

A Summary of Research into Vowel Sequences -- with Implications for Language Acquisition

Martin Gore

(Received 20 September, 2006)

Abstract.

This paper summarizes research into vocalic sequences in Japanese and English. It brings together studies conducted from diverse viewpoints (phonetic, psycholinguistic and pedagogical) and attempts to provide a concise but overall view of the difference between the diphthong and the non-diphthongal VV, implications for language rhythm, and relevance to language acquisition.

Keywords : Phonetics; perception; vowel sequence; syllable / mora; language rhythm; language acquisition.

1. Arguments in the literature

Whether a vowel sequence constitutes a diphthong or not relates to several fundamental aspects of language: whether the syllable (a unit of sonority) or the mora (a unit of timing) is predominant; the relationship between the phonological and the phonetic diphthong; and children's phonological development (Gore 2003a,b; 2005a,b; 2006b). We summarize recent studies of these below.

Though the mora is central to Japanese phonology there is still some controversy regarding its precise nature (see reviews in Warner and Arai 2001, and Gore 2006b). A possible compromise suggested in Gore (2006b) is to view the mora as a *chroneme* with *allochrones* influenced by intrinsic segmental length and prominence (Jones 1967 p.266). This allows different forms of the mora, which theoretically may account for varied perceptions of the VV. Children's earliest patterned sounds in many languages make special use of "heavy syllables" (Allen and Hawkins, 1978; Cutler, McQueen and Robinson, 1990; Jusczyk 1997; Demuth 2003), but it is by no means clear whether this a relevant unit in adult central Japanese. Young Japanese children are said to employ a mixture of moraic and syllabic approaches to the VV, but older children prefer moraic strategies (Hatano 1992). Our most recent research (Gore 2006b) suggests that phonological awareness and acquisition of kana reinforce each other to allow a reformatting of initially perceived "long syllables" into dual moras. From a

morphophonological viewpoint, one type of Japanese non-identical VV, /ai/, has some distinctive distributional characteristics (Gore 2006b): (1) it is the only morpheme-internal non-identical VV to follow the full variety of onsets that can precede a single vowel; (2) /ai/ morphemes are extremely common in modern Japanese; (3) /-ai/ morphemes are peculiarly resistant to division under abbreviation processes; (4) /ai/ is also the only non-identical VV that can occur on an undivided note among 580 songs surveyed. It thus has distributional properties that set it apart as a potential linguistic unit. However, there is no evidence that the perceived “syllable” and the distributionally significant “-ai morpheme” are related in any systematic way that might justify the use of the term “diphthong”

Whereas the diphthong in English is well established (Gore 2006b), arguments for the “diphthong” in Japanese tend to be theoretical and lacking in acoustic evidence. There are arguments for (1) a bimoraic (C)VV unit; (2) an accented (C)V followed by a weak V; (3) a type of VV morphophonological unit; and/or (4) a (C)VV morpheme; but we have not found an argument that justifies the diphthong in Japanese as a consistent category whose members are phonetically distinguishable from the V#V. The “diphthong” has mainly been described in terms of pitch (a high-pitch followed by a low on two vowels such as /ai/ (Kubozono 2001b)). But this method seems arbitrary in a moraic language for several reasons: (1) The same pitch may occur on vowels separated by a morpheme boundary (#). (2) It is difficult to see why pitch does not define a special relationship between *any* moras, e.g. the V+Q sequence such that Q (the mora obstruent) could belong to the previous or subsequent “syllable” according to pitch. (3) Pitch movements are generally described as occurring on the mora boundary, which highlights the importance of the mora rather than the syllable. (4) “The functional load of accent distinctions in standard Japanese is very low, and there are some dialects that have no accent distinctions at all” (Vance 1987 p.106). Pitch is therefore not a sufficient justification for the “syllable” as distinct from the mora or bimora. Some V moras are certainly less sonorous (just as CV moras may be) and may form the second part of a *quasi-syllabic bimoraic VV*, but there is no evidence that this is systematically related to a significant linguistic unit in Japanese.

