A Review of Perceptual Approaches to Language Rhythm

Martin Gore

(Received 19 October, 2004)

Abstract

A critical review of recent perceptual approaches to language rhythm is presented, focusing on the question of vowel-sequence segmentation. Rhythmic classes are reappraised in relation to syllable structure and vocalic proportion, evidence from speech errors is reviewed and some problems in data collection are discussed. Infant acquisition of phonological preferences and Japanese children's segmentation strategies are discussed, and some ways of analyzing the putative effect of kana segmentation on children's rhythmic perception of vowel sequences are suggested.

Keywords. Language rhythm; acquisition; phonetics; phonology; psycholinguistics; English; Japanese; vowel sequences.

1. Introduction

As part of an ongoing cross-linguistic inquiry into the timing of diphthongal vowel sequences in English and Japanese, studies and experiments have been conducted from phonetic, phonological and psycholinguistic viewpoints (Gore 2004; 2003). The present study reviews recent psycholinguistic approaches to language rhythm insofar as they address the question of vowel-sequence segmentation.

In order to obtain a comprehensive but focused view of areas and issues involved in the analysis of the perception of diphthongs, literature in many phonetic, phonological and psycholinguistic fields has been sifted for special relevance to the central question of vowel-sequence timing. The material presented in the present study draws on studies of English and Japanese phonology in so far as they relate to the perception of vowel-sequence timing, juxtapositions of vowels in the respective languages, accent type and placement relevant to the question of whether a sequence is considered a

diphthong or not, and psycholinguistic research into language rhythm and speech segmentation. There is of course considerable overlap between the phonetic, phonological and psycholinguistic aspects, but the psycholinguistic aspect is considered to be of central importance to a proper understanding of the large-scale timing of vowel sequences, and in the last resort, central to a proper definition of the concept of diphthong itself.

2. Perception-based studies of language rhythm

Allen and Hawkins (1978) note that children's earliest phonologically patterned sounds tend to be dominated by heavy syllables. Similar results have been reported for children in English, French and Japanese (Mehler et al 1988; Kubozono 2001). With the acquisition of light syllables, which tend to occur frequently in function words and polysyllabic words, children gradually develop an adult type of rhythm which may be dominated by units of quite a different nature: stress-based in English, and mora-timed in Japanese. This does not mean to say that all other units disappear, or cease to have any important effects. As Crystal (1969), Crystal and Quirk (1969) and Roach (1982) have indicated, no language is totally syllable-timed or stress-timed, and different types of timing are exhibited even by the same speaker according to context and occasion. Each language is thus a mélange of different sized chunks, and some types of chunk will probably not vary significantly across languages.

An example of cross-linguistic similarity in prosodic timing can be found in the rhythm of counting from one to ten, which tends to be based on the timing of the heavy syllable, and thus does not vary markedly from language to language, or among age-groups. Such "heavy-syllable speech" is certainly not restricted to counting. For instance, in Japanese, despite the prevailing use of mora-sized bites, it is possible for a common everyday greeting to be entirely composed of larger units, as in, "Yuu-dai-kun dai-joo-bu?" In such utterances, the heavy syllable is clearly the most prominent unit and the one that determines the rhythm of the whole phrase. In English, too, heavy syllable sentences are possible, as in the "monobeat" of the memorable soap advertisement: "Gets out dirt plain soap can't reach!" (Bolinger 1981). However, it must be admitted that these examples, though not uncommon, are not usually considered *representative* of the rhythm of their respective languages.

In typical phrases in many languages, smaller words or units (often function words) tend to alternate with larger units. It is when smaller-sized function words and larger content words are mixed together that a collective "choice" has to be made with regard to segmentation in order to integrate both small and large elements into a common perceptual lattice. The options would seem to be: to persevere with the syllable as a common unit of sound, to select a smaller common time unit (the mora), or a larger sonority-based unit (the foot). To give a concrete example, a heavy syllable followed by a light syllable can be primarily perceived, either as three moras, two syllables, or one foot, depending on the language.

2.1. Rhythmic classes reappraised

Although the concepts of language-specific rhythm and phonetic isochrony have come under attack from various sources (Uldall 1971; Roach 1982), it has been shown that listeners tend to *perceive* isochrony even in sequences of inter-stress intervals that are clearly unequal (Allen 1975; Lehiste 1977; Lehiste 1979). The origin of this reported isochronic perception is not clear. This is in part a question for the psychologist, and indeed several psychologists and psycholinguists have recently taken up the baton, and have given considerable support to the concept of rhythmic classification of the world's languages. This has been done particularly effectively from the standpoint of infant and language learner phonology.

Mehler, Dupoux, Nazzi and Dehaene-Lambertz (1996) approached infant learning of native language phonology in terms of the syllable/stress dichotomy, and suggested that infants use rhythm to distinguish native-language utterances from those of a different rhythmic class. Low-pass filtered speech was used to highlight the role of prosody (Dehaene-Lambertz and Houston 1998; Mehler et al 1988), and it was found that French children could discriminate English and Japanese rhythms, but not Dutch and English (Nazzi, Bertoncini and Mehler 1998), thereby giving support to the reality of language-specific rhythm at least as a perceptual phenomenon. It thus seems that the intuitions of phoneticians such as Abercrombie (1967), were broadly right, and that the rhythmicality of language cannot be easily dismissed.

2.2. Perception and syllable structure

Dauer (1983) reported that stress-timed and syllable-timed languages have different phonetic and phonological properties, of which the two most important are thought to be as follows: (1) Syllable structure: stress-timed languages have a greater variety of syllable types, and tend to have more heavy syllables. Stress most often falls on the heaviest syllables, whereas in syllable- (and mora-) timed languages, stress and syllable weight tend to be independent. (2) Vowel reduction: in stress-timed languages, unstressed syllables usually have a reduced vocalic system (sometimes reduced to just one vowel, schwa), and unstressed vowels are shorter and can be deleted.

These features may combine to give the impression that some syllables are more salient than others in stress-timed languages, and that all syllables are equally salient in syllable-timed languages. This, in turn, may create the impression that there are different types of rhythm.

