that the shwa is there but we cannot hear it. It is manifested as a simple relaxation of the articulatory apparatus with no acoustically measurable results.

To pursue this line of analysis further, we might claim that in (southern) British English post-vocalic -r in final position is always realised as /ə/, even though it cannot always be heard, whereas in both American and in Scottish English it is realised as some variety of rolled or trilled -r. This would have the advantage of reducing the number of vowel segments in the British inventory by 3, leaving only the 3 true diphthongs which are common to all varieties of English, and particular allophonic realisations of /ɔ:/ and /e:/ which have become ‘glides’ in the 20th century.

The matter of diphthongal /e:/ and /ɔ:/ in standard British English may be considered together with the behaviour of other “simple long vowels” which have tended to become glides in recent years. The raised vowels /i:/ and /u:/ are habitually glided to something like [iɪ] and [ʊ] in the speech of the younger generation of British speakers, with the degree of lip-spreading or rounding being considerably reduced or delayed. The feature which unites these phenomena with the British treatment of /e:/ and /ɔ:/ is one of relative tongue raising and jaw closing: all long vowels are currently articulated with as little articulatory effort as possible.

The “Principle of least effort”, discussed in Wells 1982, asserts that an articulatory simple gesture will always be preferred to a complex or effortful one. In its Wellsian form, however, it does not explain how something which is effortful in one variety or language may be comparatively effortless in another. One way out of this dilemma is to postulate that the American and British varieties of English, for instance, differ in their overall tongue configurations, or ‘articulatory setting’: a British tongue is habitually low and relatively flat, except for some raising of the front portion, whereas an American tongue is characterized by habitual raising of the central section towards the palate. This fundamental postural difference serves to make post-vocalic -r effortful in the British variety, but effortless or ‘normal’ in American English. (For further discussion of the concept of *articulatory setting*, v. Honikman 1964)

¹ This type of vocalic realisation of a continuant is also found in the southern British realisation of /l/ as /ʊ/ in such contexts as “field”, “old”, or “held”, which—in the speech of at least one former British Prime Minister—were normally realised as /fi:ʊd/, /ʊʊd/ and /heʊd/ respectively. A similar phenomenon may be observed in the Portuguese realisation of final -l as /ʊ/ in /braziʊ/.

Conclusion

We may summarize by claiming that all long vowels in British English are potentially glides from a more fronted or open position to a more central or back position: this is realized allophonically as /ə/, /ʊ/, /ɪ/ or zero, depending on the environment. This is true of the vowels /o:/ (realized as /oʊ/ or /əʊ/), /e:/ realized as /eɪ/, /u:/ (realized as /ʊə/), and /i:/ (realized as /ii/). It also holds true for the vowels /ɑ:/, /ɔ:/ and /ɜ:/ where the glide is so insignificant as to be imperceptible.

The so-called diphthongs in *beer*, *there*, *poor*, are therefore not diphthongs at all, since they contain no element of a long vowel. On the contrary, they are simple short vowels /ɪ/, /e/, and /ʊ/, followed by an allophonic variant of /r/. The three 'true diphthongs', which, as we argued above, do not form pairs with any other vowels, consist of a first element which is otherwise a long vowel in both British and American English (/ɑ/ or /ɔ/), followed by one of the short vowels /ʊ/ or /ɪ/.

Diphthongisation and its counterpart, vowel 'simplification', are ongoing processes in languages: one may predict that in diphthong-rich languages most long vowels will ultimately tend to glide (as in British and American English). Conversely, languages with a low incidence of diphthongs will ultimately be entirely free from diphthongs (as is the case in French). In articulatory terms, this means that languages which have a low degree of tension will tend to become even laxer over a period of time, while tense languages will become still more tense. These processes may of course be delayed by conservative educational intervention, but ultimately a language will tend to have a either very large number of gliding vowel phonemes (like English), or none at all (like Japanese).

References

- Giegerich, H. J. (1992) *English Phonology*, chapter 3, Cambridge: Cambridge University Press
- Gimson, A. C. (1970) *An Introduction to the Pronunciation of English*, pp. 126–146, London: Arnold
- Honikman, B (1964) 'Articulatory settings', in *In Honour of Daniel Jones*, pp. 73–84
(ed. D. Abercrombie et al), Harlow, UK: Longman
- Jones, D. (1956) *The Pronunciation of English* (4th edition), pp. 52–68, Cambridge: Cambridge University Press
- Ladefoged, P. (1982) *A Course in Phonetics* (2nd edition), pp. 219–224, New York: Harcourt Brace Jovanovich
- Laver, J. (1994) *Principles of Phonetics*, p. 112, Cambridge: Cambridge University Press
- Maddieson, I. (1984) *Patterns of Sounds*, Cambridge: Cambridge University Press
- Wells, J. (1982) *Accents of English* (3 volumes) *Volume 1: An Introduction*, Cambridge: Cambridge University Press