2. Acoustic comparisons with English

In English there is a clear difference between /a#i/ and /ai/ that rests largely on the duration of transition between the two parts of the sequence, with a threshold at 29% of total vocalic duration above which all tokens are diphthongs; F0, F1, F2 pitch and amplitude are less significant factors (Gore 2005c). The English “diphthong” is thus a distinct and self-contained category definable in terms of transition duration. If we accept Lehiste’s (1972) conclusion that temporal organization is basically independent of morphology, then the apparent strong effect of morphology on the internal temporal organization of the English VV is probably due to the fact that the English monomorphemic

sequence is typically a member of a distinctive phonemic category, namely the diphthong.

Japanese tokens, by contrast, show no phonetic difference between /a#i/ and /ai/ that could reliably assign most to one or other morphemic category (Gore 2005c; 2006a): they have very similar amplitude and F0 contours, similar peak-counts, and similarly rapid transitions between the two Vs. There is slight evidence of difference in sound between Japanese /a#i/ and /ai/ in two sets of data, depending on speech rate: (1) At very slow tempos, Japanese /a#i/ occasionally shows internal hiatus; /ai/ never shows this. This offers some support to the argument that morpheme-internal /ai/ can be a special type of unit. (2) At fast tempos, the second vowel [i] averages 51% of [a#i], but only 41% of [ai] ($P < 0.0001$; 30 tokens in each group), but slow tokens (without hiatus) show no cross-morphemic difference (Gore 2006b, p.160). However, the linguistic relevance of these findings is open to question since both depend on speech rate, and it is reasonable to view them as part of a more general potential for second high vowels to be weak and morpheme-final segments occasionally clipped, rather than as a sign that /ai/ within the morpheme is uniquely processed as a single unit.

Transition duration is statistically marginally indicative of morphemic status *regardless* of speech rate, but the two morphemic groups overlap so that 97% of tokens could belong to either group, and no transition exceeds 69ms or 34% of total VV duration (Gore 2006b). Amplitude data show no difference between Japanese [ai] and [a#i] at all. F0 data show an extremely wide variation from one token to another, and none of the monomorphemic stimuli obey Kubozono's (above) pitch-accent (High-Low) rule for "diphthongs." This agrees with assertions that pitch has a "low functional load" in Japanese (Vance 1987 p.107), and/or that pitch accents can be absent, neutralized or reversed according to context, speed and/or dialect (Hirayama 1985 p.37-69; Beckman 1986 p.35; Haraguchi 1988 p.123). There is thus no conclusive evidence for the existence of a consistent morpheme-internal diphthong (such as in English) defined by a special transition duration, frequency or amplitude.

3. Perceptual Studies

Using clapping and tail-catching methods, Gore 2006b (pp.171-188) shows that very young Japanese children employ a mixture of syllabic and moraic segmentation procedures for many V^1V^2 sequences, as is reported for long-V, VN and VQ (Hatano and Inagaki 1992), but that older (literate) children use moraic segmentation for all sequences. A possible developmental process is for the syllabic diphthong perceived by the very young to be reformatted as a dual monophthong following kana acquisition. The developed moraic segmentation strategy also seems to effect the perception of (second-language) English VV sequences (Gore 2006c): the segmentation unit used by Japanese students aged 12-15 shows the influence of the mora, but varies from one L2 English diphthong to another, suggesting that some diphthongs are perceived as two units ("moracally") while others are

perceived as one (“syllabically”). L2 English /ai/ and /ei/ tends to be perceived as a single unit, but /oi/ and /au/ are often segmented moraicly. To what extent this reflects the perception of corresponding sequences in their own language is an empirical question, but Japanese students’ unitary perception of each type of English diphthong seems in proportion to its frequency of occurrence in Japanese: /ai/ and /ei/ are common morpheme-internal VV sequences; /oi/ and /au/ are much less frequent and, in fact, absent from Sino-Japanese morphemes.