The central question for psycholinguists has been how this language rhythm is perceived and extracted from the complex pattern of everyday speech. Mehler, Dupoux, Nazzi and Dehaene-Lambertz (1996) and Ramus, Nespor and Mehler (1999) propose that infant speech perception is centered on vowels. This is thought to be due to the fact that vowels have more energy and tend to last longer than consonants. This is surely an uncontroversial view. Vowels also generally carry accent and contribute to whether a syllable is strong or weak.

Evidence has been presented that newborn infants respond more reliably to vowels than to consonants (Bertoncini et al 1988). It is therefore possible that infants perceive speech as a succession of vowels punctuated by periods of *initially* unanalyzed noise. This view parallels that presented in many coarticulation studies (e.g. Ohman 1966; Fowler 1980) that consonant articulation is superimposed on a more fundamental and continuous "diphthongal" articulation of vowels. Mehler, Dupoux, Nazzi and Dehaene-Lambertz (1996) named this the Time Intensity Grid Representation, or TIGRE.

2.3. Rhythm and proportion of vocalic interval

Ramus, Nespor and Mehler (1999) present a ground-breaking investigation that proposes a simple reliable phonetic correlate of language rhythm. They propose that

the simple division of speech into consonants and vowels will explain language discrimination observed in infants, clarify how rhythm is extracted from the speech signal and give an adequate account of the stress/syllable timing dichotomy. In Ramus, Nespor and Mehler (1999), eight languages were studied (English, Dutch, Polish, French, Spanish, Italian, Catalan and Japanese), with four speakers per language and five sentences per speaker. The duration of "vocalic and consonantal intervals" (adding up to total sentence duration) was measured. [1], below, shows the ordering of languages according to %V, i.e. the proportion of "vocalic intervals" within the sentence.

[1] Proportion of Vocalic Intervals (%V) and Standard Deviation (SD)

English	40.1	(5.4)
Polish	41.0	(3.4)
Dutch	42.3	(4.2)
French	43.6	(4.5)
Spanish	43.8	(4.0)
Italian	45.2	(3.9)
Catalan	45.6	(5.4)
Japanese	53.1	(3.4)

(from Ramus et al, 1999, p.272)

[2], below, shows the ordering of languages according to DC, the standard deviation of the duration of consonantal intervals within the sentence, i.e. "consonant flexibility."

[2] Standard Deviation of Consonant Intervals (DC)

English	5.35
Polish	5.14
Dutch	5.33
French	4.39
Spanish	4.74
Italian	4.81
Catalan	4.52
Japanese	3.56

(from Ramus et al, 1999, p.272)

It can be seen that these simple measurements of vocalic content and consonant flexibility reflect generally held views of language rhythm quite accurately. Ramus, Nespor and Mehler (1999), also point out that %V and DC are directly related to syllable variety, with English, Dutch and Polish ("more than 15 syllable types") at the top, and Japanese ("four syllable types") at the bottom. Thus their study also offers "empirical validation of the hypothesis that rhythm contrasts are accounted for by differences in the variety of syllable structures." This seems to show that "phonological properties have reliable phonetic correlates that can be measured in the speech signal, and that these correlates predict the rhythm classes."

However, it is not clear by what criteria Ramus, Nespor and Mehler (1999) judge English, Dutch and Polish to have more than 15 types and Japanese only four. In particular, the reference to Japanese as having only four syllable types is highly questionable. Campbell (1999) lists 17 phonetic syllable types for Japanese, though there is a significant drop in frequency of occurrence within his corpus after the first eight types, which are CV, CV2, V, CyV, CVN, QCV, QCV2, and V2.

In the same year as Ramus, Nespor and Mehler (1999), Campbell (1999), working separately, from a different background, with different aims (the measurement of "syllable-level accommodation" in Japanese) and with totally different sampling, measurement and data analysis procedures, nevertheless reaches some similar, or at least not incompatible conclusions: "vowels stretch to fill the syllable frame and show accommodation with the special morae [N, V^2, Q] ... consonants in Japanese are much more robust against durational change from the syllable level." Campbell emphasises the inflexibility of Japanese consonant duration "relative to the mean duration observed for all tokens of the same type," concluding, "the significant difference between the two languages [English and Japanese] lies in the flexibility of the consonants ... Japanese consonants are much less susceptible to influence from the syllable stretch."

Ramus, Nespor and Mehler (1999) showed that French subjects can discriminate English and Japanese sentences on the basis of rhythm only, without any other cues. They report "almost perfect separation of English and Japanese sentences along the %V dimension" and that English sentences with lower %V values were more reliably classified by subjects. Japanese sentences presented less variance in %V, and tended to be reliably classified. They conclude:

...correspondence between %V and the subjects' classification scores provides evidence

> for the psychological plausibility of the proposed model. The results suggest that subjects actually compute %V, and base their English/Japanese decision [threshold] at a value of approximately %V = 0.46 .... the ones having a %V further from the decision threshold being easier to classify.
>
> ...the overall pattern of success and failure to discriminate languages has been entirely simulated and predicted, on the basis of a simple model. This model assumes that subjects can compute the vowel/consonant temporal ratio ... and that their [rhythmic] categorization of sentences and languages is based on %V.

The physical measurements of speech sound in Ramus et al. thus suggest that the often-cited rhythm classes are indeed significant physical categories. They suggest that such groupings not only appeal to our intuitions about rhythm, but also reflect an actual physical property of speech in languages, namely the percentage of vowel sounds in the speech signal.

[2] seems to divide the eight languages into three natural groups: English, Polish, Dutch with approximately 40~42%V; French, Spanish, Italian and Catalan with 44~46%V; and Japanese with 53%V. This distribution leaves space for a possible class between Japanese and the middle group. There is a possibility that Korean, a language with some similarities to Japanese but a slightly higher proportion of CVC syllabic influence, might fill this gap. Comparative studies of Korean have been reported with results that were inconclusive in this respect, but not wholely incompatible with this view (e.g. Byung-jin Lim 2000). Levelt and van de Vijver (1998) propose five classes of syllable complexity, three of which seem to correspond to the above classes. Of the five proposed classes, one would seem to correspond to the gap noted between syllable-timed and mora-timed languages, and the remaining proposed class (consisting of pure CV languages) would presumably have an even higher %V value than Japanese. However, since sufficient data are not yet available for such languages, this remains unsubstantiated conjecture.

