PERCEPTION OF MORA TIMING BY ENGLISH LEARNERS OF JAPANESE

Chiharu Tsurutani and John Ingram

Griffith University University of Queensland

ABSTRACT

English learners of Japanese at two levels of proficiency in L2 were tested for their ability to perceive moraic contrasts in Japanese vowels (long vs short) and stop consonants (single vs geminate) under listening conditions where the speaker and the rate of speech varied unpredictably on a short carrier phrase (*Sore wa* ____ *desu.*). The aim was to assess the acquisition of moraic timing by second language learners and the strategies they use for temporal normalization in the face of speech rate and speaker variation. Learners performance was compared with that of native speakers of Japanese. Results indicated that advanced learners were able to normalize speech rate, while beginners failed to do so and perceived moraic contrasts reliably only in slow speech. However, perception of moraic timing contrasts of consonants in slow speech was found to be difficult even for advanced learners.

INTRODUCTION:

Perceptual accommodation for variation in speech rate is a precondition for the reliable perception of phonological contrasts of quantity or duration in spoken language (such as distinctions between long and short vowels or geminate and non-geminate consonants). Native listeners probably acquire this facility at an early stage of first language acquisition as they attune their auditory system to the prosodic characteristics of the native language (Ioup & Tansomboon:1987, Jusczyk et al,:1993). Second language learners' ability to make appropriate accommodations for time-dependent phonological contrasts in L2 has not been studied in any detail, but is of considerable interest for practical pedagogy of second language teaching and for basic questions about the process of perceptual learning after the age of primary language acquisition. Prosodic transfer effects, or the projection of temporal normalization strategies used in the first language upon the L2 listening situation, appear in the initial stages of second language learning. But how open to modification are these temporal normalization strategies once set in place early in first language acquisition? And, assuming that they are modifiable by later L2 learning, what are the steps by which such re-learning takes place?

In one of the few studies to investigate temporal normalization by second language learners, Ingram and Park (1997) compared Korean and Japanese learners of English ability to perceive contrasts between short and long (tense and lax) front vowels in Australian English in the context of speaker and speech rate variation. Both groups varied in their prior exposure to English. The Korean listeners were also divided into older and younger groups. The age distinction amongst the Korean group was important because it reflects a dialect split that has developed in standard Seoul Korean post WWII, whereby a phonological contrast of vowel length has been lost by younger speakers. We found that Japanese and older Koreans (aged 50+ years) transferred their temporal normalization strategies for the perception of vowel length contrasts in L1 to the perception of length contrasts in Australian English vowels L2. The young Korean listeners on the other hand, responded to the absolute duration of the vowel stimuli, without accommodating for speech rate variation. In terms of ability to perceive the foreign vowel length contrast, the older Korean group with no significant exposure to English performed better than the younger Korean group with at least 12 months immersion in an Australian English speaking environment, and as well as the intermediate aged group (average age approximately 30 years) who were highly fluent in English and had several years of immersion in an English language environment. This experiment attests to the strength of L1 transfer effects when it comes to a basic perceptual accommodation strategy which is required for the perception of a phonological contrast such as vowel length or quantity.

In the present study, we examine the complementary case of Australian English learners accommodating to speech rate variation for the perception of moraic timing in Japanese. English and Japanese differ greatly in their word and phrase prosody. English words may not be said to have a moraic timing structure. Rather, spoken English words in phrases have a rhythmic structure that is usually described as foot or stress timed, complicated by the possibilities of vowel reduction. On the other hand, both languages have phonologically contrastive vowel length. Also, while English does not possess a general contrast between geminate and non-geminate consonants, such contrasts can arise at word boundaries and in some restricted phonological environments (e.g.: *in-numerable* [ɪnjuməɹəbl] vs. *un-numerbable* [ʌnnʌmbəɹəbl]).

In short, while there are prosodic differences between the two languages which might be expected to defeat any simple transfer of time-normalization strategies from English to Japanese, it might be predicted that English L1 strategies for the perception of vowel length contrasts may be readily adaptable to Japanese one versus two mora vowel contrasts, but that the perception of consonantal moraic contrasts would be difficult for English learners in the face of speaker and speech rate variation. A second goal of the study was to observe the possible effects of Japanese *pitch accent* pattern upon the perception of moraic contrasts by English learners. It was hypothesised that English listeners may respond to pitch accent contrasts as *word stress* contrasts and that this may have consequences for the perception of moraic timing by English listeners. Consequently the pitch accent pattern in the Japanese stimulus words was systematically varied in the experiment, along with their moraic structure, the rate at which the items were spoken (two rates), and the identity of the speakers (three native speakers: two female and one male).