4. Some proposals

Phonological inventories tend to be symmetrical. The fact that the /-ai/ morpheme occasionally patterns like a single unit certainly sets it apart from many other V^1V^2 sequences in Japanese, but it is odd in a language that has a large number of such sequences to classify one of them as “occasionally a diphthong.” Since /ai/, an often proposed “possible” diphthong (see Gore 2006b, pp.3-5), is in most cases acoustically identical to /a#i/ in Japanese, there is little or no advantage in giving it a separate phonological category; it can be a cohesive vowel sequence according to where a break may occur (*before* or *after* the VV, rather than between the two Vs), but at normal speech rates it is usually acoustically indistinguishable from a sequence that is divisible.

In child phonological development, it is conceivable that the process of moraic segmentation develops *piecemeal*: generally reinforced by the acquisition of *kana*, but less so in some cases such as /ai/ (and /ei/, and the long vowels). In the /ai/ tail-catching tests summarized above, preschoolers showed mixed single-diphthong and dual-monophthong perception, with a tendency to perceive /(C)ai/ as a unit, but older schoolchildren preferred moraic segmentation; however, in cross-language perception by 12-15 year old students, English /ai/ was usually perceived as a single unit but /au/ was often segmented moraicly, suggesting influence from Japanese /au/ which does not exist as an SJ morpheme. We speculate that there are both morphological and acoustic reasons: (1) the existence of /ai/ as a common morphological unit that resists abbreviation (Gore 2006b, pp.91-95), and (2) a bias in favour of /ai/ as a combination of two acoustically stable vowels (Stevens 1989) which form an especially stable (prominent + non-prominent) pairing. The fact that these two characteristics are seen together in some sequences does not necessarily imply special phonological status.

The phonological development of Japanese children can be summarized according to VV perception: As Inagaki and Hatano (1992) report, and as our experiment on non-identical VV (Gore 2006b, pp171-188) suggests, preschoolers’ perceptions are *mixed*, i.e. both unitary and divided, but during and after the seventh year *moraic* (divided) segmentation dominates both perception and production. We propose a notional third stage that allows for two modes side-by-side: the standard moraic mode, and a subsidiary unitary mode whereby Vv (long vowels and some instances of /ai/ with

weak V²) are an undivided perceptual unit (often assigned, as noted above, to a single note in song). Here we use the “Vv” accent notation employed in Gore (2006b, p.119). Some of these Vv (accented vowel sequences such as /ai/ that are likely to be perceived as a unit) are morpheme-internal and thereby fulfill Kubozono’s definition of “diphthong” (1999; 2001b); others occur across a morpheme-boundary (Vance 1987). Thus an undivided perception of some VVs continues (or resurfaces), and the child also learns that some /ai/ sequences are especially cohesive in a morphological sense (mostly Sino-Japanese morphemes), though countable as two moras. There is, however, no evidence that perception of the syllable corresponds systematically (more than 50%) to these morphologically cohesive sequences, hence the problematic nature of “diphthong” as a phonological term.

The arguments and experiments summarized above leave several questions unanswered, which suggest paths for future research: (1) The morphological contrast in V² [i] duration in *fast* tokens calls for experiments on running speech at various tempos. (2) The developmental change in VV perception calls for experimental clarification of the roles of kana acquisition and age-related change, but it will be difficult to isolate these factors. (3) Japanese speakers perceived some English diphthongs (mainly /ai/) as more unitary than others, but to what extent this applies to VVs in their own language remains to be tested. A new experimental technique must be devised to test this.

