409

LANGUAGE AND SPEECH, Vol. 25, Part 4, 1982

SOME ACOUSTIC AND PHYSIOLOGICAL OBSERVATIONS ON DIPHTHONGS[*]

RENÉ COLLIER[**]
University of Antwerp
FREDERICKA BELL-BERTI[†]
St. John's University
and
LAWRENCE J. RAPHAEL[†]
Herbert H. Lehman College, The City University of New York

This paper presents an analysis of some articulatory properties of (Dutch) diphthongs, attempting to correlate articulatory inferences based on perceptual and acoustic data with more direct physiological measurements (recordings of EMG activity). Evidence is presented that supports a distinction between "genuine" and "pseudo" diphthongs: the two classes appear to differ (1) in openness and advancement at their onsets and offsets, (2) in the harmony of tongue position between the beginning and ending configurations, and (3) possibly also in the number of articulatory gestures involved.

INTRODUCTION

It has long been customary to transcribe diphthongs using two phonetic symbols that, used separately, represent simple vowel and semivowel segments. To judge from these impressionistic transcriptions, any two diphthongs may differ minimally in either their onset or offset qualities. For example, in Dutch the diphthong /εi/ is said to end with a high front vowel, whereas the diphthong /aj/ is said to end with an acoustically similar semivowel. In such instances one might ask whether these transcriptions – that reflect perceptual differences between two sounds – also reflect measurable differences in acoustic structure and articulatory strategy. Furthermore, we might ask whether the symbols used in the impressionistic transcription of the diphthongs have the same acoustic and articulatory values as do the simple vowel and semivowel segments that they represent. Finally, does conventional transcription practice reflect the perceptual impression that these sounds are composed of two separate segments and, if so, are they produced as a sequence of two articulatory gestures? These questions may best be addressed in a language containing the simple vowels and semivowels used in transcribing

---

[*] *This work was supported by NINCDS grants NS-13617 and NS-05332, BRS grant RR-05596 to Haskins Laboratories and by the National Science Foundation of Belgium (NFWO).*

[**] *Also Institute for Perception Research (IPO).*

[†] *Also Haskins Laboratories*

its diphthongs.

We have chosen to study Dutch because it is a language containing a sufficient number of diphthongs to allow one to answer the questions we have raised above. In fact, it is claimed that Dutch has two types of diphthongs: "genuine" ones (/ɛi, ʌy, ɑu/) and "pseudo" ones (/aj, oj, uj, iw, ew/).[1] There has been little consensus among Dutch phoneticians and phonologists as to what characterizes each class of diphthong. Matters have been further complicated by the existence in Dutch of "long" or "tense" vowels that tend to be diphthongized as well, [e^{i}, ϕ^{y}, o^{u}], possibly in still a different way (Koopmans-van Beinum, 1969; 't Hart, 1969).

A good survey of how phoneticians and phonologists have interpreted the nature of Dutch diphthongs is given in Zonneveld and Trommelen (1980). It appears that from the end of the nineteenth century until about 1940 most phoneticians did not make a principled distinction between diphthongs and (long) vowels, and – *a fortiori* – did not differentiate between genuine and pseudo diphthongs. Yet they realized that diphthongs consist of two (or more) elements and can be classified according to the relative openness of their first component and (or) the frontness vs. backness of their second. There was also some discussion as to whether the components correspond to vowels that can occur in isolation. The structural phonologists of the thirties raised the question of whether diphthongs should be given a monophonemic or biphonemic representation. They tended to agree that the genuine diphthongs are single phonemes whereas the pseudo ones consist of two phonemes each. This point of view was still endorsed by Van den Berg (1959), whereas Cohen *et al.* (1959) considered all diphthongs to be biphonemic. Generative phonologists, too, have generally preferred a biphonemic underlying representation for the Dutch diphthongs, but they have shown a wide divergence of opinion as to the nature of the two segments involved.

In recent years better instrumental and experimental techniques have produced a more reliable phonetic specification of the genuine Dutch diphthongs. A perceptual analysis has resulted in the following characterization:

> [ɛi] is the Dutch vowel [ɛ], followed by *movement in the direction of* [i]; [ʌy] is the English vowel [ʌ] (as in *cup*) – and not the Dutch [œ] – followed by *movement in the direction of* [y]; [ɑu] is the Dutch vowel [ɑ] – not [ɔ] – followed by *movement in the direction of* [u]. The endpoints [i, y, u] are reached only in careful, isolated pronunciation, with no final consonant. Usually the endpoints are [ɪ], [ϕ] and [o]. ('t Hart, 1969, p. 172. Our translation, his italics).

---

[1] *Possible occurrences of these diphthongs in Dutch words:*

/ɛi/	kei	*'pebble'*	*/aj/*	maai	*'mow'*
/ʌy/	lui	*'lazy'*	*/oj/*	mooi	*'beautiful'*
/au/	rauw	*'raw'*	*/uj/*	snoei	*'trim'*
			/ew/	leeuw	*'lion'*
			/iw/	nieuw	*'new'*

Another pseudo diphthong, /yw/ as in duw *'push', was not included in our utterance set.*

Thus, we find a new emphasis on the dynamic character of the genuine diphthongs and a shift away from the traditionally assumed importance of onset and offset qualities. Spectrographic analysis has revealed that the genuine diphthongs are mainly characterized by a relatively unchanging F_2 and an avalanche-like decrease of F_1 (Mol, 1969).

Cohen (1971, p. 288) summarizes the results of these acoustic and perceptual studies as follows:

> There are a number of arguments . . . for accepting the diphthongs of the Dutch ɛ*i*, ʌ*y*, a*u* type as vocoids, recognizable as such and distinguishable from the other vocoids of the long and short classes, on account of their peculiar, dynamic character.

Pols (1977, p. 103) summarizes his own recent findings by noting that:

> A diphthong can be described as quite a long steady-state onset part followed by a fast specific transition to an offset area where no steady-state part is necessary. The diphthong [ɑu] starts at [ɑ] and terminates at [ɔ, o]; [ɛi] starts at [ɛ] and goes to [ɪ, e]; and [ʌy] starts at [a] and goes to [œ, ø]. So, none of the three Dutch diphthongs reaches the vowel position indicated in its phonetic transcription.

Pols also notes that the acoustic variability of these diphthongs is very large. This variability correlates well with the fairly large perceptual tolerance observed by Slis and Van Katwijk (1963), who studied the acceptability of two-formant synthetic diphthongs having a great variety of beginnings and endpoints in the F_1-F_2 plane.