3. A combined phonological and perceptual approach

Kubozono (2001) takes a broad view of the mora, and considers its role in Japanese

in four distinct ways: a) as a unit of phonological weight or distance, b) as a unit of temporal regulation, c) as a unit of segmentation in speech production, and d) as a unit of segmentation in perception. He notes that the last three attributes of the mora are not reported widely in other languages.

His approach involves evidence from speech errors (1989), stuttering (1994), word blending tests (1995), speech segmentation (1996), crosslinguistic perspectives on the mora/syllable, the bimoraic foot, loanword accent (1999), vocabulary and syllable structure (2000), and issues such as epenthetic vowels and accent, the syllable as a unit of prosodic organization, and temporal neutralization (2001). Nine papers discussing these questions have been republished under the title, "Mora and Syllable in Japanese" (2001). The following paragraphs evaluate elements of Kubozono's arguments that may be relevant to the present investigation into vowel-sequence and diphthong timing.

3.1. Speech errors

Kubozono (2001) presents evidence to show that mora boundaries, rather than syllable boundaries, are the most common switch point in Japanese speech errors. It is an empirical question whether speech error "switch point" is an important aspect of the rhythmic difference that exists between Japanese and English. However, in view of the statistical imbalance between short and long syllables that Kubozono has cited, it must be said that his finding that most switches occur between moras (short syllables) is, in itself, not surprising.

Kubozono (1989; 2001) presents evidence from substitution/transposition errors to show that 64% of long syllables are replaced by another long syllable, 22% by two short syllables, and only 14% by a single short syllable. Kubozono (1989; 2001) observes from this evidence, "The syllable weight of long syllables is ... equivalent to the weight of two short syllables combined." However, "equivalence" of one long syllable with two short syllables is surely questionable, even from Kubozono's own evidence, since as many as 14% of long syllables were actually replaced by one short syllable, as against only 22% replacement by two short syllables. It might seem more appropriate, given the statistical evidence presented, to say that the long syllable is intermediate in weight between one and two short syllables, but closer

to that of two syllables. The reluctance to draw such a moderate conclusion may stem from the fact that it might be interpreted as undermining the mora argument. Kubozono adjusts his position later (2001) by saying, "Long syllables do not consist of two morae of equal status."

Kubozono's comparisons of English and Japanese speech errors throw some interesting light on syllable structure, including diphthongal sequences. Unlike English (Fromkin 1973) Japanese errors "often involve a split of complex (i.e. long or diphthongal) vowels as well as long consonants." Two examples of the splitting of diphthongal vowel sequences are given (p. 4):

[3]

a. mo.o-ta.a ba.i-ku 'motorbike'
   mo.i-ta.a ba.i-ku
b. ko.o-zu.i tju.u-i-ho.o 'flood warning'
   ko.o-zu.u tju.u-i-ho.o

Kubozono uses these examples to suggest that "complex vowels and long consonants in Japanese, unlike English, need not be interpreted as unitary in phonological description." However, the evidence presented does not constitute a convincing argument that this is really "unlike English." Of the two examples, in (a) *motorbike* is mistakenly pronounced /moitEbaik/ in English, too. (Since the first vowel in RP is usually /E/, a more exact analogy might be *ought I?*, which can be pronounced /oitai/). If this is accepted, then it is difficult to extrapolate from such evidence to a general phonological difference between English and Japanese. In (b), /ko.o-zu.u tju.u-i-ho.o/ shows not a "split" of a complex vowel but merely a smoothing of /u.i...u.u/ to /u.u...u.u/, which is also not unknown in English. Therefore this evidence does not, in itself, unequivocably substantiate Kubozono's suggestion of a clear difference between Japanese and English in the segmentation of the diphthong, though many would agree with his general thrust that there does seem to be difference.

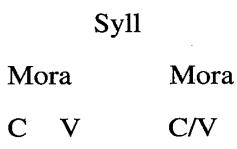
Another example given of metathesis within a syllabic frame, /kankei/ — /kaikei/ (p.4), may also in a strict phonological sense be moraic metathesis, but /kan/ and /kai/ are phonetically close. Since other clearly moraic substitutions such as /kao/ or kae/ would be extremely unlikely, even in Japanese, it does seem that there are limits to "moraic metathesis" as such. If such limits are accepted, this may also

be interpreted as evidence for existence of diphthongs such as /ai/. Kubozono, however, interprets this evidence in the opposite way: since this sort of metathesis involves "splitting of diphthongal vowel clusters," he says such vowel sequences in Japanese "unlike comparable units in English, need not be interpreted as unitary in phonological description" (p.4).

Thus, there seem to be several cases where the evidence Kubozono presents does not unequivocably support the conclusions he reaches. However, one important conclusion that the evidence clearly does support is that these interactions involve only the final segments of long syllables. In other words, the second half of the diphthong or long vowel is without doubt more mutable than the first. This is shown in the high proportion of replacement or metathesis of a nasal and the second element of a long vowel or diphthong. It can thus be said that "the second elements of long vowels and diphthongs occupy the same position within the syllable as do the first part of long consonants and the syllable-final nasal."

Since there is "little or no evidence in Japanese for a syllable constituent corresponding to the rhyme" and since there is "cohesiveness between the onset and the peak rather than the peak and the coda" (p.6), Kubozono proceeds to offer the following "un-Germanic" model for the Japanese syllable:

[4]



Accordingly, errors can be seen "not as segmental errors but as moraic errors, in which the non-syllabic mora of one long syllable (that is, the second mora of a bimoraic syllable) is replaced by the non-syllabic mora of another long syllable." And later, (p.15) "the reality of the syllable is substantiated ... by the finding that there is a strong constraint on inter-mora interactions ... the vast majority of errors involve replacement of a non-syllabic mora by another non-syllabic mora." A conclusive argument is Kubozono's observation that, "Without some unit that corresponds to the syllable, there is no way to distinguish between the two types of mora."

Kubozono then asks whether there is any evidence in Japanese for a "syllable

unit corresponding to the rhyme, the existence of which is supported in English and other languages." He suggests mora and rhyme are incompatible, since the former implies cohesiveness between the onset and the peak, whereas the latter implies cohesiveness between peak and coda. He asserts (p.17) that Japanese speech error analysis provides "no substantial evidence for the existence of the rhyme as a syllable unit," thus showing that "analysis of syllable into onset and rhyme does not represent a universal syllable structure, as often implied in the literature (cf. Fudge, 1987)."

