Phonological and Phonetic Views of Vowel Sequence Timing

—A Cross-Linguistic Review—

Martin Gore

(Received 21 October 2003)

Abstract. First language prosodic timing is a priority issue for TESOL teachers and phonologists alike (Celce-Murcia et al 1996; Cutler et al 1994). Vowel sequences are important because they reveal several crucial aspects of this timing in a microcosm. However, cross-linguistic comparisons need to be reassessed according to whether the focus is on micro-phonetic phenomena such as the transition between two adjacent vowels, or broader phonological indices such as vowel length or accent.

Conclusions hinge on a few central observations. Phonologists tend to think of phonetic nucleus as being characterized by phonological accent. However, the comparisons of Japanese and English vowel sequences found in the literature and tested by experimental studies (Gore 2003) show that phonetic nucleus and phonological accent do not necessarily go hand in hand. It seems possible for a phonological diphthong to influence accent placement in Japanese, while being phonetically indistinguishable from a sequence which has no effect on accent; and it also seems possible for a sequence not normally considered a phonological diphthong to show more diphthongal phonetic characteristics than one which does constitute a phonological diphthong.

1. Introduction

Vowel sequences known as "diphthongs" are said to exist in the syllables of many languages, such as English, French, German and Chinese. Whether such "diphthong syllables" may be said meaningfully to exist in Japanese, a mora-timed language in which the short vowel or CV is generally assumed to be the main unit of prosodic timing, is the subject of debate. Indeed, the very concept of "diphthong syllable"

begs the question in several important ways. It presupposes that both diphthongs and syllables (as opposed to moras or segments) have independently verifiable and definable roles in the language in question, either as phonetic phenomena or as phonological units or as both, and it also presupposes that diphthongs and syllables co-occur in the sound structure of the language in a readily definable, non-random way. It is all too easy to make these assumptions without due consideration as to whether they are warranted in the particular language under discussion.

Modern Japanese definitely has sequences of vowels. The mystery is how to describe them. The motivation behind the present investigation derives in the first instance from the discovery of unexpectedly large timing differences between Japanese and English vowel sequence transitions (Gore 2002), and from the expectation that unraveling the mystery of how to account for and describe these differences will not merely lead to a more accurate description of vowel sequences and their transition timing in Japanese and English, but will throw light on other broader phonological differences between English and Japanese, such as language-specific prosodic timing in general. It is hoped that this, in turn, will not only lead to better linguistic understanding, but also prove to be of practical benefit to English and Japanese language teachers in their attempts to unravel the various phonological influences of L1 over L2.

Though pedagogy is beyond the scope of this study, in language education, too, prosodic timing is increasingly becoming the focus of investigation. For instance, Celce-Murcia suggests that the first question the TESOL teacher should ask about pronunciation is: "Is the student's first language stress timed or syllable timed?" (1996, p324). If this line of questioning is considered important it must be even more so with regard to Japanese/English, since Japanese is said to be mora timed or "mora counting," thus providing a particularly stark contrast with English, which is said to show a tendency towards stress timing, at the opposite end of the prosodic timing continuum.

In tandem with this new (or renewed) pedagogical awareness of the importance of prosodic timing, Japanese-English cross-linguistic timing studies have seen something of a boom in recent years (e.g. Cutler and Ohtake 1994). However, prosodic timing has rarely been studied specifically from the point of view of vowel sequences as such, and conversely vowel sequences have rarely been studied cross-linguistically from the point of view of timing. Several studies investigating the ratios of long and short Japanese vowels, for example Hoequist (1983) and Campbell (1992), have

presented material showing that long vowels are generally not long enough to constitute two entire moras, but these studies have not targeted diphthongs as such. Thorough investigations of diphthongs do exist, such as the pioneering work of Lehiste and Peterson (1961), Holbrook and Fairbanks (1962), Gay (1968) and Maddieson and Emmorey (1985), but these have not targeted diphthongs in relation to language-specific timing. Thus the possible variations between languages in vocalic inter-segmental transition timing and its possible relationship to language-specific prosodic timing have received little attention. One prominent exception is the study of cross-linguistic differences in diphthongs by Lindau, Norlin and Svantesson (1990), building on work by Gay (1968), Kent and Moll (1972) and Maddieson (1984) and comparing Chinese, English and Hausa. They conclude that transition duration is a language-specific trait, taking up 60% of English [au], 50% of Chinese and only 20% of Arabic. These findings are echoed in English/Japanese comparisons (Gore 2002). However, such studies are rare.