As for the pseudo diphthongs, there has been little or no controversy over their essential characteristics. They have been and still are considered to be sequences of a "tense" vowel and a semivowel. They start with a vowel whose quality is the same as that of the separately occurring vowels [a, e, o, i, y] and move into the glides [j] and [w]. Phonetically they are the sum of their components.

Comparing the characteristics of the genuine and pseudo diphthongs, we find that they differ in a number of respects, including: (1) the degree of "openness" at onset; (2) the degree of change in tongue advancement between onset and offset; and (3) the degree of harmony between lip position at onset and offset. For example, each of the genuine diphthongs starts with a relatively open vocal tract and ends with a relatively closed one. A pseudo diphthong, on the other hand, may start with an open, half open, or closed vocal tract, before ending with a semivowel. Furthermore, each of the genuine diphthongs ends with a vocal tract shape in which tongue advancement and lip position are approximately the same as they were at the beginning of the diphthong. Each pseudo diphthong, however, ends with a vocal tract shape in which tongue advancement and, usually, lip position are different than they were at the start of the diphthong.[2] In addition, the genuine diphthongs are characterized by relatively continuous and gradual changes in formant structure, whereas the pseudo diphthongs are produced with more abrupt changes in formant structure (Figure 1).

Since there were no physiological data on the production of Dutch diphthongs, the

---

[2] *Although /aj/ is said to begin with the low front vowel [a], the data we offer below imply a substantial back-to-front movement during this diphthong.*

Fig. 1. Single token examples of the genuine diphthong [ɛi] and the pseudo diphthong [aj], spoken in isolation.

available acoustic and perceptual information led us to hypothesize that there must also be significant differences between the two classes of diphthongs in the articulatory domain. Therefore, the primary aim of our study was to explore how changes in vocal tract configuration are brought about in each of these diphthongs, in order to determine whether physiological descriptions would support their traditional separation into two classes on the basis of acoustic, perceptual, and articulatory phonetic descriptions. To this end, we have, necessarily, described their production in some detail, to provide a base for making the relevant comparisons.

PROCEDURES

We simultaneously recorded both acoustic and electromyographic (EMG) signals from one speaker of Dutch.[3] The EMG potentials were recorded from four muscles known to affect the position of the tongue and the mandible: the genioglossus, styloglossus, mylohyoid, and anterior belly of the digastric. Previously reported physiological data have led us to three groups of assumptions.

Assumptions concerning the functions of the muscles studied

The genioglossus is the only muscle known to contribute significantly to tongue advancement (Smith, 1971; Kakita, 1976; Alfonso and Baer, 1982). It has also been implicated in tongue bunching/raising gestures, although its activity in this regard accompanies activity of other intrinsic and extrinsic tongue muscles (Miyawaki *et al.*, 1975; Raphael and Bell-Berti, 1975; Raphael *et al.*, 1979). The styloglossus is primarily responsible for retraction of the tongue body (Smith, 1971; Raphael and Bell-Berti,

---

[3] *Our subject, the senior author, speaks the Belgian variant of Standard Dutch.*

1975). Both genioglossus and styloglossus act with the mylohyoid to elevate the tongue, with the mylohyoid providing the greatest portion of the vertical thrust (Raphael *et al.*, 1979). The mylohyoid may also act to stabilize the hyoid bone, in conjunction with the activity of the anterior belly of the digastric, which assists in lowering the mandible (Raphael *et al.*, 1979).[4]

Assumptions concerning the relationship between the acoustic signal and vocal tract shape for vocoids

It is possible to calculate formant frequencies from a given vocal tract shape and, given a set of formant frequencies, to infer characteristics of the vocal tract shape that produced it (Chiba and Kajiyama, 1941; Delattre, 1951; Stevens and House, 1955, 1961; Fant, 1970). The methods of calculating formant frequencies have been sufficiently refined over the years to generate a near-unique solution for any tract shape. Although the inference of tract characteristics from formant frequencies is less certain, it is widely accepted that the frequency of F_1 is primarily dependent upon the degree of vowel openness, and the frequency of F_2 is primarily dependent upon the length of the front cavity (Stevens and House, 1955, 1961; Fant, 1970; Kuhn, 1975). Thus, for instance, a more open vowel will have a higher F_1 than a more closed one, a fronted vowel will tend to have a higher F_2 than a retracted one, and a rounded vowel will tend to have a lower F_2 than an unrounded one.

Assumptions concerning temporal relationships between EMG potentials and movement

EMG potentials precede their mechanical effect (cf. Harris, 1981). The "contraction times" for the muscles included in this study are of the order of 70-100 msec; that is, movements associated with EMG potentials begin about 70-100 msec after the electrical activity begins.

Pairs of bipolar hooked-wire electrodes were inserted into the genioglossus (anterior fibers), mylohyoid, styloglossus, and anterior belly of the digastric muscles, using standard procedures which are described elsewhere (Hirose, 1971; Raphael and Bell-Berti, 1975). The nonsense test utterances were of the form [də'pDpapə], where D = /aj, oj, uj, ew, iw, ɛi, ʌy, ɑu/, and [ə'pVp], where V = /i, u, e, ɛ, a, ɑ, y, œ, o, ɔ]. The subject read from randomized lists of the utterances until he had produced 16 tokens of each. The recordings of all tokens of each of the eight utterance types were aligned with reference to the onset of vocal fold vibration in the diphthong. The EMG potentials were rectified, integrated, and computer-sampled, and ensemble averages of the EMG potentials were then calculated for each channel for each utterance type. The EMG data processing system is described in greater detail in Kewley-Port (1973).

In addition to the EMG analysis, we performed acoustic analyses with a digital-waveform and spectral-analysis system. Ensemble averages of both the amplitude envelope of the audio waveforms and of digital spectrograms were also calculated.

---

[4] *We recognize, of course, that more than four muscles are involved in positioning and shaping the tongue, and that the articulatory description provided here is, of necessity, a simplified one.*

RESULTS

We shall describe the EMG and acoustic data in relation to the traditional articulatory phonetic, perceptual, and acoustic descriptions, provided above, concerning the differences between the genuine and pseudo diphthongs. As we have explained above, there is no one-to-one correlation between articulator position and muscle potentials, nor may a unique vocal tract shape be derived from a set of formant values. Hence, we will not attempt to specify absolute articulator position (i.e., vocal tract shape) on the basis of our acoustic or physiological data. Rather, we will compare the data on the diphthongs among themselves and with the data on simple vowels, to infer relative differences in the articulatory parameters. We shall consider first the hypothesis that the two groups differ in openness and advancement at their onsets and offsets. In addition, the onsets and offsets of the genuine diphthongs differ in openness and advancement from the simple vowels and semivowels described as their starting and ending positions, whereas the pseudo diphthongs do not. The second hypothesis is that the groups differ in the harmony of tongue position between the beginning and ending configurations.[5] Finally, we shall examine the hypothesis concerned with whether or not the two groups of diphthongs are specified as different numbers of discrete gestures; that is, that the genuine diphthongs are specified as single gestures whereas the pseudo diphthongs are specified as two discrete, concatenated gestures.