The developmental changes we have found in VV perception challenge conventional rhythmic class distinctions, particularly the notion that segmentation strategies are fixed at an early age and uninfluenced by later developments such as orthography acquisition (Trubetskoy 1969; Jusczyk 1997; Ramus, Nespors and Mehler 1999). There is also a problem concerning the terminology for the rhythmic classes, “mora-timed, syllable-timed, and foot-timed.” The irregularity of the foot (Roach 1982) suggests that “time” is not the best device to sort languages into groups. A more usable yardstick might be whether the rhythmic grouping in the language is based on the perception of a unit of time, *or* on the perception of a sonority contour; i.e., whether the perceptual and productive segmentation strategy of the language-user is predominantly *time-based* or *sonority-based*. Just as there are two main approaches to segmenting music (beating time *vs.* choreographing the dynamics), we suggest there are two approaches to spoken language: the language user may count each CV (or similar small unit) or focus on the dynamics. This view, suggested by the present comparison of Japanese VVs and English diphthongs, may provide a useful framework for future studies of rhythm.

5. Implications for language acquisition

Roach writes that for native speakers of languages such as Japanese, Hungarian and Spanish, “which do not have weak syllables to anything like the same extent as English,” it can be helpful to “practise repeating strongly rhythmical utterances since this forces the speaker to concentrate on

making unstressed syllables weak” (Roach 2000 p.138); Dauer says that “Non-native speakers of English can improve their rhythm enormously by reducing unstressed syllables” (Dauer 1987). Celce-Murcia, Brinton and Goodwin (1996) suggest that the first question the TESOL teacher should ask about pronunciation is: “Is the student’s first language stress timed or syllable timed?” (1996 p.324). Though there are problems with this terminology, since Abercrombie (1967) it has been recognized that language rhythm is a central issue for teachers and learners. However, the issue of the perception and parsing of vowel sequences has been largely ignored. Since the Japanese unit is not a unit of sonority but of time, by which adjacent vowels are always processed separately, the effectiveness of the training to which Roach refers clearly depends on whether the learner understands the essential difference between mora and syllable.

Studies that have implications for language rhythm must also have implications for acquisition. This applies on two scales, large and small. Vowel-chart displays Gore (1996; 1998) show that (1) Japanese students’ English diphthongs often “jump over” the central area (representable as [a.i]); (2) they have difficulty pronouncing /ə/, both as a phoneme and as a transitional element in a diphthong, and (3) they acquire the schwa phoneme more easily after practising slow-gliding in diphthongs (representable as [aəi]) -- and vice versa; i.e. the diphthongal sequence and the /ə/ phoneme reinforce each other. Acoustic experiments (Gore 2005c) show that slow gliding is absent from Japanese both across morpheme boundaries and within the morpheme; and we know that schwa is absent from the Japanese phoneme inventory. These facts suggest that the process of acquiring the slow glide necessary to create an English-type /ai/ diphthong will present especially acute problems to the learner, both from a syllabic and a phonemic point of view. Conversely, English learners of Japanese must make special efforts to acquire an ability to pronounce sequences of short tense vowels, which do not exist in their mother tongue, *without* slow transitional gliding.

This paper has reviewed phonetic, phonological and psycholinguistic approaches to the VV. However, there are strong indications that the rhythmic contrast between the diphthong and the vowel sequence will, as a key aspect of language rhythm, be a fruitful area of research from the point of view of language acquisition, too.

References

- Abercrombie, D. (1967). *Elements of General Phonetics*. Edinburgh University Press.
- Allen, G. & Hawkins, S. (1978). The development of phonological rhythm. In Bell, A. & Hooper, J.B. (eds.) *Syllables and Segments*. 173-185. Amsterdam: North-Holland.
- Beckman, M. (1982). Segment duration and the ‘mora’ in Japanese. *Phonetica* 39: 113-135.
- Beckman, M. (1986). *Stress and Non-Stress Accent*. Dordrecht, Holland: Foris Publications.
- Celce-Murcia, M., Brinton, D. & Goodwin, J.M. (1996). *Teaching Pronunciation*. Cambridge University Press.