This is in contrast to the large number of detailed phonetic studies that have been made comparing vocalic coarticulation in VCV sequences (Ohman 1966, Butcher and Weiher 1976, Recasens 1984, Farnetani et al 1985, Magen 1997, Manuel 1999 etc). The assumption seems to be in most quarters that the characteristics of adjacent vowel-to-vowel movements are relatively invariable across languages, or at least show so little variation as to be unremarkable or unimportant when considered against the broad backdrop of phonological theory and phonetic enquiry. What, after all, can a small fluctuation in a diphthong tell us about the sound structure of the language as a whole? The present investigation offers the answer: more than one might expect. Indeed, since movements from one vowel to another are liable to be influenced by many disparate aspects of phonology such as prosodic timing, vowel distribution, stress, tone and syllable structure, it may be argued that diphthongs encapsulate (or should be capable of revealing) these aspects in a microcosm. However, since it is impossible to focus on all of these aspects at once, the present study will make a start by dealing primarily with aspects related to timing, though other facets of phonology will also be taken into account.

2. Mora Timing in Japanese

There is considerable evidence and a broad consensus of opinion, though not total agreement, that Japanese, or at least standard Japanese, is fundamentally a "mora-

timed," "mora-based" or "mora-counting" language (Bloch 1942, Han 1962, McCawley 1968, Hoequist 1983, Haraguchi 1996, Kubozono 2001). Accordingly, "the duration of a word stays very close to a target duration that depends on the number of moras in it," and "words with increasing numbers of moras increase in duration by nearly constant increments" (Port et al 1986). According to this view, since every V or CV constitutes a mora in itself, a sequence of two vowels is to be analyzed as two moras, regardless of phonetic quality and regardless of whether the sequence derives etymologically from a diphthong (of, say, Chinese or English origin) or from a native VCV where the intervening consonant has become elided over the years.

Both Kubozono (2001) and Warner and Arai (2001a) offer comprehensive overviews of mora timing, its limitations, and the role of other timing units such as the syllable and the foot (Poser 1990). Warner and Arai note that the correlation between word duration and number of moras might be expected anyway since all moras have segmental content, and that there is no consistent criterion for how linear the correlation must be, and they concur with Beckman (1982) that the negative correlation between adjacent segments sometimes cited as proof of "compensation" to preserve mora timing will also appear if there is any measurement error in placing boundaries. They conclude that the restricted syllable structure, the prevalence of light syllables, the lack of pitch effect on duration, the variability in the number of syllables that can occur between accents, the frequent lack of any accent, the lack of quality difference with durational distinction, and the lack of allophony conditioned by accent are typical phonological characteristics of Japanese that may be related to (or give the cumulative impression of) mora-timing. They also agree with Cutler and Ohtake (1994, 1999) that the mora is thus "not an indivisible unit of speech perception," but it may help the listener to "locate boundaries, and thus to choose between possible parses of the signal." Warner and Arai (2001b) present additional intriguing results that "segments are, by one count, statistically equivalent to moras in the strength of their relationship to word duration." These results confirm Beckman's view (1992 p458) that "mora-timing" should not be taken literally, because "mora durations in Japanese differ as a function of the number of component phonemes" much as "interstress intervals in English vary as a direct function of the number of constituent syllables." (See Beckman 1982; Lehiste, 1977.)

McCawley (1968), Vance (1987), Haraguchi (1991, 1996) and Kubozono (2001) describe in detail the conditions whereby syllables, rather than moras, may become

the basis for accent placement. It is clear that a final nasal, the first consonant of a geminate and the second half of a long vowel cannot bear an accent for obvious articulatory reasons. On the other hand, it is less clear which dual-vowel syllables should be considered as "diphthongs" (where only the first mora is said to bear an accent or influence accent placement) and which should be considered as a juxtaposition of two light monophthongs each capable of bearing or influencing accent in its own right. Are articulatory factors at work here, too? In any case, determining "diphthong" in terms of "heavy syllable" smacks of circularity and lacks explanatory force (see Section 4).

3. Phonetic aspects of English/Japanese vowel sequences

From the purely phonetic viewpoint, recent studies of common Japanese vowel sequences using spectrographs and synthesis (Gore 2002) demonstrate that Japanese VVs do not typically have the gliding phonetic characteristics of English diphthongs but rather the non-gliding or "jumping" quality of two juxtaposed monophthongs. Such non-gliding juxtaposed monophthongs can also be seen occasionally in English, particularly at word boundaries such as *draw even,* which never approaches a diphthongal realization even when pronounced at great speed, in loan words such as *piano* or *Zaire* and occasionally at morpheme boundaries such as in *drawing, pawing* etc., though in these cases there is usually some gliding or, perhaps more frequently, rhoticization. In Japanese, however, the dual monophthongal quality is usually of great clarity and is heard even within a single morpheme, indeed even within a single syllable. (The nature and relevance of the syllable needs to be further defined since its importance vis-à-vis the diphthong cannot be ignored; this, however, is a far from simple process, especially in the cross-linguistic context.)