Hypothesis 1: Openness and Advancement

Openness

Traditionally, the genuine diphthongs of Dutch were described as proceeding from relatively open to relatively close articulatory positions, whereas the pseudo diphthongs proceed from various degrees of open to close articulatory positions. Thus, the articulations of the genuine diphthongs were said to begin with relatively open positions (similar to those of /ɛ, ɔ, œ/) and to end with the close positions of [i, u, y], respectively. In contrast, the articulations of the pseudo diphthongs were said to begin with the appropriate degrees of openness for the vowels /a, o, u/ and /e, i/, and to end with the close positions of the semivowels /j/ and /w/, respectively.

Genuine diphthongs. As stated in the introduction, perceptual analyses of the genuine diphthongs have revealed that these diphthongs – especially [ɑu] and [ʌy] – tend to be more open at their beginnings than are the simple vowels used in former transcriptions. This point is fairly well supported by our acoustic and physiological data. The acoustic data in Figure 2a and Table 1 indicate that [ɑu] and [ʌy] have higher F_1 values at their onsets than the simple vowels [ɔ] and [œ]. Hence they are likely to be more open at their beginnings. In fact, [ɑu] has the same F_1 onset value as [ɑ].[6] On the other hand

---

[5] *We will not address the question of whether the genuine and pseudo diphthongs differ in maintaining harmony of lip position between starting and ending configurations.*

[6] *We are unable to compare the relative openness, or frontness, of the beginning of [ʌy] with [ʌ] because of the absence of this latter vowel in Dutch.*

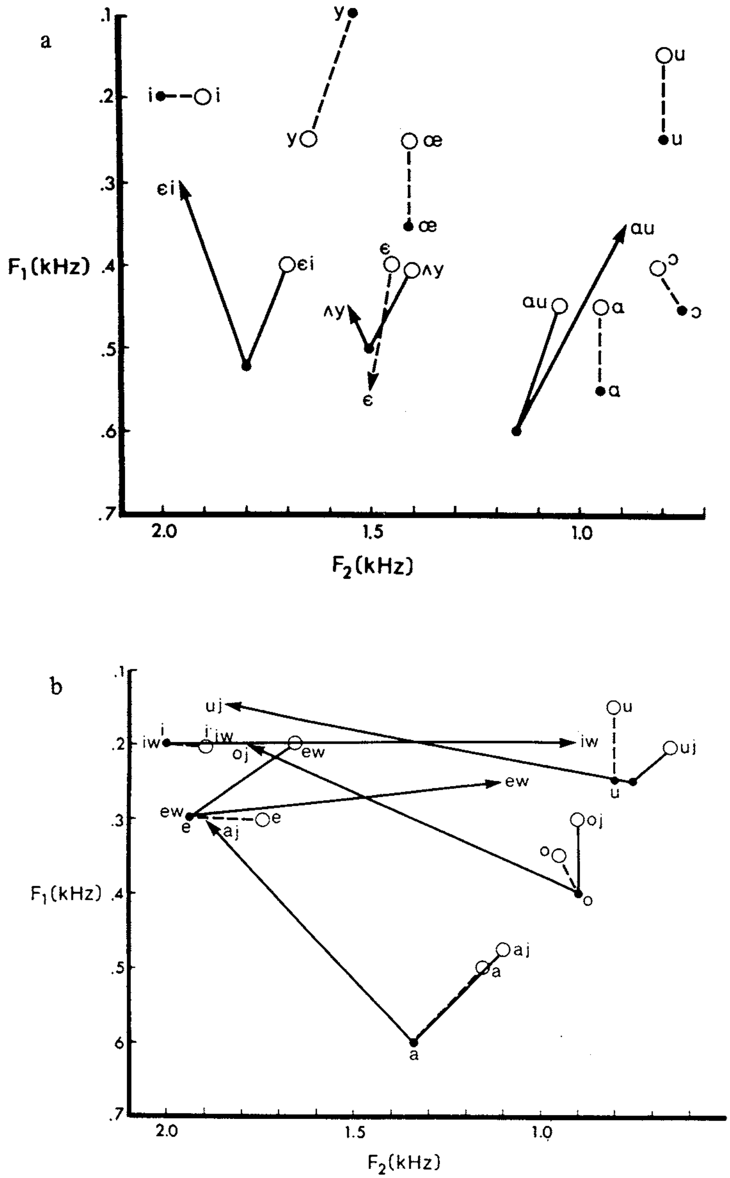


Fig. 2. Formant trajectories, in F_1-F_2 plane, of genuine diphthongs (a) and pseudo diphthongs (b), and simple vowels traditionally said to begin and end them. Open circles indicate onset values, filled circles indicate midpoint values, arrowheads indicate offset values (shown only for diphthongs). Solid lines connect diphthong values, dashed lines connect simple vowel onset and midpoint values.

TABLE 1

Averaged formant values (one speaker, 16 repetitions) for three genuine and five pseudo diphthongs, recorded during the experiment. Measurements are based on sections at onset, midpoint, and offset, and are compared with formant values of simple vowels recorded during the same experimental session.

A. Genuine Diphthongs

	ONSET			MIDPOINT			OFFSET	
	[εi]	[ε]		[εi]	[ε]		[εi]	[i]
F_1	400	400		525	550		300	200
F_2	1700	1450		1800	1500		1950	2000
	[ʌy]	[œ]		[ʌy]	[œ]		[ʌy]	[y]
F_1	400	250		500	350		450	250
F_2	1400	1400		1500	1400		1550	1650
	[ɑu]	[ɑ]	[ɔ]	[ɑu]	[ɑ]	[ɔ]	[ɑu]	[u]
F_1	450	450	400	600	550	450	350	250
F_2	1050	950	800	1150	950	750	900	800

B. Pseudo Diphthongs

	ONSET		MIDPOINT		OFFSET	
	[aj]	[a]	[aj]	[a]	[aj]	[i]
F_1	475	500	600	600	300	200
F_2	1100	1150	1350	1350	1900	2000
	[oj]	[o]	[oj]	[o]	[oj]	[i]
F_1	300	350	400	400	200	200
F_2	900	950	900	900	1800	2000
	[uj]	[u]	[uj]	[u]	[uj]	[i]
F_1	200	150	250	250	150	200
F_2	650	800	750	800	1850	2000