Kubozono attempts to illustrate this with blend-error statistics. According to Kubozono, in Japanese, onset-peak boundary is the least preferred switch point. Only 2 out of 9 unambiguous cases showed onset-peak switch point (his *Table 2*). However, if we compare his statistics for the switch point in English blend errors with those for the switch point in blend errors (his *Table 3*: "ambiguous and unambiguous combined errors"), 59% *may* be onset-peak boundary switches, which is nearer the English 68% quoted; the remaining imbalance may be related to the fact that tense short vowels in English tend to be followed by a consonant, which would reduce the number of possible peak-offset boundary metatheses in English.

3.2. The problem of data collection

However, the problem with the approach outlined above may not be in the analysis, but in the collection of the data. Collection of speech error data will always have subjective elements, unless conducted by a computer trained to spot errors in human speech. As a native English speaker with fluent Japanese, I myself have listened out carefully for mistakes in Japanese for over 30 years and have heard quite a large proportion of onset-peak metatheses resulting in rhyme preservation analoguous to *down train / town drain* in English. The fact that I have noticed so many such errors may be because my native English background has made me especially sensitive to spoonerisms. However, if that is so, it follows that a Japanese linguist brought up in Japan would be *less* tuned in to "rhyme" errors, and less inclined to interpret an ambiguous error on the basis of rhyme.

In other words, if there is a tendency for a certain type of mistake to be rather rare in a language, that would affect not only the speaker *but also the collector of the data*. It follows that a mistake that does not fit the listener's preconceptions

may easily be overlooked. This factor would multiply the effect on the data collected. A similar multiplication effect, in reverse, may have occurred historically with stress "isochrony" in English, where not only the speaker but presumably also the phonetician conspired to create a theoretical isochrony which could not be justified in physical terms. Thus, rhyme-related errors may indeed be few in Japanese, but they are also noticed less, and even where they are noticed they tend not to be categorized as rhyme-related errors. So if, in reality, onset-rhyme switch-point errors constitute 10% of total Japanese errors, only 10% of these (i.e. 1% of the total) will be recorded by Japanese linguists looking for "moraic" segmentation.

This multiplication effect must be taken into account when we read Kubozono's blanket assertion that "speech error data in Japanese provide no substantial evidence for the existence of the rhyme as a syllabic unit" (p.15).

3.3. Non-moraic segmentation

The following are non-moraic errors which I have heard in university department meetings in Japan. All, in spite of Kubozono's assertions, seem to be motivated by a force that acts to preserve the rhyme as a unit.

[5]

a. jizeN-jigo shidoo — jigeN-jizo shidoo
b. dooki zuke — zooki-zuke
c. biizu kuQshon — biizu buQshon

One way of viewing [5] (a) is to say that the phonemic segment /g/ has metathesized from the second foot to the first, while /z/ has moved in the opposite direction to take its place. Another way is to say that /gen/ is a common SJ syllabic triphone and that /jigen/ is also a common SJ word; this would make the mistake more likely to occur. However, in either case it is not the mora which has predominated, but the phoneme or the syllable. It is particularly interesting that the rhymes /-en/ and /-o/ have been preserved in both feet, which conflicts with the notion that rhymes are not important in metathesis in Japanese.

Strict moraic metathesis would give: /jigon-jize shidoo/, which is, according to

several Japanese informants an "inconceivable" error. Why moraic metathesis should sound "inconceivable" is an intriguing question that challenges Kubozono's theory that metathesis is moraic in Japanese. It seems that part of the answer may be that rhyme does play a role, after all. But this conflicts with Kubozono's repeated assertion that there is little or no evidence for rhyme. Later (p. 20) we find an even stronger denial: "there is no form of verse in Japanese in which either the onset or the rhyme plays a crucial part," yet we know that onset and rhyme do play an important part in both Kanshi (Chinese or Sino-Japanese verse) and in Buddhist sutras, though it is true that Kanshi are often read in a quasi-Japanese manner.

However, even some of Kubozono's own YJ examples of moraic metathesis actually preserve the rhyme in a way that does not sound strange to English ears at all. For instance, /noren ni udeosi/ --- /uden ni noreosi/ (p.54) preserves the rhyme /–en/, but this is not remarked upon. It is surely the sameness of the *vowel,* /e/, not the mora (/de/ changing to /re/), that encourages the metathesis to happen in this particular case. It is also interesting that the undisputably moraic metathesis in /tekkin konkuriito/ --- /kokkin tenkuriito/ (p.54) occurs in the first syllable, whereas the /-en/ rhyme preservation mentioned above occurs in the second syllable. From an English point of view there is nothing strange about this. While no-one would dispute that moraic metathesis is generally the predominant form in Japanese, an effort must be made to explain cases of segmental or syllabic metathesis which, in preserving the rhyme, not only break the moraic rule, but *cannot be made to obey it without considerable unnaturalness.*

We propose that rhyme *does* play a role in Japanese, though it is peripheral rather than central. The role is small when compared with that of the mora, and it is more common in SJ contexts. In such contexts it is in order to maintain the rhyme that the onset consonant rather than the mora metathesizes. If these exceptions cannot be explained except by reference to onset/rhyme segmentation, and if the corresponding mora-based error is deemed inconceivable, as in the present case, extant solutions cannot be viable and more research is needed. It is unfortunately beyond the scope of the present investigation to conduct a thorough-going collection or analysis of such error data. If many more such examples are found they will constitute an argument in favor of syllabic prosody in Japanese which does not rely on accent placement. Experiments reported by Hatano and Inagaki (1992) show that the syllable seems to be more important for young children, and we already know that the syllable

is important in Sino-Japanese compounds. This may have an effect on rhyme as a (non-moraic) sub-syllabic unit. Kubozono suggests (personal communication) that errors involving onset/rhyme segmentation may be more common in children and/or in Sino-Japanese compounds, but in adults "are much less common than errors involving a mora-sized unit."

All this points to the influence of kana orthography (which is not used by very young children or in Sino-Japanese) as a mora-reinforcing factor, and as a factor that makes syllable, and rhyme as a sub-syllabic unit, difficult both to conceptualize and to express.