There thus seems to be little or no *phonetic* difference between those Japanese vowel sequences that are said to be "diphthongs" and those that occur between morpheme or word boundaries and so presumably are not qualified to be called "diphthongs" in the conventional sense of the term (defined later in this review). To English ears, the effect is rather as if every monophthongal mora were its own morpheme (though, of course, few language users are conscious of morphemes as they actually use them) and, indeed, a fair proportion of native Japanese morphemes are only one mora in length. Recent studies have shown that Japanese VVs exhibit an almost total absence of the sort of gliding the English speaker might expect to

find between adjacent vowels (Gore 1996, 2002), and yet in the literature it is often suggested that certain adjacent vowels somehow constitute a "syllable" or a "diphthong" or a "diphthong syllable." How are we to make sense of this apparent contradiction?

4. The conundrum of Japanese vowel sequences

Though many present-day Japanese phonologists and phoneticians refer to "diphthongs" and "heavy syllables" in the course of their theorizing (Haraguchi 1996, Kubozono 2001, Kawahara 2002), the actual physical nature of the sound units referred to is rarely made explicit. But without independently and physically verifiable measurements we are left with terminology which sometimes seems contradictory or circular. Whether such concepts as "diphthong" have any explanatory force, i.e., can be applied to real language without leading to internal contradiction or tautology, surely depends on definitions that make reference to physical reality. To give an instance of the sort of circularity frequently encountered in the literature, there are cases where the putative "syllable" in Japanese is defined and justified in terms of the existence of "diphthongs," while the "diphthong" in its turn is defined in terms of two vowels occupying the same "syllable." Yet, at the same time, it is usually asserted that the basic unit of timing is the mora. No reliable indication has yet been offered of why some bi-moraic VV syllables are held to be heavy syllables, i.e. phonological "diphthongs," while others are not. In such arguments, exact physical measurements and their linguistic implications are rarely mentioned. This lack of physical evidence leaves the reader wondering whether the phenomenon being discussed actually exists -- whether it is not, perhaps, a mere figment of the linguist's imagination. Churchill's "riddle wrapped in a mystery inside an enigma" applies well to the Japanese diphthong.

Much of the mystery concerning Japanese vowel sequences seems to hang on whether the term "diphthong" refers to a phonetic phenomenon such as relative speed of transition between the two adjacent vowels, or to a phonological unit like vowel length or accent, which may or may not have wider prosodic ramifications, or to some unquantified mixture of phonetic and phonological characteristics. Until more success is achieved in welding phonological theory to actual physical sound measurements, it would thus seem difficult, in view of the conceptual haze at the present time, to justify the use of the word "diphthong" with regard to Japanese. It is becoming increasingly obvious that both the segmental and supra-segmental aspects of Japanese

vowel sequences need to be investigated in greater detail and defined with greater clarity. The present investigation seeks to address these issues by suggesting practical phonetic and phonological approaches, making, where appropriate, comparisons with vowel sequences and diphthongs in other languages, particularly English.

5. Orthographic influences

To compound the confusion concerning Japanese vowel sequences, there is also at least a possibility of some degree of interaction between orthography and phonology, since of the two contrastive orthographic systems that are basic to modern Japanese, *kanji* allow the representation of a diphthong within a putative syllable (as in Chinese), whereas *kana* (though originally derived from *kanji*) are inherently short-vowel moraic and thus, even if diphthongs actually exist in some form, do not allow their orthographic representation as a single unit. As the historian, W.G. Aston succinctly observed over a hundred years ago in 1899:

> Chinese... has... diphthongs, combinations of consonants and final consonants, none of which are to be found in Japanese, where every syllable consists of a single vowel or of a single consonant followed by a single vowel.... The Japanese, in adopting Chinese vocables, modify them to suit their own phonetic system. But the process of assimilation is incomplete. The two elements harmonize no better than brick and stone in the same building.

In other words, in modern Japanese, the Chinese-derived "stone" syllable is (it seems) measured on a phonologically different scale from the indigenous "brick" or mora. It is expected that a close phonetic analysis of Japanese vowel sequences, especially with regard to the question of whether diphthongs do or do not exist, will illuminate this prosodic mishmash.

Since many Japanese words (but not all) can be written in either *kanji* or *kana,* the effects of orthography on pronunciation may be more complicated than in mono-orthographic languages. Beckman has also pointed out the effects of historical/orthographic conventions in relation to mora timing (1982), but not in direct connection with diphthongs. When it comes to the question of whether diphthongs actually exist in Japanese, this multi-layered orthography may indeed underlie the reality (if diphthongs do exist), or alternatively the misapprehension (if they do not exist). Thus, the possibility of some degree of interaction between orthography and phonology should not be ruled out, and needs to be tested carefully, if only to be disproved. This

can be done by analyzing the effects of very early *kana* education on the segmentation strategies employed by pre-school and first grade children in word games. Similar approaches have already been reported by Muto (1987), Hatano and Inagaki (1992) and Morais et al (1996). Preliminary trials (Gore 2004a, in preparation) confirm and clarify the results of the aforementioned reports (1987, 1992) that the acquisition of *kana* orthography does indeed seem to have an effect on children's segmentation strategies. Another possible approach to the question of orthographic influence on diphthong timing independent of child F1 acquisition is to test the relative vowel-to-vowel transition speeds and total sequence durations when *kanji* and *kana* are read aloud by Japanese adults (Gore 2004b, in preparation).