	[iw]	[i]	[iw]	[i]	[iw]	[u]
F_1	200	200	200	200	200	250
F_2	2000	1900	2000	2000	900	800

	[ew]	[e]	[ew]	[e]	[ew]	[u]
F_1	200	300	300	300	250	250
F_2	1650	1750	1950	1950	1100	800

[εi] has about the same onset F_1 value as [ε]. As far as the EMG data for [εi] are concerned (Figure 3), there is more anterior-belly-of-the-digastric activity at its onset than for [ε], but this tongue-lowering action may be compensated for by stronger genioglossus activity. At the onset of [ʌy] there is far more anterior-belly-of-the-digastric activity than for [œ]. The tongue-lowering effect of this action is only partly counterbalanced by the high peak in mylohyoid activity, because this comes late, mainly associated with the later portion of the diphthong. Therefore [ʌy] is likely to have a more open onset position than [œ]. The onset of [ɑu] is very similar to that of [ɑ]: the peaks of mylohyoid and anterior-belly-of-the-digastric activity are roughly the same.

At their ends the first formant frequencies of the genuine diphthongs reflect degrees of openness greater than those of the simple vowels [i, y, u]. Interpreting the EMG data, we must assume that the strong, early, jaw-and-tongue lowering activity of the anterior belly of the digastric is not entirely compensated for by the strong, later, tongue-raising activity of the genioglossus, styloglossus, and mylohyoid. In other words, [εi, ʌy, ɑu] do not terminate with the target vowels suggested in their transcriptions. This finding is in agreement with the perceptual analysis by 't Hart (1969).

Pseudo diphthongs. The pseudo diphthongs appear to achieve relatively stable first formant frequency values (Figures 5b and 5c), which reflect openness positions equivalent to those of the simple vowels said to begin them (Figure 2b). The EMG data (Figure 4), while somewhat less straightforward, do not contradict the inferences drawn from the acoustic measurements. At the onset of [aj] the EMG values are very similar to those for [a], except that genioglossus activity begins later for the diphthong. [ew, iw, oj] appear to begin with the same balance of tongue raising and lowering activity as [e], [i], and [o], respectively. For instance, at the onset of [iw] there is less tongue fronting and raising activity in the genioglossus than for [i], but by much stronger mylohyoid contraction. Similarly, the antagonistic forces of styloglossus and anterior belly of the digastric are reversed at the beginning of [oj] as compared to [o]. At the beginning of [iw] the earlier and stronger mylohyoid activity probably compensates for the reduced genioglossus activity in comparison with [i]. Only in the case of [uj] is there no apparent compensation for the reduced activity of the mylohyoid when compared with [u], but possibly the early onset of genioglossus contraction (associated with [j]) contributes to early tongue raising for this diphthong.

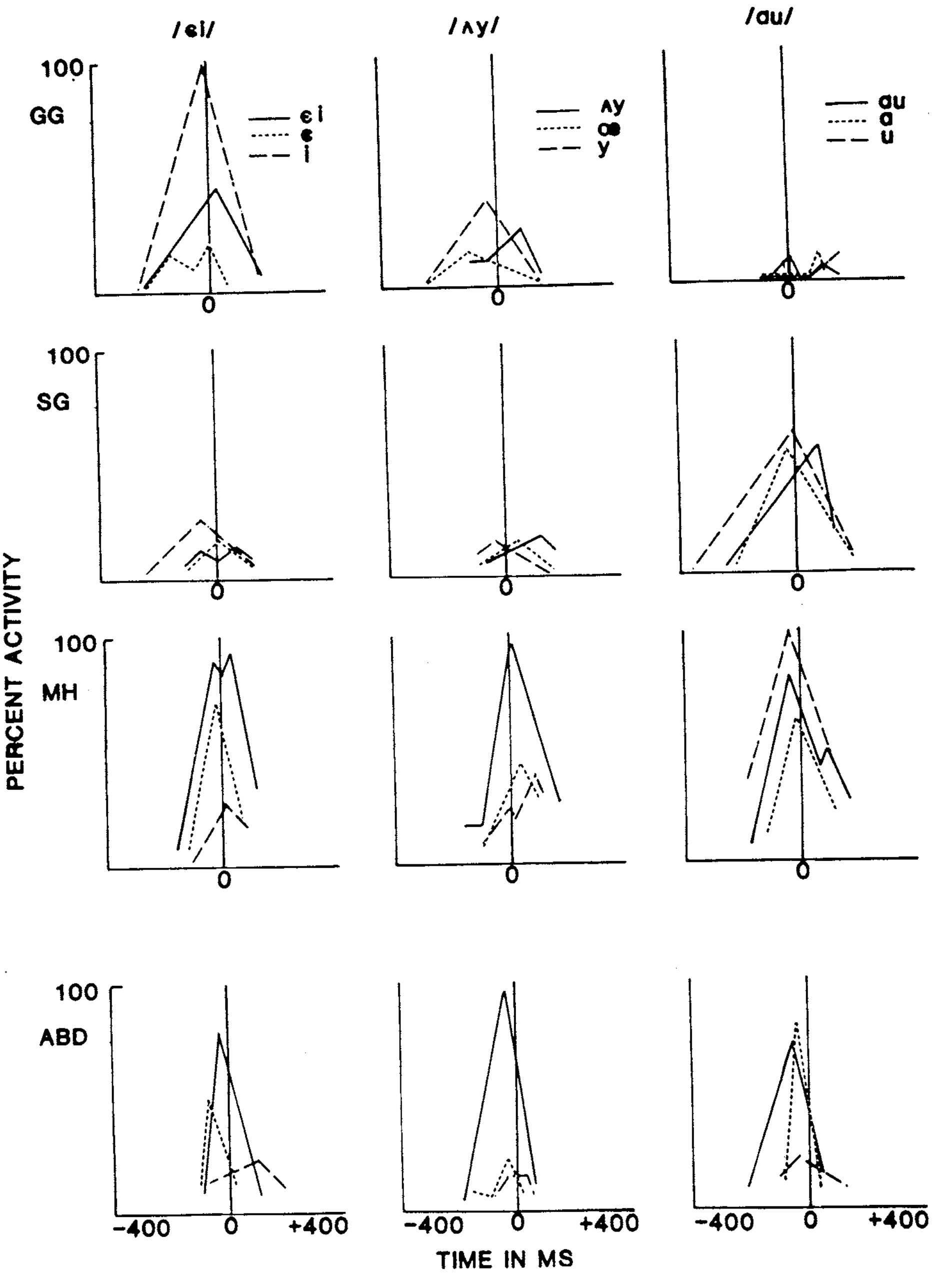


Fig. 3. EMG data for genuine diphthongs and simple vowels used in describing them. Each graph is a schematized representation of the time course of EMG activity in a given muscle, expressed as a percentage of the overall range of that muscle's activity across utterance types. Zero on the abscissa represents the acoustic onset of the diphthongs and the simple vowels said to begin and end them.