- Cutler, A., McQueen, J. & Robinson, K. (1990). Elizabeth and John: sound patterns of men's and women's names. *Journal of Linguistics* 26 (2): 471-82.
- Dauer, R. M. (1987). Phonetic and phonological components of language rhythm. *Proceedings of the 11th International Congress of Phonetic Sciences*, Tallinn, Estonia, 5: 447-450.
- Demuth, K. (2003). The status of feet in early acquisition. *Proceedings of the 15th International Congress of Phonetic Sciences*, Universitat Autònoma de Barcelona, 151-154.
- 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. (2003a). What is a diphthong? *Journal of the English Phonetic Society of Japan*. 6. 217-226.
- Gore, M. (2003b). 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.
- Gore, M. (2005a). Language rhythm and syllable type. *Journal of the English Phonetic Society of Japan*. 7. pp.61-71.
- Gore, M. (2005b). Rhythmic perception and the VV. *Journal of the English Phonetic Society of Japan*. 8. pp.77-85.
- Gore, M. (2005c). Vowel sequences in Japanese; diphthongs in English. *Proceedings of the International Congress of Phoneticians of English*. pp.154-163.
- Gore, M. (2006a). A cross-morphemic comparison of timing in VVs. *Bulletin of the Faculty of Education, Kagoshima University*. 57 (Studies in Education). pp.1-18.
- Gore, M. (2006b). *English and Japanese Diphthongs and Vowel Sequences*. PhD thesis. Department of Applied Linguistics, University of Reading.
- Gore, M. (2006c). Segmentation of English diphthongs by Japanese schoolchildren. *Journal of the English Phonetic Society of Japan*. (submitted).
- Haraguchi, S. (1988). Pitch accent and intonation in Japanese. Hulst, H.v.D. & Smith, N. (eds.), *Autosegmental Studies in Pitch Accent*. Dordrecht, Holland: Foris Publications.
- 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.
- Hatano, G. & Inagaki, K. (1992). *Yoozi no onin ishiki - sono 2: shiritori wo tooshite* [Phonological awareness in Japanese children 2: tail-catching]. 56th Annual Convention, Japanese Association of Psychology.
- Hirayama, T. (1985). *Zen-Nihon no hatsuon to akusento*. [The Dialects of Japanese]. *NHK-hen Hatsuon Akusento Jiten* (NHK Dictionary of Japanese Pronunciation) 37-69. Tokyo: Nihon Hoosoo Kyookai.
- Inagaki, K. & Hatano, G. (1992). *Yoozi no onin ishiki - sono 1: go no bunkai wo tooshite*. [Phonological awareness in Japanese children 1: word segmentation]. 56th Annual Convention, Japanese Association of Psychology.
- 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.
- Jones, D. (1967). *The phoneme*. Cambridge: Heffer.
- Jusczyk, P. (1997). *The Discovery of Spoken Language*. Cambridge MA: MIT Press.
- Kubozono, H. (1999). Mora and syllable. In Tsujimura, N., (ed.) *The Handbook of Japanese Linguistics*. 31-61. Oxford: Blackwell.
- Kubozono, H. (2001a). *Mora and Syllable*. Kobe University.
- Kubozono, H. (2001b). On the markedness of diphthongs. *Kobe Papers in Linguistics [Koobe Gengogaku Ronsoo]* 3, 60-73. <http://www.lit.kobe-u.ac.jp/linguistics/kubozono.pdf> Kobe University.
- Kubozono, H. & Matsui, F.M. (2003). Phonetic vs. phonological control of speech: closed syllable vowel shortening in Japanese dialects. *Proceedings of the 15th International Congress of Phonetic Sciences*, Universitat Autònoma de Barcelona, 2425-2429.
- Lehiste, I. (1972). The timing of utterances and linguistic boundaries. *Journal of the Acoustical Society of America* 51 (6): 1972-2024.

- 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.
- Stevens, K. (1989). On the quantal nature of speech. *Journal of Phonetics* 17: 3-45.
- Trubetsky, N.S. (1969) Principles of Phonology. University of California Press. Translated from (1939) *Grundzuge der Phonologie*. Transactions du Circle Linguistique de Prague.
- Vance, J.V. (1987). *An Introduction to Japanese Phonology*. Albany NY. State University of New York Press.
- Warner, N. & Arai, T. (2001). Japanese mora timing: a review. *Phonetica* 58: 1-25.