6. Tone accent

Many researchers imply that Japanese "diphthongs" can be defined adequately in terms of High-Low tone groups (Haraguchi 1996, Kubozono 2001, Kawahara 2002). Indeed, discussion of the lexical importance of High and Low tones has for a long time been one of the most popular general areas of linguistic investigation by Japanese scholars of their own language. (In the early history of the language, the use of tones in Japanese may have been affected both by Chinese tones as such, and also by accent realignment conditioned by the assimilation of Chinese syllables into a basically moraic phonological system, but such historical considerations are beyond the scope of the present study.) However, in many situations in modern Japanese, the very tones upon which much present-day phonological theory is based are effectively neutralized (for instance, in long phrases) or even reversed (in some compounds). Tone accents also vary a great deal depending on the particular dialect of the speaker, and there are several well-documented non-tonal or "nonaccentual" (Haraguchi 1988) dialects (as in Kagoshima and other areas of Kyushu) where tone is comparatively fluid and is thought to have little or no lexical relevance. The important point here is that it has not been made clear in the literature hitherto whether the supposed "diphthong" somehow becomes a "non-diphthong" when the tone-accent system on which it is assumed to depend is, for any of these reasons, changed or neutralized.

However, assuming that the "diphthong" does exist as a meaningful entity in Japanese, it does seem that vowel sequences that might be described as "diphthongs" in certain dialects (including the standard Tokyo dialect) do occur with accent placement characteristics that are perhaps best explained in terms of phonological syllable

weight (Haraguchi 1996, Kubozono 2001), though precisely when and why this should occur is not clear, and there is no evidence yet that they are phonetically distinguishable in themselves from sequences which have no effect on accent.

7. Preferred vowel juxtapositions in Japanese

When considering the status of diphthongs in a language such as Japanese where virtually any juxtaposition of monophthongs is possible, it may be relevant to consider whether there are any significant patterns of vowel combination, regardless of whether any such combinations can be considered diphthongs from other phonetic or phonological viewpoints.

Any juxtaposition of vowels seems to be possible in Japanese words, though some, such as [eo] and [ea], seem to be of rather more limited occurrence than others. In the Sino-Japanese lexicon, however, same-syllable different-vowel sequences are mainly restricted to [ai], [ei] (with /ee/ allophone), [ui], and [ou] (with /oo/ allophone), all of which occur with an extraordinarily high frequency. This, in itself, creates a stark distributional imbalance between the two lexicons, Sino-Japanese on the one hand, and the "native" Yamato-Japanese on the other.

The most intriguing source of imbalance is that [au], a common combination in both Chinese and Japanese, does not occur in the Sino-Japanese lexicon at all. This extraordinary fact has rarely been noted in the literature. Why this imbalance should have come about is beyond the scope of the present short review, but the fact of [au]-absence in the modern Sino-Japanese lexicon may have a bearing on the diphthong debate, and must thus therefore be born in mind.

One possible phonetic ramification may be that the [a] of [ai] (especially in Sino-Japanese) may sometimes be closer to [i], than that of [au], [ae] and [ao], which only occur in native Japanese (Keating & Huffman 1984). If such a coarticulaory effect occurs predominantly on one pair of vowels, such as [ai], this may be an argument in favor of considering that pair to be a diphthong. This phonetic effect may spill over from Sino-Japanese usage into [ai] sequences in native Japanese words. Experiments must be done to test these possibilities.

8. Allophonic diphthongs, phonological diphthongs and phonetic diphthongization

Some definitions of "diphthong" in Japanese have recourse to whether a particular vowel sequence has a single vowel or long vowel allophone (analogous

to the diphthong/vowel allophony in the second syllable of English *again*), but such allophones, though common, are optional rather than predominant in Japanese and do not exist in all dialects. Nevertheless, this type of definition, though of rather limited application and slightly different from English ideas of what might constitute the definition of a diphthong (discussed in Section 9), has the advantage of being at least readily understandable and of having few internal contradictions, and it is probably for these reasons that it forms the basis of many definitions of "diphthong" in Japanese dictionaries such as the Kokugogaku Jiten.