Advancement

Genuine diphthongs. Acoustically (in terms of F_2 values), the genuine diphthongs [ɛi] and [ɑu] appear to begin with a more fronted tongue position than do the simple vowels said to begin them ([ɛ] and [ɑ]) (Figure 2a). The second formant frequency of [ɑu] indicates that it ends with a slightly more fronted tongue position than does the simple vowel [u]. On the other hand, F_2 measurements imply that [ɛi] and [ʌy] end with slightly more retracted tongue positions than do [i] and [y]. That is, all the genuine diphthongs appear to be centralized at their endpoints, when considered in relation to the simple vowels [i, y, u].

The EMG activity (Figure 3) of the genioglossus and styloglossus supports the acoustically-based observation that the tongue is more fronted at the beginning of [ɛi] and [ɑu] than [ɛ] and [ɑ]: genioglossus activity is stronger for the early part of [ɛi] than it is for [ɛ], and styloglossus activity (which retracts the tongue) is weaker for [ɑu], especially in its earlier portion, than for [ɑ]. The EMG data also support the acoustically-based inferences about tongue position at the ends of these diphthongs: genioglossus activity is much weaker for [ɛi] than [i] and for [ʌy] than [y], implying less extreme fronting for the diphthongs. In parallel with this difference is the slightly weaker styloglossus activity for [ɑu] than for [u], implying slightly less tongue retraction (i.e., more fronting) for this genuine diphthong.

Pseudo diphthongs. The acoustic analyses, in particular the F_2 values, indicate that the first portion of each pseudo diphthong reaches the formant frequencies of the simple vowel said to begin it, but that the second portion of each falls short of its expected semivowel endpoint: the "fronting" diphthongs [aj, oj, uj] fail to reach the second formant frequency values of [i] (Figure 2b), and the "retracting" diphthongs [ew, iw] fail to reach those of [u] (Figure 2b).

Electromyographically, relative activity of the genioglossus and styloglossus for the early part of the fronting diphthongs [aj, oj, uj] is essentially the same as found for the simple vowels [a, o, u] (Figure 4a). The relative activity levels of these muscles for [ew, iw], on the other hand, might lead one to expect slightly less fronting than is inferred for [e] and [i], respectively (Figure 4b). The greater activity of the mylohyoid (which raises the tongue) at the beginnings of [ew] and [iw], than of [e] and [i], suggests that the genioglossus is devoted primarily to tongue advancement, although contributing secondarily to tongue raising.

All five of these diphthongs end "short" of the F_2 values for [i] or [u], (Figure 2b), and this, too, is reflected in the relative EMG activity level of the genioglossus and styloglossus muscles (Figures 4a & 4b).

Kakita *et al.* (1976) have observed that there is less genioglossus activity for [j] than for [i] and their X-ray data indicate that the tongue root is indeed less advanced for the semivowel. In our own data this more centralized tongue position for [j] may explain why there is less genioglossus activity for the offset of [aj] and [oj]. Of the fronting diphthongs only [uj] has genioglossus activity as strong as that for [i]; this activity is comparatively brief, though, and follows shortly after strong retracting action by the styloglossus. Among the retracting diphthongs, styloglossus activity is not nearly