There is another phenomenon in Japanese that may be considered a sort of "supra-syllabic rhyme" whereby the last syllable *or syllables* of a reduplicative word are preserved while the first mora only is changed. One such is /tinpun-kanpun/ "hocus-pocus." Another example is /sidoro-modoro/ "tongue-tied." The high-vowel/low-vowel alteration is, incidentally, similar to that commonly found in English reduplicatives, such as "shilly-shally" and "pitter-patter." There are similar examples where rhyme is preserved in the conventional sense as a subsyllabic unit, too, e.g. /nantoka-kantoka/.

4. Orthography and the perception of diphthongs in Japanese

When it comes to the question of whether diphthongs actually exist in Japanese, Japanese multi-layered orthography may underly the reality (if diphthongs do exist), or alternatively the misapprehension (if they do not). Thus, the possibility of some degree of interaction between orthography and phonology needs to be tested.

4.1. Pre-literate and literate phonology

Pre-school Japanese children have little awareness of the inherently moraic and prescriptive kana orthography, and thus may receive a greater influence from syllables or other segments than the mora. The speech of the very young and "carer language" have been shown (not just in Japanese) to have a special bias towards heavy first syllables. These are defined as containing a nucleus of one mora plus a post-nuclear element of one degenerate mora (Kubozono 2001), which may be a continuation

of the vowel, the second part of a diphthong, a final nasal (N) or a gemination (Q) of the following consonant. Thus "cat" is not "neko" but "nyaNko," and phonologically distinctive syllabic sequences such as "haihai" (crawling) and "oQpai" (breast/milk) proliferate.

If pre-school or young children show the influence of such units, this may provide some grounds for believing that diphthong syllables do exist at a certain conceptual/ phonological level, as an acoustic pattern or 'trace' in the infant lexicon.

4.2. Infant acquisition of phonological preferences

Jusczyk (1997) discusses how human infants acquire the phonetic skills needed to perceive and produce language, and shows how a large number of possible language contrasts are discriminable by infants shortly after birth. Children gradually develop a preference for the sounds of their own language and an inability to distinguish contrasts which are distinctive only in other languages. The narrowing-down of the ability to perceive foreign contrasts is related to the attempt to coordinate the perceptual and articulatory capabilities as children focus on the sounds they are learning to pronounce. The ability to distinguish most of the non-native sounds is then gradually lost because there is no articulatory-perceptual link.

Whether and/or when this affects the application of mother language *rhythm* to the infants' speech segmentation strategies is a matter for debate. There seems to be a loose consensus that it does: i.e children's speech segmentation strategies are thought to become fixed in much the same way as their preferences and inabilities regarding phonetic distinctions become fixed, but it is doubtful whether all these processes occur at the same stage of development.

Let us first summarize Jusczyk's model of infant speech perception, Word Recognition and Phonetic Structure Acquisition (WRAPSA), insofar as it may relate to the present inquiry. This is a model of "innately guided learning" where young children are programmed to learn particular kinds of things in particular ways. The infant's cognitive system uses innate programming to acquire the language it hears by focusing on features that are phonologically significant in the target language. However, just as correct pronunciation of a language involves a specific articulatory posture, learning to perceive includes language-specific tuning of the perceptual facilities, so as the

brain grows, the movement is away from general-purpose recognition of phonetic differences to a specific focusing on only those differences that are contrastive within the infant's language.

The lowest stage of analysis stores the features of every utterance that the system is capable of extracting. Multiple tokens will thus have multiple representations. This allows young infants to detect differences irrelevant to the target language, and "the description that emerges is neutral with respect to the language that is spoken." (p.215). This representation is thought to decay rapidly.

The next stage focuses attention on language-specific features. This may be mildly stimulated at birth, but increasing language input causes a language-specific warping of perceptual space. The third stage allows the extraction of patterns (syllable shape and number, and prosodic information) from the weighted output of the second and stores these patterns in memory, where they comprise the infant lexicon. The fourth compares new patterns with old ones to attempt a match. Feature similarities cause old patterns (traces) to be stimulated. As more tokens of each utterance are collected, more traces will be activated by new tokens, so recognition of patterns will soon become more efficient and will lead to extraction of words. All these stages are thought to occur within the first two or three years of life.

Yet it is probably true to say that as old traces are stimulated they also become susceptible to *reformatting*. It is also probable that the finer the phonological distinction, the later it will be incorporated into the system. This may be relevant to the acquistion of mora timing in Japanese, particularly in relation to heavy syllables, i.e. those that contain Q, N, or V^2. In the above synopsis of Jusczyk's WRAPSA model (1997), after the fourth stage where old patterns are stimulated and a match is attemped with new patterns, it is possible that a process of *reformatting* occurs, which may, in rhythmic terms, be more gradual and subtle than has been thought hitherto, and which may well be triggered by knowledge of the orthography. This would be akin to discovering that a tune one thought from an early age to have a simple up-down rhythm (2/4) was actually written in a rhythm of four (4/4) -- such a discovery would surely then spread rapidly to affect one's perception of all musical rhythm. In other words, one's perception of "rhythm" itself would have become enlarged or reformatted to include new possibilities. This sort of reformatting would be triggered by orthography. In Japanese, some diphthongal vowel sequences may continue to be perceived as syllables, like long vowels, until quite a late stage of phonological

development. It is only then that, triggered by the discovery of kana, they become reformatted, as dual monophthongs.

4.3. Japanese children's shift in segmentation strategy

Inagaki and co-workers (1992; 2000) present evidence that Japanese children's segmentation strategies shift gradually around the age of six from syllable-based (or mixed syllable/mora-based) to almost exclusively mora-based segmentation in association with kana acquisition, and their conclusion is "Japanese speech segmentation is influenced by the acquisition and use of kana literacy." Further evidence to support this has been reported by a small number of researchers including Hatano and Inagaki (1992), Kubozono (1998; 2001) and Aoyama (1999; 2002). Very young (preschool) Japanese children do not seem sensitive to moraic timing distinctions (such as between /kite/ and /kiite/, or Aoyama's example of /hana/ and /hanna/) in quite the same way that older (school) children are. It is certainly conceivable that the moraic nature of these timing differences is reinforced by kana acquisition.

5. Experimental studies of segmentation strategies

Several studies since the 1980s have shown that speakers of languages in different rhythm classes use different segmentation strategies. As we have seen, French speakers listening to French are reported to use a syllabic segmentation procedure (Mehler, Dommergues, Frauenfelder and Segui 1981), whereas English speakers listening to English, which is stress-based, use a procedure that is said to depend on stress timing (Cutler and Norris 1988). The following sections review studies of segmentation strategy relating to Japanese.