In adjectives ending in *-ai* or *-oi*, the allophone *-ee* (or *-e*) (Kubozono 2001, Kawahara 2002) may be used in some dialects including standard Tokyo Japanese. The final *-i* may be a separate morpheme, historically *-shi* or *-ki* in many cases e.g. *na.ki* (historical) – *na.i* – *nai* – *nee* – *ne*, but even so this reduction or allophony has lead many Japanese phonologists to refer to [ai] and [oi] in this position as diphthongs. Whether the fact that the second vowel, *-i*, forms a separate morpheme has any influence on the phonetic realization of the VV sequence, as would be expected in English (see below), can be tested by spectrographically comparing within-morpheme versus across-boundary sequences. Word-boundary tests should also be conducted.

High-speed reiteration is another experiment that might reveal phonetic diphthongal phenomena in Japanese. In spite of the fact that there are fewer articulatory restrictions on speed in vowel sequences, it has been shown that mora timing is usually maintained to a fairly high degree, except at the very highest of speeds where the articulatory restrictions on CV moras become so great that the sense of isochrony with single V moras seems to break down. If this breakdown point is found to come earlier in some VV combinations than in others, this would be an argument in favor of the existence of phonetic diphthongization in certain contexts. To what degree such "phonetic diphthongization" co-occurs with putative "phonological diphthongs" is yet another question. The discussion so far and the results of previous experiments would suggest that phonetic diphthongization in Japanese is rare, and when it does occur it is not highly correlated with the presence of phonological diphthongs. It is this aspect of Japanese that contrasts most starkly with our phonological and phonetic expectations of diphthongs in English.

9. Some definitions of diphthong in English

With regard to English Vowel sequences, Lehiste and Peterson (1961) and Holbrook

and Fairbanks (1962) did much groundbreaking work. Lehiste and Peterson defined the diphthong as "a vocalic syllable nucleus containing two target positions" and showed that (especially in [aɪ], [aʊ] and [ɔɪ]) "the transitions between targets are longer than either target position." Holbrook and Fairbanks (1962) gave detailed measurements of GA at five points along the course of each diphthong, showing the formants to be continually and gradually changing. Kent (1992) also refers to diphthongs as "dynamic sounds in which the articulatory shape (and hence formant pattern) slowly changes during the sound's production."

Abercrombie (1967) defines diphthong as "a vowel of continually changing quality" but remarks, "No harm is done by thinking of a diphthong as a sequence of two vowels, provided that they occupy only one syllable. A sequence of two vowels which occupy two syllables, as in *gnawing* or *ruin,* is not a diphthong."

Abercrombie's examples of non-diphthongs are intriguing, and may lead us to a closer understanding of why certain vowel sequences are held to be diphthongs. The reason why *gnawing* is not a diphthong is surely different from the reason why *ruin* is not a diphthong. *Gnawing* comprises two separate morphemes, though the relevance of the morpheme boundary is not without controversy, as we have seen in the case of Japanese final *-i.* In any case, the fact that the single morpheme *ruin* has not become a diphthong would seem to be due more to a combination of distributional and articulatory factors, rather than any morphemic boundary factor.

Catford, 1988, (p115) defines diphthong as a:

> ... sequence of different vowels within one and the same syllable... English examples are [aɪ] in *high,* [aʊ] in *how,* [ɔɪ] in *boy* and in very many types of English [eɪ] in *day,* [oʊ] or [aʊ] in *go,* etc. In a diphthong, the two vowels, the starting point (or 'first element') and the finishing point ('second element'), such as [a] and [ɪ] in *high,* are not perceived as two separate vowels, but the diphthong is perceived as a transitional gliding sound starting at the first element and gliding towards the second element. A diphthong... occurs within a single syllable... performed with a single stress-pulse or pulse of initiator power.

From this definition, four concepts that are crucial to defining the English diphthong can be gleaned: occurrence within a single syllable; perception as a single sound; performance with a single stress-pulse; and presence of transitional gliding. It is at this point that several questions begin to arise: What is a syllable? What is a single sound? What is a stress-pulse? What is a glide? What if only some of the necessary conditions are met?

Catford's concept of syllable is not defined until much later (p179), where he arrives at "a definition of the *syllable as a minimal pulse of initiatory activity* bounded by a momentary retardation of the initiator, either self-imposed, or, more usually, imposed by a consonantal type of articulatory stricture." However, in the very next section, entitled "The Foot" (p180), Catford adds: "initiatory activity seems to be parcelled out into relatively equal chunks that are often longer than the length of a syllable." This section concludes with the remark, "Not all languages are like English. In French and Japanese for example, the parcelling out of initiator power is done differently."