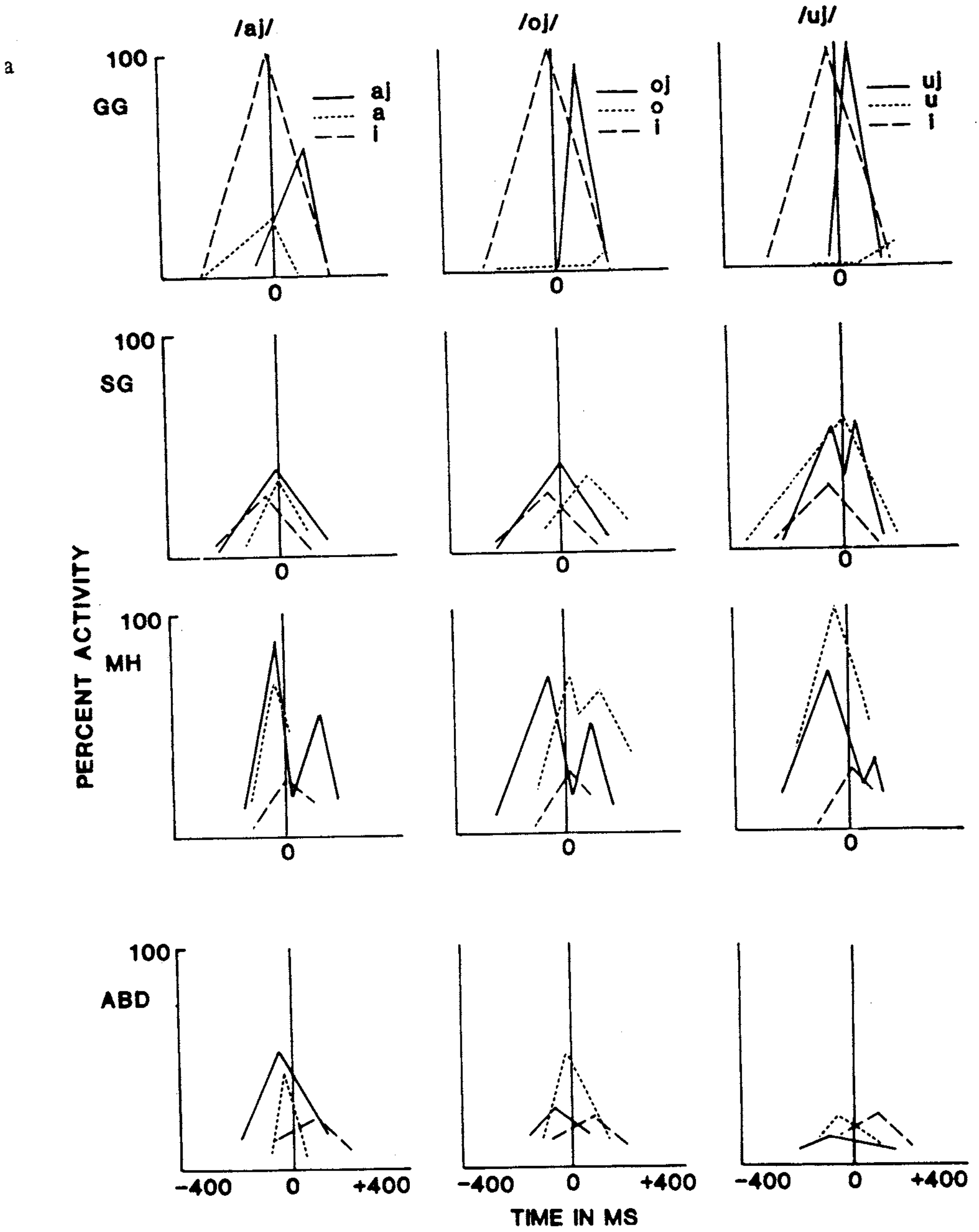
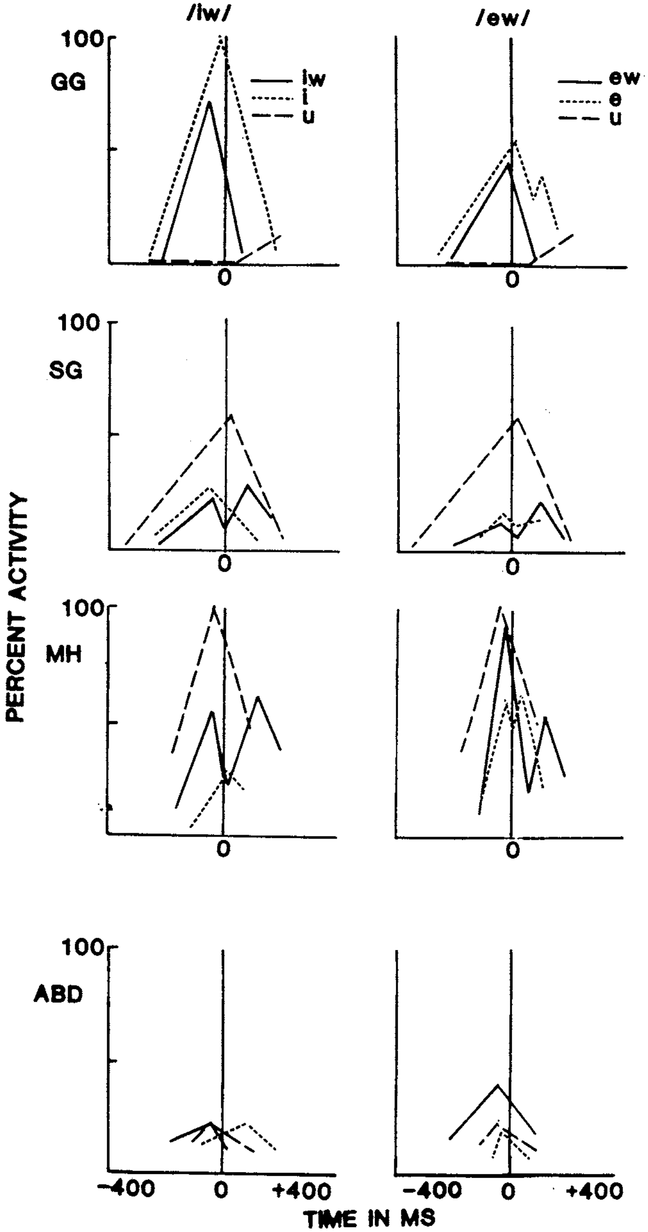


Fig. 4. EMG data for the pseudo diphthongs and simple vowels: [aj, oj, uj] in (a), [ew, iw] in (b). Each graph is a schematized representation of the time course of EMG activity in a given muscle, expressed as a percentage of the overall

b



range of that muscle's activity across utterance types. Zero on the abscissa represents the acoustic onsets of the diphthongs, the simple vowels said to begin them, and the vowels /i/ and /u/ that approximate the glides that end them.

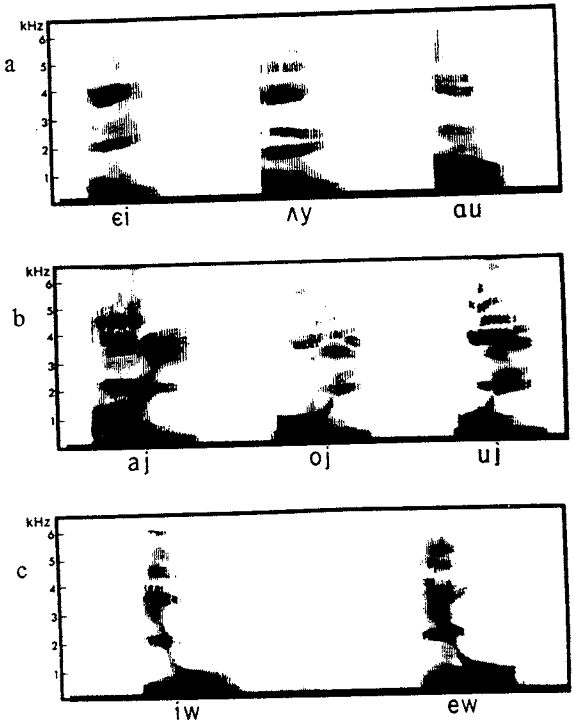


Fig. 5. Single tokens of each of the diphthongs studied, pronounced in isolation. Note difference in the rate of formant frequency change in genuine diphthongs (a) and pseudo diphthongs (b and c).

so strong as that found for [u]. That [iw] and [ew] probably end with a relatively retracted tongue position despite the low level of styloglossus activity at their offset may be due to the fact that the tongue has been strongly raised in their first part (for [i] and [e]), so that it requires less styloglossus action to pull the tongue back for their second part. In short, the EMG data suggest that all the pseudo diphthongs end more centrally than the vowels [i] and [u].

Finally, let us consider the observations made above, in so far as they relate to the basic distinction between diphthong types, viz. that the realization of fixed targets is essential for the pseudo, but not for the genuine diphthongs. We take this to mean that there should be acoustic and EMG differences between the patterns of the genuine diphthongs and those of the simple vowels which, at least in older phonetic transcriptions, are said to compose them. Further, such differences should not be found between the purported simple vowel components of the pseudo diphthongs and the pseudo diphthongs themselves.

Looking for these differences in the acoustic data for the various vowels, we find some support for this distinction between diphthong groups. As we have already seen, the averaged F_1 and F_2 values for the genuine diphthongs differ from those of their simple initial "components," whereas there is a very close correspondence between the F_1 and F_2 values of the pseudo diphthongs and their simple initial components (just before the abrupt change in second formant frequency).