5.1. Japanese segmentation strategies

In spite of Beckman's (1982) challenge to the phonetic status of the mora in Japanese, it is still generally believed that the mora is an important phonological

unit. Recent psycholinguistic experiments have added support to this position by showing that native adult speakers of Tokyo Japanese segment their speech using the mora as the most basic unit. For example, Otake, Hatano, Cutler and Mehler (1993), performing target monitoring experiments with college students as participants, found that a CV target (e.g., ta) was recognized in words beginning with a CV initial (e.g., tanishi) or CVN initial (e.g., taNshi) more readily than a CVN target (e.g., taN) was. Finding a CVN target (e.g., taN) in words starting with CVNV (e.g., tanishi) was extremely hard. In other words, the CVN embedded in the first two morae was difficult to recognize.

In their study, Japanese students listened to Japanese CVCVCV and CVNCV words, and pressed a response key when they heard CV or CVC sequences. From a syllabic hypotheses, CV targets in CVCVCV words and CVN targets in CVNCV words would be easy to detect, perhaps equivalently so, while CV targets in CVNCV words and CVN targets in CVCVCV words would, again, pattern similarly and would cause difficulty for listeners. However, from this experiment, the predictions of the syllable hypothesis for Japanese are not confirmed. In fact, CV targets in CVNCV words were as easy to detect as CV targets in CVCVCV words, while CVN targets in CVNCV words in fact produced the slowest responses. Thus the syllable hypothesis cannot account for Japanese listeners' performance. The mora hypothesis, on the other hand, can account for the results. Because the initial mora of both word types is the same (CV), the mora hypothesis predicted that CV targets in both CVCVCV and CVNCV words should be easy to detect, and equivalently so. Indeed, there was no difference between these two conditions either in error rates or in response times. The pattern of results in this experiment thus appears to offer support for the mora hypothesis but none to the syllable hypothesis. On the other hand, there may have been some bias in the experiment towards moraic segmentation. The target words (/tanishi/ "mud snail" and /tanshi/ "terminal/outlet") are extremely rare and not easily recognizable, even by university graduates. This may have biased the results in favour of smaller segmentation units, since rare words tend to demand clearer pronunciation (Wells 1990) or spelling-visualization (Bolinger 1981). Without challenging the accuracy of their results, it is indeed strange that the experimenters chose such rare words upon which to build such an all-embracing argument concerning language rhythm.

It is undeniable that the mora is central to Japanese phonology. In Japanese,

the postvocalic portion of a syllable, such as the nasal coda, is a separate mora, and is not closely connected to the preceding vowel, whereas the syllable onset cannot readily be separated from the vowel with which it makes a mora, because without the vowel the onset has no phonological status. It is generally thought that accurate perception and production of adult Japanese speech (especially Tokyo speech) depends on the approximately equal duration of the mora, not on any attribute of the syllable. In other words, the unit of timing, rather than sonority or dynamics, is the primary segmentation marker. Words such as /kite/, /kiite/, /kitte/ and /kitee/ are distinguished by contrastive duration, which is held to be measurable in a moraic unit. Otake and Yoneyama (1994) point out that if the nasal coda in /tenoto/, a nonsense word, is prolonged, the word is taken as /teNnoto/ by Japanese adults. It is therefore not unreasonable to conclude that, in standard (adult) Japanese, mora-based segmentation is more efficient than syllable-based segmentation or other procedures. Here again, however, there may be an influence of spelling visualization. The use of rare or nonsense words may encourage the listener to try to visualize the spelling, which would inevitably implicate the mora-based kana system. It does seem that preschool children, who are less aware of kana letters, are rather late to appreciate some differences in timing, such as between /hana/ and /hanna/ (Aoyama 1999).

5.2. Effect of kana on children's perception

It is debatable whether the use of a mora-based segmentation procedure can be attributed to purely "phonological" aspects of Japanese. Phonological awareness may also be influenced by acquisition and utilization of a writing system (Morais, Bertelson, Cary and Algeria 1986). This effect may be heightened in Japanese by the peculiarly prescriptive nature of one of the writing systems, especially the fact that it is prescriptive in a temporal sense. Since the borrowed Chinese characters had meaning but little phonetic content, kana were needed, and still are needed, as symbols of phonetic transcription not unlike the IPA. In transcribing Chinese, a choice had to be made for each character with regard to length. The kana phonetic symbols therefore have had a peculiarly temporally prescriptive effect that is not found in other systems, except perhaps, to a certain extent, in Korean hangul and in IPA long marks. It was decided at a historically early stage not to include final consonants in the kana,

even though they may have existed in speech (and still do in many forms and in many areas of Japan.) This has an added constraining effect. Each item of the basic kana inventory graph represents a single V or CV, and each occupies an equal space, both on paper and, from the purist's viewpoint, in spoken time. Therefore, it is not inconceivable that the mora-based segmentation strategy used by Japanese adults may actually be a product of the highly rhythmical kana chanting, reading and writing which begins in earnest at the age of 6. Inagaki, Hatano and Otake (2000) suggest that "increased experience in reading and writing changes the preferred segmentation procedure used by native speakers, including the basic unit of speech segmentation."

Against this, there is a general belief that speech segmentation strategies do not change significantly after the first year of childhood. Studies on infants have reported that speech segmentation procedures are set very early in development and are thereafter immobile. This is because infants are highly sensitive to the rhythm of the language spoken by the people around them (Mehler et al. 1988, Jusczyk 1997), and it is therefore probable that children develop a segmentation procedure adapted to their native language long before they learn to read and write. Indeed this process is thought to be the very essence of "segmentation strategy." Advanced phonological awareness and readiness to learn to read result from increased oral vocabularies, which require increasingly elaborate phonological representations of words (Jusczyk 1997; Goswami, Porpodas and Wheelwright 1997; Metsala and Walley 1998), but it has hitherto not been thought that such elaboration would include different segmentation units. Even "bilinguals" find it difficult, according to some reports, to switch segmentation procedures when they switch from one language to another (Cutler, Mehler, Norris, and Segui 1992). All in all, at the present time, the majority of research seems to support the position that speech segmentation procedures are not modified by experiences such as reading and writing.