It is easy to see how a "single stress-pulse" would *allow* a "glide" within a single unit; but the two parts of Catford's definition (the stress-pulse and the gliding) are surely not dependent on each other; indeed, a comparison of Japanese and English spectrograms of diphthongs in the respective languages (Gore 2002) shows that they are essentially independent qualities. In other words, whether a glide between elements occurs or not seems to have no inherent connection with stress-pulse. To be more specific: Japanese vowel sequences *within* a putative syllable seem to show the same lack of gliding as those that are plainly not of the same syllable (though more intensive and extensive spectrogram analysis must be done to clarify this under various conditions). Strictly speaking, since a total absence of formant gliding is virtually impossible, what we are really saying here is that the *timing* of the glide (slow or rapid) has no inherent physical connection with the *dynamism* of the stress-pulse. Phrased in this way, we can see that there is an alternative possible connection which may have eluded us until now: that the timing of the formant transition may be influenced by the dominant mode of prosodic timing of the language as a whole, while being relatively unaffected by contextual variations in Syllable/Mora stress placement; in other words, stress-timed languages may tend to have a slow glide as the default timing, whereas mora-based languages may tend to have a relatively fast transition time between vowels regardless of whether they constitute phonological groupings such as diphthongs or not. It is thus clear from the discussion so far, particularly in view of the comparisons that have been made between Japanese and English, that neither phonological nor phonetic characteristics alone are adequate to define the concept of diphthong.

In Fundamental Problems in Phonetics (1977), Catford presented a view which is intriguing from the point of view of English/Japanese cross-linguistic timing studies:

In point of fact, a diphthong may, indeed, consist of two distinct, discrete 'elements', with a relatively rapid transition between them. On the *other* hand, it may be more correctly characterized as a continuous, gliding movement from a starting point to a finishing point. There are thus two extreme types of diphthong, a 'sequential' type, which may be represented diagrammatically by _⌐ and a 'gliding' type, represented by _╱ with, of course, a continuum of possible gradations between these extremes.

Catford goes on to write: "In many types of English the 'narrower' diphthongs, such as the [eɪ] of *day* or the [əʊ] of *go,* are more often than not of the gliding type, while the 'wider' diphthongs, such as the [aɪ] of *buy* or the [ɔɪ] of *boy,* are of the more distinctly segmented sequential type."

This division of diphthongs into two types, "gliding" and "sequential," is of some interest (though not uncontroversial) in English, but is illuminating when applied to the cross-linguistic context, as Catford proceeds to suggest.

The shorter the duration of the 'non-syllabic' element and the more sudden the transition to or from it, the more like a consonantal semivowel it is. Thus, in principle, the difference between English *pie* and the French *paille* lies in the rapidity of transition and the duration of the second element: English *pie* _╱ or _⌐ [paɪ], and the French *paille* _╱^ [paːj].

This is still in accordance with Abercrombie. It is interesting to note that O'Connor (1991) presents a slightly different view:

A diphthong... is a vowel glide or a sequence of two vowel segments which functions as a single phoneme. Not all such sequences or glides are to be considered as diphthongs. In French, for example, the sequences [ai] and [aːi] occur in *hair* ('to hate') and *paille* ('straw') but neither constitutes a diphthong because neither occurs in the variety of contexts which other vowel phonemes of the language can occupy; so [ai] is taken to be a sequence of vowels /a/+/i/, just as /ʊɪ/ is in English *suet,* and [aːi] is interpreted as vowel+consonant, /a/+/j/...

Not only is O'Connor's view of *paille* stricter than Catford's, but the interpretation put on the final /ʊɪ/ example also seems to contrast with Catford who allows /ʊɪ/ as one of the 'Major Diphthongs' in American English, as in *buoy* (1988. p215), though many Americans do not use this pronunciation.

Now, English *suet* is usually thought of as having two syllables. But if *paille* is one syllable, then the reason for the sudden transition between these vowels would seem to have little to do with prosodic timing unless the syllable were considered

to be segmentable, as in the Sino-Japanese lexicon, into moras which reflect some counting measure without necessarily affecting or being affected by dynamic factors such as "initiator activity," "nucleus" or "accent."

10. Conclusion

We can now see four key concepts that define the English diphthong: the single syllable, the perception as a single sound, the gliding, and the "occurrence in the variety of contexts which other vowel phonemes of the language can occupy." However, each of these concepts has, at best, severely limited application to the Japanese mora-based phonology, though this is not to say that *syllables* do not exist in Japanese as prosodic units. What we can say about Japanese vowel sequences in connection with syllable structure is that though "phonological diphthongs" may be considered to exist and constitute a single syllable (as in English), they are not necessarily phonetically distinguishable from juxtaposed monophthongs; whereas in English one of the defining qualities of a diphthong is that it is *phonetically* quite a different thing from the other vowel phonemes and yet occurs in a similar variety of contexts.