Comparing the offsets of the pseudo diphthongs with their simple components yields data sets that are not strictly comparable, because these diphthongs end in semivowels, of which there are no other examples in our data. In Table 1, however, we have included the frequencies of the first two formants at the midpoints of the vowels [i] and [u] on the assumption that the semivowels [j] and [w], respectively, might well approximate these simple vowels acoustically. We find no exact matches and, in several instances, considerable discrepancies in formant values, particularly for the second formants. That is, the second formant frequencies of the pseudo diphthongs fall short of those of [i] and [u], suggesting that the diphthongs are more centralized than are these simple vowels. On the other hand, with the exception of [aj], the first formant values for four of the five pseudo diphthongs are equal to or smaller than those for [i] and [u], suggesting a degree of opening at least as small as that of the most closed vowels.

The acoustic data for the genuine diphthongs, on the other hand, suggest that they end with a more open and central articulation than [i, y, u], supporting the claim that the genuine diphthongs do not match the qualities of the simple vowels that conventional transcriptions suggest as their initial and terminal components. The pseudo diphthongs, in contrast, do match the qualities of the simple vowels that are said to initiate them, although the greatest acoustic similarities occur near the midpoints of the diphthongs and the simple vowels, and not at their onsets. Their offsets approximate the semivowels [j] and [w] rather closely in terms of openness, but tend to be more centralized.

With few exceptions, the EMG data support the inferences drawn from the acoustic data about the differences in starting and ending positions between the genuine and pseudo diphthongs. It is worth noting that the strong correlation between the acoustic and physiological data holds not only for rather gross differences between the two groups of diphthongs: Details of these data support the differentiation of the members of each diphthong class as well. For instance, the F_1 values at the end of the fronting pseudo diphthongs indicate an increasing degree of openness from [uj] to [oj] to [aj]. This gradation is reflected in decreasing levels of genioglossus activity associated with the semivowel. Also the formant values for the offset of [iw] suggest that this diphthong ends with a somewhat higher and more retracted tongue position than [ew]. This correlates with the more pronounced second peak of styloglossus and mylohyoid activity for the former.

These detailed correspondences between the acoustic and the physiological parameters lend support to our assumptions concerning the functions of the muscles studied.

Hypothesis 2: Harmony

The claim that there is harmony of tongue advancement for the genuine, but not necessarily for the pseudo, diphthongs is also substantiated by both acoustic and EMG data. The second formants of [ɛi], [ʌy], and [ɑu] display minimal changes in frequency, indicating an absence of extreme changes in tongue advancement (Figures 2a and 5a). In contrast, the second formants for [aj], [oj], [uj], [iw], and [ew] show dramatic frequency shifts, implying the presence of considerable horizontal tongue movement (Figure 2b).

The activity of the muscles responsible for tongue fronting (genioglossus) and backing (styloglossus) also indicates that there is less horizontal tongue movement for the genuine than for the pseudo diphthongs. The genioglossus is moderately active throughout [ɛi], while the styloglossus exerts almost no backward pull; for [ʌy] both muscles are relatively inactive, suggesting a predominance of vertical movement (which is positively indicated by mylohyoid and anterior-belly-of-the-digastric activity); and for [ɑu] the styloglossus is moderately active throughout, while the genioglossus is relatively inactive. In contrast, among the pseudo diphthongs we see patterns of activity in which the genioglossus and styloglossus muscles are alternately active. Thus, for [aj], [oj], and [uj], we find early peaks of styloglossus activity and late peaks of genioglossus activity, indicating fronting of the tongue from a backed position; for [iw] and [ew], we find the reverse sequence of genioglossus and styloglossus activity, indicating that the tongue is being retracted from a fronted position.

In summary, we find that our data support claims that distinctions between genuine and pseudo Dutch diphthongs include differences in harmony between the first and second elements with regard to tongue advancement.

Hypothesis 3: Single or Concatenated Gestures

Let us turn next to the description that maintains that a genuine diphthong is best characterized by a single articulatory gesture whereas a pseudo diphthong is best characterized as a sequence of two articulatory gestures. The EMG data suggest that there is a difference in the number of gestures for each of the two types of diphthongs. The data cited above, concerning the alternation of genioglossus and styloglossus activity for the pseudo diphthongs, are also relevant here. They depict articulations controlled by two muscles, acting successively first to retract and then to front the tongue ([aj], [oj], [uj]) or to front and then to retract the tongue ([iw], [ew]). The reciprocal timing in activity of these muscles reflects a sequence of opposing motor commands. Further, each pseudo diphthong is produced with two discrete peaks of mylohyoid activity (only in the case of [uj] is the second peak somewhat less pronounced). Each of these peaks is closely aligned in time with a peak of activity in either the genioglossus or the styloglossus muscle, suggesting that the mylohyoid muscle discretely supports the successive fronting and retracting tongue gestures.

In contrast, we would conclude from the EMG data that the genuine diphthongs are characterized as single gestures dominated by the activity of the genioglossus in the case of [ɛi] or by the styloglossus in the case of [ɑu], supported by mylohyoid activity. This supporting activity is less evidently "double peakéd" than with the pseudo diphthongs. In the case of [ʌy], where both muscles, as we have noted earlier, are relatively inactive and vertical movement predominates, the mylohyoid muscle displays a single peak of activity, suggesting, once again, a single articulatory gesture. Further research, using articulatory synthesis techniques, is needed to strengthen this hypothesis. Meanwhile, some support for it can be derived from the acoustic data.

The acoustic analysis reveals abrupt changes in second formant frequency of the pseudo diphthongs. For instance, over the first half of its duration the F_2 of [aj] shows a gradual rise in frequency of 250 Hz; over the second half of its duration the increase is 550 Hz, suggesting a rapid movement of the articulators. The analogous frequency changes for [oj] are 100 Hz and 800 Hz; for [uj], 100 Hz and 1200 Hz; for [iw], no change over its first half, and then a decrease of 1100 Hz; and for [ew], an increase of 300 Hz over its first half, and then a drop of 850 Hz. The genuine diphthongs show no such rapid shift in formant frequency in either half of their duration (Figure 5a). Acoustically, then, we do find support for the notion that the pseudo diphthongs are sequences of articulatory gestures.[7]

DISCUSSION

The articulatory data tend to support the acoustic and perceptual separation of the diphthongs into two groups. The genuine ones are characterized by a gradual increase in the activity of those muscles that either cause or support the smooth movement of the tongue in an upward and forward or backward direction. The pseudo diphthongs are characterized by a rather sharp increase in the activity of those muscles that either cause or support the abrupt movement of the tongue from a vowel into a semivowel in which the tongue moves horizontally across the vowel space. In other words, genuine diphthongs behave more like "unitary" segments, while pseudo diphthongs behave like sequences of two segments.