5.3. Reasons for suspecting orthographic influence in diphthong segmentation

The most basic reason for suspecting orthographic influence in Japanese diphthong segmentation are that diphthongs, like long vowels and the other "special" syllables in Japanese, can only be represented by two separate moraic graphs (kana), which

are each placed in a separate box on the boxed writing paper. At elementary school, manipulative activities centered on these kana boxes, such as chanting, reading and writing, may well encourage children to "reformat" their segmentation of sequences that may previously have been perceived as a single syllable. It is possible that they thus learn, rather late in their development, to give each "fraction" an isochronic time value, and to make a distinction between long and short syllables based on that value.

6. Ways of analyzing the possible effect of kana literacy on the perception of diphthongal sequences

One way to examine whether there is an effect of kana literacy on Japanese speech segmentation is to conduct experiments using illiterate adult speakers as participants, as Morais and his co-workers have done (Morais, Bertelson, Cary and Alegria 1986; Morais, Cary, Alegria and Bertelson 1979). Second and later generation Japanese living in other countries could be also tested (Otake and Yoneyama 1998), but their segmentation strategies may be influenced by the language of the country of domicile. Indeed there is anecdotal evidence that second and third generation Japanese in America do show more stress on heavy syllables and more gliding on diphthongal sequences than is considered normal in Japan, which points to the influence of English prosodic timing.

Another experimental method to probe the effect of orthographic knowledge is to use preschool children who have not learned kana or who are relatively unskilled in manipulating kana. Young children may show a greater or a lesser tendency to use mora-based segmentation than older children. We know that older school children use mainly moraic segmentation when playing *shiritori,* and that any deviation from moraic segmentation is considered a "mistake." The fact that this is considered a mistake is, in itself, indicative that there is something to be corrected. This "something" may be syllabic perception.

Hatano, Inagaki and Otake (2000) report that the lowest level children in their study based their segmentation on the syllable and the mora with approximately equal frequencies, and that children at higher levels of kana recognition showed increasingly mora-based patterns. 100% moraic segmentation of (non-diphthongal) special syllables

was not achieved until the seventh year. They suggest a connection with school entry at six, and the concomitant sudden increase in reading, writing and chanting of the kana. Children who do not use a mora-based segmentation procedure can divide words into moras when instructed. They suggest there may be a delay between the ability to segment words into moras and the habitual use of the mora as the basic segmentation unit.

If these conclusions are true, there are considerable theoretical implications. More experiments and a thorough discussion of the implications are the subject of a separate paper in preparation. It should be possible to test the unit that young children use in diphthongal sequences, too, by asking them to segment a word by clapping, and/or by using tail-catching (*shiritori)* with a diphthongal sequence in the tail. Tests have been done for sequences containing N, Q and long vowels (e.g. Inagaki and Hatano 1992; Hatano, Inagaki and Otake 2000) but these have not tested words containing diphthongal sequences. The reasons for this are not clear, but may be connected with the complexity of the subject matter and the difficulty of delimiting the field. Other types of "syllable" are relatively easy to define in Japanese, but there are many ways of defining "diphthong," and since any juxtaposition of vowels is possible in the Japanese word, this is not an easy task. It is also possible that diphthongs were not considered a likely candidate for "segmentation shift," because from an adult perceptual standpoint it is very difficult to view any non-identical Japanese vowel sequence as a single sound, whereas the sounds chosen for analysis by Hatano and Inagaki are generally recognized as being a special set: unique components of the so-called "special" syllables represented orthographically by "special" kana such as the small "tsu" that indicates the silent moraic obstruent (Q). Diphthongs are a notoriously difficult set to define in Japanese, because any combination of vowels is legal, and because there is little agreement on criteria for determining which are (or may be perceived as) heavy syllables as opposed to two juxtaposed light syllables.

There are also considerable distributional limitations on the choice of target material. Whereas /ai/ occurs frequently, neither /au/ nor /oi/ occurs in SJ morphemes, and these sequences are consequently less common in Japanese as a whole. Apart from /ai/, there are few other diphthongal sequences that occur in nouns regularly used by children, except /ei/, and few other diphthongal sequences can begin or end a noun. In spite of all these limitations, it is surely possible to use shiritori and clapping to test diphthongal sequence perception, given an appropriately defined

field. It may also be possible to test the relative diphthongality of different sequences.

In spite of certain practical limitations imposed by recognizability and the children's restricted vocabulary, test procedures used hitherto may be enhanced by using pictures as the stimulus. This would help to elicit the child's own segmentation strategy and prevent contamination from the tester's pronunciation. Experiments to test diphthongal perception and the effect of kana on Japanese children's perception of diphthongal sequences are in progress, and results will be reported shortly.

References:

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

Allen, G. D. (1975). Speech rhythm: its relation to performance universals and articulatory timing. *Journal of Phonetics* 3: 75-86.

Allen, G. & Hawkins, S. (1978). The development of phonological rhythm. In Bell, A. & Hooper, J.B (eds.) *Syllables and Segments.* 175-85. Amsterdam: North-Holland.

Aoyama, K. (1999). Acquiring mora timing: the case of the Japanese coda nasal. *Japanese/ Korean Linguistics* 9: 64-71. Stanford, CA. http://webpages.acs.ttu.edu/kaoyama/papers.html

Aoyama, K. (2002). Quantity contrasts in Japanese and Finnish: Differences in adult production and acquisition. *Studies in Language Sciences (2): Papers from the Second Annual Conference of the Japanese Society for Language Sciences.* 121-135. Tokyo: Kuroshio. http://webpages.acs.ttu.edu/kaoyama/papers.html

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

Bertoncini, J., Bijeljac-Babic, R., Jusczyk, P.W., Kennedy, L.J. & Mehler, J. (1988). An investigation of young infants' perceptual representations of speech sounds. *Journal of Experimental Psychology: General* 117 (1) 21-33.

Bolinger, D. (1981) *Two Kinds of Vowels, Two Kinds of Rhythm.* Indiana University Linguistics Club, Indiana.

Campbell, N. (1999). A study of Japanese speech timing from the syllable perspective. *Journal of the Phonetic Society of Japan* 3 (2) 29-39.

Crystal, D. & Quirk, R. (1964) *Systems of prosodic and paralinguistic features in English.* The Hague: Mouton.