A rudimentary graphic summary of some of the main cross-linguistic comparisons can be made as follows, using Catford's convenient short-hand to indicate speed of formant movement:

English [aɪ]	French [aːj]	Japanese [ai]
Slow glide		Fast transition
⟋	⎯⟋ᐱ	⎽⌐ (phonetically like 2 monophthongs)
Single pulse	Double count; but may form single "heavy" syllable	
Schwa audible	Central Vs absent due to "jump" over central area	

A clear view of these differences between Japanese and English diphthongal vowel sequences and the differences in their intersegmental timings is surely an essential part of a full understanding of the language-specific prosodic timing of English on the one hand and Japanese on the other.

The definitions of English diphthongs reviewed in Section 9 seem to be relatively uncontroversial, probably owing to the high degree of agreement between phonological and phonetic findings. However, the Japanese definitions reviewed in Sections 3~8 give the general impression of being ad hoc, inconsistent and unrelated. This is not

to say that diphthongs do not exist in Japanese, but rather that our conceptual equipment is not quite in tune with the phenomena that we observe. If we expect to find a diphthong with English characteristics then we are looking in the wrong place. How we relate disparate but interleaving definitions based on etymology, orthography, tone, accent placement, prosodic timing and transition speed to each other in a coherent system is the major problem we still face. McCawley's view of Japanese as a "mora-counting but syllable-accenting" language (1968) (broadly followed by Haraguchi and Kubozono) may still give us our best approach to the Japanese diphthong, but Beckman's orthographical cautions (1982) and Warner and Arai's segmental findings (2001) are worthy of investigation. Much experimental work still needs to be done to underpin, and perhaps even unify, the various theoretical approaches to vowel sequences in all their forms.

References

Abercrombie, D (1967). *Elements of General Phonetics*. Edinburgh University Press.

Aston, W.G. (1899). *The History of Japanese Literature*. Book 2 (Nara Period) Chapter 2. Kelly and Walsh. Yokohama. http://www.lily.sannet.ne.jp/s_noriko/Aston_02.htm

Beckman, M. (1982). Segment duration and the 'mora' in Japanese. *Phonetica* 39: 113-135.

Beckman, M. (1992). Evidence for speech rhythms across languages. Tohkura, Varikiotis-Bateson, Sagisaka eds., *Speech perception, production, and linguistic structure*, pp458-463. Ohmsha, Tokyo.

Bloch, B. (1942; 1950). Studies in Colloquial Japanese IV: Phonemics. *Language* 26: 86-125.

Berg, T. (1985). The Monophonematic Status of Diphthongs Revisited. *Phonetica* 42: 198-205.

Butcher, A., & Weiher, E. (1976). An electropalatographic investigation of coarticulation in VCV sequences. *Journal of Phonetics* 4, 59-74.

Campbell, N. (1992). Segmental elasticity and timing in J speech. Tohkura, Varikiotis-Bateson, Sagisaka eds., *Speech Perception, Production and Linguistic Structure*. ATR Auditory and Visual Perception Research. Tokyo: Ohmsha.

Catford, J.C. (1977). *Fundamental Problems in Phonetics*. Edinburgh University Press.

Catford, J.C. (1988). *A Practical Introduction to Phonetics*. Oxford University Press.

Celce-Murcia, M., Brinton, D. & Goodwin, J.M. (1996). *Teaching Pronunciation*. Cambridge University Press.

Cutler, A. and Otake, T. (1994). Mora or phoneme? Further evidence for language-specific listening. *Journal of Memory and Language* 33: 824-844.

Farnetani, E., Vagges K. & Magno-Caldognetto, E. (1985). Coarticulation in Italian /VtV/ sequences: a palatographic study. *Phonetica* 42: 78-99.

Gay (1968) Effect of speaking rate on diphthong formant movements. *Journal of the Acoustical Society of America* 44 (6), 1570-1575.

Gore, M. (1996). Formant frequency characteristics of English central vowels acquired by Japanese university students. *Collected Articles on the English Language* 30, Tokyo.

Gore, M. (1998). Phonetic analysis of Japanese university students' pronunciation of English

vowels. *Bulletin of the Faculty of Education, Kagoshima University* 49.

Gore, M. (2002). Vowel-to-vowel coarticulation in Japanese and English diphthongs. *Proceedings of the 7th Annual Conference of the English Phonetic Society of Japan.*

Gore, M. (2003). Language-specific prosodic timing and intersegmental transition speed. *Bulletin of the Faculty of Education, Kagoshima University* 54.

Han, M.S. (1962). *Japanese phonology: an analysis based upon sound spectrograms.* Tokyo: Kenkyusha.

Haraguchi, S. (1988). Pitch accent and intonation in Japanese. In Hulst, H.v.D. & Smith, N. (eds.), Autosegmental Studies in Pitch Accent. Foris Publications. Dordrecht, Holland.

Haraguchi, S. (1991). *A Theory of Stress and Accent.* Foris Publications. Dordrecht, Holland.

Haraguchi, S. (1996). Syllable, mora and accent. In Otake, T. & Cutler, A. (eds.), Speech Research 12. *Phonological Structure and Language Processing.* Cross-linguistic Studies. Mouton de Gruyter.