METHOD

The subjects were asked to listen to the tape and choose the appropriate transcription from possible three choices CVCV, CVVCV and CVQCV which were written in Japanese orthography, kana, depending on the sound they heard. The test was conducted in a quiet room and the tape was played without interruption.

Subjects

Ten subjects each from three groups, native listeners, advanced learners and beginners, participated in the experiment. Beginners have been studying in the second year basic Japanese subject and advanced learners have been studying Japanese for 6~7 years, including one-year stay in Japan. They were in their second year at Griffith University at the time of experiment. The ten native speakers were Japanese university students who were visiting Australia to attend an ESL programme for a few weeks.

Stimuli

"kiko" and "sote" were chosen, as a base form, CVCV word. Gemination and vowel lengthening were then inserted between two syllables (CVQCV, CVVCV) bearing in mind the following considerations:

- Out of five Japanese vowels the high front vowel [ ] and a non-high back vowel [o] were chosen. A vowel at word final position did not receive any restriction for its selection because it does not affect gemination or vowel lengthening.

- Voiceless consonants were used, which are a crucial factor for the occurrence of gemination, and also for the convenience of acoustic analysis.

By assigning the pitch pattern HL , LH and HH to six words, fourteen patterns of nonsense words were made. "sotte" is a real word (one conjugated form of verb "sou") but it is not part of the vocabulary of the subjects who participated in this experiment.

kiko	(HL / LH)	sote	(HL / LH)
kikko	(HL / LH)	sotte	(HL / LH)
kiiko	(HLL / HHL / HHH)	soote	(HLL / HHL* / HHH)

H = high pitched mora, L = low pitched mora

* Although this pitch pattern does not exist in Japanese, it was included in the list in order to see the influence of pitch change on the duration of the first heavy syllable.

- The HH pattern in CVCV and CVQCV and the LH pattern in CVVCV do not occur in Japanese.

The words were read by three native speakers T, Y (female) and N (male). They were in their thirties and all spoke standard Japanese. A stimuli tape contained 84 items made up of the fourteen items spoken by three speakers at two different speaking rates, slow and fast. The speaking rates were controlled individually by each speaker according to the following guideline provided: Fast = the speed used for chatting with Japanese friends Slow = the speed used for talking to beginners in first year Japanese class

The items were presented in the carrier sentence "Sore wa....desu.(It is....)" , since isolated words presented at randomly varying speech rates would not constitute a realistic listening situation. It also makes it easy for subjects to accommodate speaker variation. After recording, the best reading for each item was chosen using a digital wave form editing programme " Sound Recorder " and presented twice, with an inter-stimulus interval of one second and an inter-trial interval of two seconds.

RESULTS

Overall level of performance

The performance of the native listener group was essentially error free. None of the Japanese listeners had difficulty accommodating to the speaker and speech rate variation. The advanced learner group also made few errors of moraic perception.

The likely cause for the low overall incidence of timing errors was that the speech rate manipulation did not oblige subjects to adopt a time-normalizing listening strategy for most of the items. Despite the speech rate manipulation, it proved possible to correctly classify most of the stimuli into the three moraic response categories: VC (mono-moraic vowel and consonant), VCC (geminate consonant), VVC (bimoraic vowel) on the basis of absolute duration of the medial consonant and vowel (see Fig. 1.).

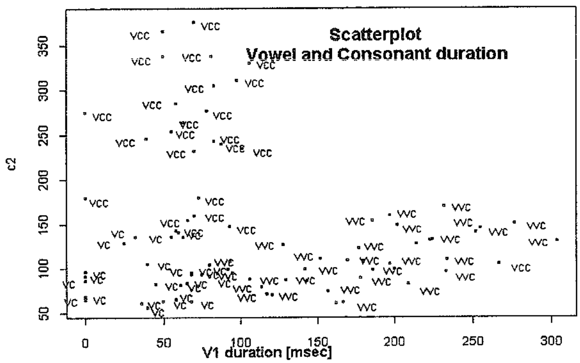

Fig. 1 V1/C2 plot with output of Gaussian classifier.

Thus, only a small proportion of the stimuli required listeners to make appropriate speech rate compensations in judging moraic length. Nevertheless, there were statistically quite robust differences in the rate of moraic errors between the beginning and the advanced learner groups.