Crystal, D. (1969) *Prosodic systems and intonation in English.* Cambridge, UK: Cambridge University Press.

Cutler, A., Mehler, J., Norris, D. G., & Segui, J. (1986). The syllable's differing role in the segmentation of French and English. *Journal of Memory and Language* 25: 385-400.

Cutler, A., Mehler, J., Norris, D. G., & Segui, J. (1992). The monolingual nature of speech segmentation by bilinguals. *Cognitive Psychology.* 24: 381-410.

Cutler, A., & Norris, D. G. (1988). The role of strong syllables in segmentation for lexical access. *Journal of Experimental Psychology: Human Perception & Performance* 14: 113-121.

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

Dauer, R.M. (1983) Stress-timing and syllable timing reanalyzed. *Journal of Phonetics* 11: 51-62.

Dehaene-Lambertz, G., & Houston, D.M. (1998). Faster orientation latency toward native language in two-month-old infants. *Language and Speech* 41: 21-43.

Fowler, C. (1980). Coarticulation and theories of extrinsic timing. *Journal of Phonetics* 8: 113-133.

Fromkin, V. (1973). *Speech Errors as Linguistic Evidence.* The Hague: Mouton.

Fudge, E. (1987). Branching structure within the syllable. *Journal of Linguisitcs* 23: 359-377.

Gore, M. (2004). Phonological and phonetic views of vowel sequence timing: a cross-linguistic review. *Bulletin of the Faculty of Education, Kagoshima University 55* (Studies in Education). 83-99.

Gore, M. (2003). Phonetic, phonological and pedagogical views of Japanese and English diphthongs. *Proceedings of the 8th Annual Conference of the English Phonetic Society of Japan.* 52-57.

Gore, M. (2003). What is a diphthong? *Journal of the English Phonetic Society of Japan.* 6. 217-226.

Gore, M. (2003). The relationship between language-specific prosodic timing and intersegmental transition speed. *Bulletin of the Faculty of Education, Kagoshima University* 54. (Cultural & Social Sciences). 147-169.

Goswami, U., Porpodas, C. & Wheelwright, S. (1997). Children's orthographic representations in English and Greek. *European Journal of Psychology of Education, 3.*

Inagaki, K. & Hatano, G. (1992) *Phonological awareness in Japanese children (1) Word segmentation.* 56th Annual Convention, Japanese Association of Psychology. (Japanese).

Hatano, G. & Inagaki, K. (1992) *Phonological awareness in Japanese children (2) Tail-catching.* 56th Annual Convention, Japanese Association of Psychology. (Japanese).

Inagaki, K., Hatano, G. & Otake, T. (2000). The effect of kana literacy acquisition on the speech segmentation unit used by Japanese young children. *Journal of Experimental Child Psychology* 75: 70–91.

Hayes, B. (1995). *Metrical Stress Theory: Principles and Case Studies.* Chicago: University of Chicago Press.

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.

Jusczyk, P.W. (1997) *The Discovery of Spoken Language,* Cambridge MA: MITPress.

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

Kubozono, H. (1989). The mora and syllable structure in Japanese, evidence from speech errors. *Language and Speech* 32: 249-278.

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

Kubozono, H. (2002). Prosodic structure of loanwords in Japanese: syllable structure, accent and morphology. *Journal of the Phonetic Society of Japan* 6 (1) 79-97.

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

Levelt, C. & Van de Vijver, R (1998). Syllable types in cross-linguistic and developmental grammars. In *Fixing Priorities: Constraints in Phonological Acquisition,* Cambridge: Cambridge University Press.

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

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

Mehler, J., Dupoux, E., Nazzi, T., & Dehaene-Lambertz, G., (1996). Coping with linguistic diversity: the infant's viewpoint. In Morgan, J.L. & Demuth, K., *Signal to Syntax: Bootstrapping from Speeech to Grammar in Early Acquisition.* 101-116. Mahwah, NJ: Lawrence Erlbaum Associates.

Mehler, J., Dommergues, J.-Y., Frauenfelder, U., & Segui, J. (1981). The syllable's role in speech segmentation. *Journal of Verbal Learning and Verbal Behavior* 20: 298-305.

Mehler, J., Jusczyk, P.W., Lambertz, G., Halsted, N., Bertoncini, J. K., & Amiel-Tison, C. (1988). A precursor of language acquisition in young infants. *Cognition* 29: 143-178.

Metsala, J., & Walley, A. (1998). Spoken vocabulary growth and the segmental restructuring of lexical representations: precursors to phonemic awareness and early reading ability. In J. Metsala & L. Ehri (eds.), *Word recognition in beginning literacy* (pp.89-120). Mahwah, NJ: Erlbaum.

Morais, J., Bertelson, P., Cary, L., & Algeria, J. (1986). Literacy training and speech segmentation. *Cognition* 24: 45-64.

Morais, J., Cary, L., Alegria, J., & Bertelson, P. (1979). Does awareness of speech as a sequence of phonemes arise spontaneously? *Cognition* 7: 323-331.

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.

National Language Research Institute. (1972). *Reading and writing ability in preschool children.* Tokyo: Tokyo Shoseki. [in Japanese]

Nazzi, T., Bertoncini, J. & Mehler, J. (1998). Language discrimination by newborns: towards an understanding of the role of rhythm. *Journal of Experimental Psychology: Human Perception and Performance* 24 (3) 756-766.

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

Otake, T., & Yoneyama, K. (1994). A moraic nasal and syllable structure in Japanese. *Proceedings of the International Conference on Spoken Language Processing* 3: 1124-1128.

Otake, T., & Yoneyama, K. (1998). Phonological units in speech segmentation and phonological awareness. *Proceedings of the International Conference on Spoken Language Processing* 5: 2179-2182.

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.

Ramus, N., Nespor, M. & Mehler, J. (1999) Correlates of linguistic rhythm in the speech signal. *Cognition* 73: 265-293

Roach, P. (1982). On the distinction between 'stress-timed' and 'syllable-timed' languages. In D. Crystal, *Linguistic controversies.* London. Edward Arnold.

Uldall, E.T. (1971). Isochronous stresses in R.P. *Form and Substance.* Hammerich, L.L. et al. (eds.), 205-210. Copenhagen: Akademisk Forlag.

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

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

Warner, N. & 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

Wells, J.C. (1990) *Longman Pronunciation Dictionary.* Longman.