The observed articulatory differences cannot be explained by the difference in the distances an articulator must move between the beginning and the end of the diphthongal gesture in the two groups of diphthongs. Rather, we find that in [ɛi, ʌy, ɑu] tongue movement is primarily vertical, while in the pseudo diphthongs tongue movement is primarily horizontal.[8] In terms of "articulatory distance," therefore, the two classes are not necessarily very different. However, the "closing" gesture of the genuine diphthongs is achieved through synergistic action of the mylohyoid and genioglossus or styloglossus, whereas fronting or backing gestures of the pseudo diphthongs are achieved through the sequential antagonistic actions of the genioglossus and styloglossus.

---

[7] *This difference in the rate of change of the formant frequencies is perceptually less relevant than the correct timing of the onset of that change (Collier and 't Hart, in press).*

[8] *We should note that even in the case of [aj] our acoustic and EMG data indicate that [a] is articulated more similarly to back vowels, such as [ɑ], than to front vowels, such as [ɛ]. Indeed, the change in* F_2 *for [aj] is more than twice as great as the largest change in* F_2 *for the genuine dipthhongs. Thus, a more accurate transcription of our subject's version would be [ɑːj].*

This synergism versus antagonism is reflected in the differences in temporal pattern of formant frequency change between the two classes of diphthongs. In the genuine diphthongs formant frequency change is nearly continuous throughout the entire course of the diphthong; in the pseudo diphthongs a nearly stable initial portion of substantial duration is followed by a period of rapid formant frequency change (especially in F_2).

The contrastive muscle activity patterns associated with genuine and pseudo diphthongs lend support to Cohen's (1971, p. 228) proposal to treat [εi, ʌy, ɑu] as "unitary segments, requiring a feature specification of their own, rather than allow for this problem to be circumvented in a treatment which results in a phonetically arbitrary segmentation by assigning one part as dominated by a vocalic and a second one by a sonorant (i.e. non-vocalic, non-consonantal) feature." A biphonemic interpretation only seems plausible for the pseudo diphthongs.

REFERENCES

ALFONSO, P.J. and BAER, T. (1982). Dynamics of vowel production. *Language and Speech*, **25**, 151-173.

CHIBA, T. and KAJIYAMA, M. (1941). *The Vowel: Its Nature and Structure* (Tokyo).

COHEN, A. (1971). Diphthongs, mainly Dutch. In L. Hammerich, R. Jakobson and E. Zwirner (eds.), *Form and Substance* (Copenhagen), pp. 277-289.

COHEN, A., EBELING, C.L., ERINGA, P., FOKKEMA, K. and VAN HOLK, A.G.F. (1959). *Fonologie van het Nederlands en het Fries* ('s-Gravenhage).

COLLIER, R. and 'T HART, J. (in press). The perceptual relevance of formant trajectories in diphthongs. In M. van den Broecke and V. van Heuven (eds.), *Studies for A. Cohen* (Dordrecht).

DELATTRE, P. (1951). The physiological interpretation of,sound spectrograms. *Publications of the Modern Language Association of America*, **66**, 864-875.

FANT, C.G.M. (1970). *Acoustic Theory of Speech Production*, 2nd ed. (The Hague).

HARRIS, K.S. (1981). Electromyography as a technique for laryngeal investigation. *ASHA Reports No. 11*, 70-87.

'T HART, J. (1969). Fonetische steunpunten. *Nieuwe Taalgids*, **62**, 168-74.

HIROSE, H. (1971). Electromyography of the articulatory muscles: Current instrumentation and technique. *Haskins Laboratories Status Report on Speech Research*, **SR-25/26**, 73-86.

KAKITA, K. (1976). Activity of the genioglossus muscle during speech production: An electromyographic study. Doctoral dissertation, University of Tokyo.

KAKITA, K., HIROSE, H., USHIJIMA, U. and SAWASHIMA, M. (1976). A preliminary report on the electromyographic study of the production of the Japanese semivowel /j/. *Annual Bulletin of the Research Institute of Logopedics and Phoniatrics, University of Tokyo*, **10**, 37-46.

KEWLEY-PORT, D. (1973). Computer processing of EMG signals at Haskins Laboratories. *Haskins Laboratories Status Report on Speech Research*, **SR-33**, 173-183.

KOOPMANS-VAN BEINUM, F.J. (1969). Nog meer practische zekerheden. *Nieuwe Taalgids*, **62**, 245-250.

KUHN, G.M. (1975). On the front cavity resonance and its possible role in speech perception. *Journal of the Acoustical Society of America*, **58**, 428-433.

MIYAWAKI, K., HIROSE, H., USHIJIMA, T. and SAWASHIMA, M. (1975). A preliminary report on the electromyographic study of the activity of lingual muscles. *Annual Report of the Research Institute of Logopedics and Phoniatrics, University of Tokyo*, **9**, 91-106.

MOL, H. (1969). Fonetische steunpunten. *Nieuwe Taalgids*, **62**, 161-167.

POLS, L.C.W. (1977). Spectral analysis and identification of Dutch vowels in monosyllabic words. Doctoral dissertation, University of Amsterdam.

RAPHAEL, L.J. and BELL-BERTI, F. (1975). Tongue musculature and the feature of tension in English vowels. *Phonetica,* **32,** 61-73.

RAPHAEL, L.J., BELL-BERTI, F., COLLIER, R. and BAER, T. (1979). Tongue position in rounded and unrounded front vowel pairs. *Language and Speech,* **22,** 37-48.

SLIS, I.H. and VAN KATWIJK, A.F.W. (1963). Onderzoek naar de Nederlandse tweeklanken. Institute for Perception Research, Report no. 31 (Eindhoven).

SMITH, T.ST.J. (1971). A phonetic study of the function of the extrinsic tongue muscles. Doctoral dissertation, University of California, Los Angeles.

STEVENS, K.N. and HOUSE, A.S. (1955). Development of a quantitative description of vowel articulation. *Journal of the Acoustical Society of America,* **27,** 484-493.

STEVENS, K.N. and HOUSE, A.S. (1961). An acoustical theory of vowel production and some of its limitations. *Journal of Speech and Hearing Research,* **4,** 303-320.

VAN DEN BERG, B. (1959). *Foniek van het Nederlands* (The Hague).

ZONNEVELD, W. and TROMMELEN, M. (1980). Egg, Onion, Ouch! On the representation of Dutch diphthongs. In W. Zonneveld, F. van Coetsem and O.W. Robinson (eds.), *Studies in Dutch Phonology* (The Hague):