Specific factors affecting perception of mora timing

A series of analyses of variance were conducted to test for specific factors affecting the perception of mora timing. For these analyses, the unit of observation was a given stimulus token which was associated with two mora perception error scores, one for the beginning and another for the advanced learners summed over the subjects in each group, for a maximum possible error score of 10/10 and a minimum of zero.

Pitch accent pattern: Pitch accent pattern apparently had no affect either singly or in combination with other factors for the perception of mora timing. No main or significant interaction effects were found for this variable.

Speaker: No significant main effect of speaker was found. However, there was a significant speaker x speech rate interaction (F = 4.35, d.f. = [2, 162] p=.015), which is not of particular interest for the present discussion.

Mora type (VC, VCC, VVC): There was no significant main effect for mora type upon the rate of errors of mora perception, but there were two significant interaction effects: a) a highly significant interaction of mora type with rate of speech (F = 13.71, df = [2, 156], p<.00001) that was entirely predictable (see Fig. 2), and b), a significant interaction between mora type and level of Japanese proficiency (F = 9.14, DF = p<.0002,) discussed below (see Fig. 4).

Level of Proficiency: There was a highly significant main effect for Proficiency Level (F = 28.07, df = [1,156], p<. .000001) and interaction effects for: c) Proficiency Level with Speech Rate (F = 5.22, df = [1, 156], p<.05), and, Proficiency Level with Mora Type (previously mentioned as b) above). The main effect of Proficiency Level indicated that the advanced learner group performed better than the beginner group on correctly identifying mora timing. There was a small main effect of Speech Rate suggesting that performance was better at the slow rate of speech. However, there was an overriding interaction effect between Proficiency Level and Speech Rate indicating that any differences in accuracy of mora perception associated with speech rate were entirely attributable to the beginner level group making fewer errors under the slow speech listening condition. The advanced learners performed at ceiling for both speech rates (see Fig. 3)

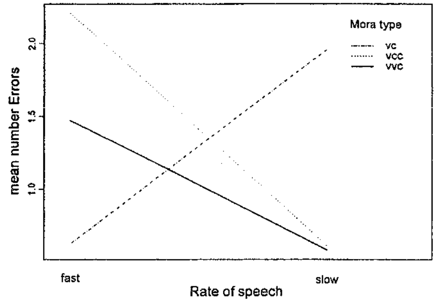

Fig. 2. Interaction plot: Speech Rate and Mora type

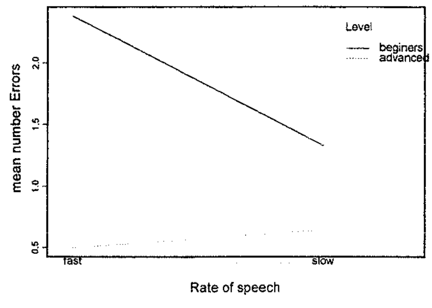

Fig. 3 Interaction plot: Speech Rate and Proficiency level

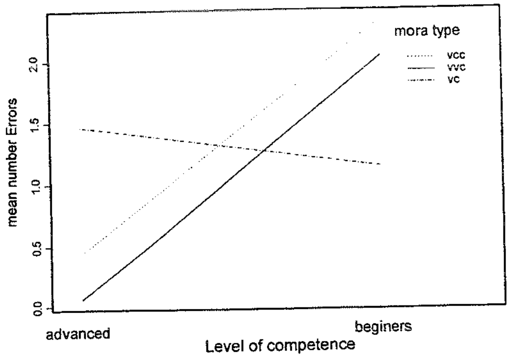

Fig.4 Interaction plot: Proficiency Level and Mora type

DISCUSSION

Interest centers on the interaction effects. The interaction of speech rate and mora type (Fig.2) indicates, as expected, that mono-moraic (VC) segments will tend to be perceived by learners as bimoraic at slow speech rate and that bi-moraic segments tend to be perceived as mono-moraic at fast speech rate. This is what would be predicted if listeners were failing to adopt a rate-normalizing listening strategy (judging segments on the basis of absolute duration). The interaction of Speech Rate and Proficiency level (Fig. 3) shows that while the beginner group performed better at slow speech rate in the perception of mora timing, the advanced group was at ceiling performance for both slow and fast rates of speech.