Hardcastle, W.J. & Hewlett, N., (eds.) (1999). *Coarticulation: theory, data and techniques.* Cambridge Studies in Speech Science and Communication. Cambridge: Cambridge University Press.

Hatano, G., & Inagaki, K. (1992) *Phonological awareness in Japanese children.* Annual Convention, Japanese Association of Psychology.

Hoequist, C. (1983). Syllable duration in stress-, syllable- and mora-timed languages. *Phonetica* 40:203-237.

Holbrook, A. & Fairbanks, G. (1962). Diphthong formants and their movements, *Journal of Speech and Hearing Research* 5-1. In Lehiste, I. (1967), Readings in Acoustic Phonetics. Cambridge MA: MITPress.

Keating. P.A. & Huffman, M.K. (1984). Vowel Variation in Japanese. *Phonetica* 41: 191-207.

Kent, R. D. (1992). *The Acoustic Analysis of Speech.* San Diego: Singular Publishing Group.

Kent, R.D. & Moll, K.L. (1972). Tongue body articulation during vowel and diphthong gestures. *Folia Phoniatrica* 24: 278-300.

Kubozono, H. (2001). *Mora and Syllable.* Kobe University.

Lehiste, I. & Peterson, G.E. (1961). Transitions, glides and diphthongs. *Journal of the Acoustical Society of America* 33 (3): 268-277. In Lehiste, I. (1967), Readings in Acoustic Phonetics. Cambridge MA: MITPress.

Lehiste, I., (ed.) (1967). *Readings in Acoustic Phonetics.* Cambridge MA: MITPress.

Lehiste, I., (ed.) (1977). Isochrony reconsidered. *Journal of Phonetics* 5: 253-263.

Lindau, M., Norlin, K. & Svantesson, J. (1990). Some cross-linguistic differences in diphthongs. *Journal of the International Phonetic Association* 20:1.

Kawahara, S. (2002). *Similarity among Variants: Output-Variant Correspondence.* Thesis at International Christian University. http://www-unix.oit.umass.edu/~kawahara/ov_faith.pdf.

Kokugogaku Jiten (Dictionary of Japanese Linguistics). (1955). Kokugogakkai. Tokyodo Shuppan.

Maddieson, I. (1984). *Patterns of Sounds.* Cambridge: Cambridge University Press.

Maddieson, I. & Emmorey, K. (1985). Relationship between semivowels and vowels: cross-linguistic investigations of acoustic difference and coarticulation. *Phonetica* 42: 163-174.

Magen, H.S. (1997). The extent of vowel-to-vowel coarticulation in English. *Journal of Phonetics* 25: 187-205.

Manuel, S. (1999). Cross-Language Studies: relating language-particular coarticulation patterns to other language-particular facts. In Hardcastle, W.J. & Hewlett, N. (eds.), *Coarticulation: theory, data and techniques. Cambridge* University Press. 179-199.

McCawley, J.D. (1968) *The phonological component of a grammar of Japanese.* Mouton, The Hague.

Morais, J., Kolinsky, R. & Nakamura, M. (1996). The psychological reality of speech units in Japanese. In Otake, T. & Cutler, A. (eds.), Speech Research 12. *Phonological Structure and Language Processing.* Cross-linguistic Studies. Mouton de Gruyter.

Muto, Y. (1987). *Relationship between kana reading and verbal game playing in Japanese young children.* 9th Biennial Meeting of the International Society for the Study of Behavioral Development, Tokyo.

O'Connor, J.D. (1991). *Phonetics.* London. Penguin.

Ohman, S.E.G. (1966). Coarticulation in VCV utterances: spectrographic measurements. *Journal of the Acoustical Society of America* 39: 151-168.

Otake, T., Hatano, G., Cutler, A., and Mehler, J. (1993). Mora or syllable? Speech segmentation in Japanese. *Journal of Memory and Language* 32: 258-278.

Port, R.F., Dalby, J. & O'Dell, M. (1987). Evidence for mora timing in Japanese. *Journal of the Acoustical Society of America* 81 (5): 1574-1585.

Poser, W.J. (1990). Evidence for foot structure in Japanese. *Language* 66: 78-105.

Recasens, D. (1984). Vowel-to-Vowel coarticulation in Catalan VCV sequences. *Journal of the Acoustical Society of America* 76: 1624-1635.

Vance, J.V. (1987). *An Introduction to Japanese Phonology.* Albany. SUNY Press.

Warner, N. and Arai, T., (2001a). Japanese mora timing: A Review. *Phonetica* 58: 1-25.

Warner, N. and Arai, T., (2001b). The role of the mora in the timing of spontaneous Japanese speech. *Journal of the Acoustical Society of America* 109 (3): 1144-1156.