The interaction effect which merits most attention is the interaction of Proficiency Level and Mora Type (Fig. 4). It indicates that while perception of mora length for the bi-moraic stimuli (VCC and VVC) is more accurate for the advanced learners over the beginners, there is no corresponding performance increase for perception of mora timing in mono-moraic (VC) stimuli. In fact, there is a small (but not statistically significant) *increase* in errors of mora timing perception for mono-moraic stimuli by the advanced group. An analysis of the error patterns that typified the two learner groups was undertaken to clarify the mechanism underlying this unexpected finding.

Error Analysis

Two error patterns were particularly noticeable in each group of learners. a) bi-moraic CVQCV and CVVCV stimuli in fast speech were misperceived as mono-moraic CVCV by many beginners (Fig5(5) and 5(6)); and b) mono moraic CVCV items tended to be heard as CVQCV particularly in *slow* speech (Fig. 5(1)). The large number of Pattern a) errors suggests that beginners rely on the length of target phonemes to judge moraic length and cannot adjust their perception to a different speech rate. That is, beginners' perception criteria for moraic timing do not shift from those used to perceive slow speech first encountered at the beginning of acquisition. As a result, long vowels (C<u>VV</u>) and geminate consonants (CV<u>CC</u>) in fast speech are perceived as short (CV).

The pattern b) errors in which CVCV words tended to be misperceived as containing geminate consonants are interesting precisely because they are more prevalent in advanced learners. It is expected that this error would be more prevalent in the slow speech condition if listeners were not applying rate compensation in their judgement of mora length. But failure of rate compensation is unlikely to be the cause of error in the case of advanced learners.

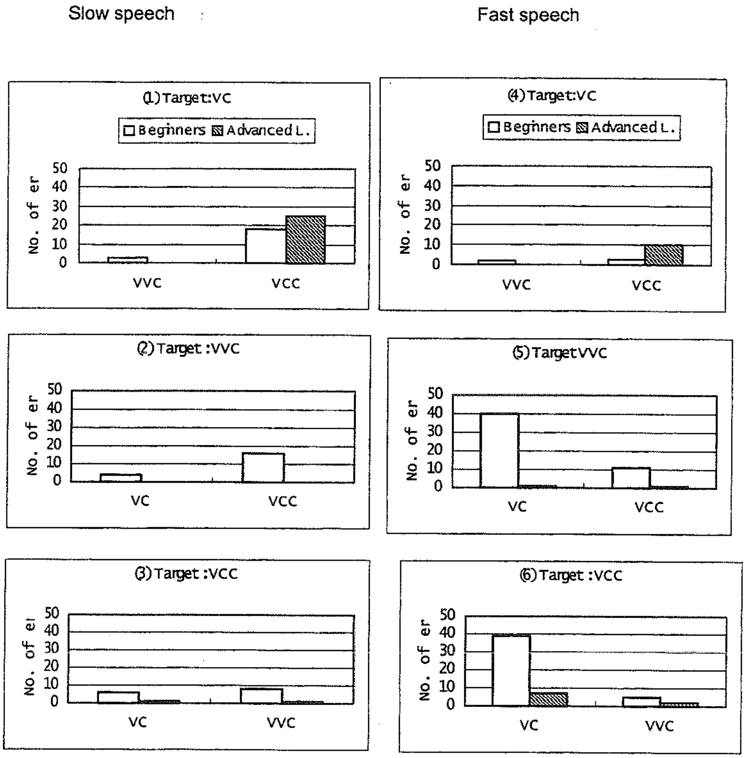

Fig.5 Patterns and occurrences of duration misperception in fast and slow speech

We interpret this greater incidence in the tendency to perceive geminate consonants as arising from a heightened awareness of their likely occurrence on the part of the more advanced learners. Language learners are thus more likely to make false positive identifications of geminate consonants, which are more difficult for English learners to perceive than geminate vowels. We also conclude that strategies for speech rate compensation are probably more fragile for geminate stops than for long vowels. This is consistent with findings mentioned in the introduction to this paper where it was found that first language temporal normalization strategies generalized to the perception of phonetically related contrasts in a foreign language.

REFERENCES

Ingram, J & Park, S (1997) Cross-language vowel perception and production by Japanese and Korean learners of English, *Journal of Phonetics* 25, 343-37

Ioup, G. & Tansomboon, A. (1987) The acquisition of tone: A maturational perspective in Ioup and Weinberger (eds.) *Interlanguage Phonology* 333-349

Jusczyk, P. W. (1993) How word recognition may evolve from infant speech perception capacities in Altmann Shillcock (ed.) *Cognitive Models of Speech Processing* 27-55