Acoustic speech analysis of second language English stress system

MAMIKO AKITA

University of Durham, UK

ABSTRACT

L2 learners' timing of the falling pitch and how they mark word prominence with prosodic features is investigated based on longitudinal production data collected from three adult Japanese EFL learners. Data show that although the learners' prosody appeared to be distinctly non-native throughout the data collection, they did exhibit changes over time: two of the subjects started to produce reduced vowels and their timing of the falling pitch became target-like. It is suggested that the acquisition of the reduced/neutral vowel is the key to acquisition of prosodic features (i.e. timing) of English, revealing the interaction of a segmental process and a prosodic process. It is concluded that L2 learners follow the developmental sequences observed in L1 acquisition.

1. PRELIMINARIES

The area of phonology for which fossilisation seems to be more apparent than elsewhere in second language acquisition, namely, prosodic features, is the focus of this paper. Leather and James' (1991) study shows that transfer persists well into advanced stages of acquisition with respect to prosodic features. Akita (1998, 1999) confirms this proposal with perception data from Japanese EFL learners.

Languages are held to involve differences in how word prominence is marked, and are often given typological labels such as 'stress language', 'pitch accent language' and 'tone language'. However, this simple classification has been found rather misleading: simple distinctions between stress languages and pitch accent languages have been difficult to find at the phonological level (Haraguchi 1994, Halle and Vergnaud 1987); McCawley (1987) rejects the sim-
ple classification of tone languages and pitch accent languages based on a discussion of Japanese and other languages.

Sugito (1969) argues that English and Japanese share a similarity in that ‘pitch’ plays a crucial role in marking stress or accent. No doubt, Japanese accents are marked with pitch, but what is worth mentioning is that it is the movement of the falling pitch (rather than the relative pitch height) which marks accent. Despite the general impression that stress languages make use of stress (loudness) to mark accent, this is actually a mixture of at least four acoustic parameters: speech power, fundamental voice frequency, phonetic quality and duration. And at least in English, pitch more than loudness or duration is likely to be responsible when a syllable is perceived as ‘stressed’ or ‘emphasised’ (e.g. Lieberman 1960, Sugito 1969, Fry 1958).

Although Japanese and English share similarity in that they use pitch to mark accent, there are several features which differentiate the prosodic features of English and Japanese. Although both languages use pitch to mark accent, the differences lie in the actual timing of the falling pitch in both languages. In English, for a word with penultimate accent such as *summit* (HL), the falling pitch starts from the first syllable, while in Japanese, for a word with penultimate accent, such as *kame* (HL) ‘turtle’, the falling pitch starts from the second syllable (see Figure 1).

There are two other features which differentiate English and Japanese. English has a tendency to shorten unaccented syllables. Parmenter and Trevino (1935) reported that accented syllables are longer than accentless syllables by 50%. Because the vowel inherently is the longest segment in a syllable, it is the portion of the vowel, rather than the consonant, that is susceptible to reduction. On the other hand, in Japanese, the vowel does not differ in length depending on whether it is accented or not. In addition, in English, certain vowel quality differences have particular significance: the centralisation of a vowel, e.g. substitution with the neutral vowel [ə] is observed for unaccented syllables such as in *photograph* [fəətəɡraɾf] (cf. *photo* [fətəʊ]; Japanese does not exhibit this difference.

The objective of this study is to investigate whether L2 learners have the falling pitch appropriate for English (Question 1, henceforth Q1), and if not, whether learners show any changes over time (Q2); in addition, the study investigates whether learners produce accented/unaccented syllables with appropriate

---

1. Please note that in Figure 1, [s] and [t] in *summit* and [k] in *kame* do not actually show up on the pitch contour. I have added those segments in the labelling for the ease of comprehension.
length and quality (Q3), and if not, whether the learners exhibit any changes over time (Q4). To answer these four questions, L2 learners’ timing of the falling pitch and how they mark word prominence with prosodic features was acoustically investigated based on longitudinal data collected from three adult Japanese EFL learners.

The remainder of this paper is organised as follows: in section 2, the methodology of the study is presented. In section 3, results of the data collection are reported. In section 4, I first summarise studies in L1 acquisition with respect to prosodic features (4.1), and compare them with the data in L2 acquisition (4.2). In section 4.3, I summarise the findings of the study along with possible future directions.

2. METHODOLOGY
The subjects for this study were three Japanese females (referred to as NI, MO and SK) at a U.K. university, either reading for a postgraduate degree or enrolled in a university exchange student scheme. In addition, I had one Japanese native speaker (JNS) and two English native speakers (ENS) as controls.

The L2 learners had had English from six to nine years at school and university in Japan before they arrived in the U.K. None of the subjects had had native-speaking teachers except for occasional visiting fellows. The main method of instruction was grammar translation. They were first recruited from a pre-sessional intensive English programme where they were enrolled to improve their English to reach the university requirement of IELTS 7.0. All were at the same placement level, i.e. high intermediate, in the programme. The age of the subjects ranged between 21 and 30. NI was on a three-month course and MO and SK were on the two-month course. NI was an undergraduate student on a 12-month exchange programme. She had no previous experience of living abroad. MO was a graduate student on a 14-month exchange programme. She had lived in the U.S.A. for a year when she was 13 years old, but reported that her family lived in a Japanese-speaking community and had very little contact with native speakers of English and she went to a Japanese school. SK was a graduate student on a 12-month masters course. She had lived in the U.S.A. for 2 years when she was 14 years old.

The data were obtained in a series of interviews during which the subjects met individually with the investigator. Meetings took place at the investigator’s home in a relaxed atmosphere. Data collection sessions were held at approximately one-month intervals for all subjects for over a year. Subjects performed several tasks testing their perception and production ability in English. Speech production was tape-recorded using a Sony WM-D6C. The first data collection
session was held at the earliest date the subjects were available after they arrived in the country (average two weeks). Production data involving 6 nouns (e.g. parapet, advocate, calibre) and 6 verbs (e.g. titrate, sever), are investigated in this paper.

Several factors were examined to determine the learners’ timing of the falling pitch and how they mark prominence: 1) their pitch contours were extracted, 2) the values of the first three formants, and the amplitude for each vowel were measured, 3) vowel and consonant duration was measured, and 4) the ratio of vowel to syllable duration was established.

Analyses were conducted either on a Pentium II 30 MMX workstation or on a laptop computer IBM ThinkPad 560 E, using either Multi-Speech Model 3700 version 2.2 or WinSAL-V Version 1.2 speech analysis software.

For vowel duration, the criteria used in determining the beginning and the end of a vowel were similar to those described by Peterson and Lehiste (1960). In general, vowel onset and offset criteria included the initiation and cessation of voicing from waveform displays; defined as the interval between the onset (zero crossing) of the first fully voiced glottal period and the end of the final glottal period of the target syllable. When vocalic onset was difficult to determine, information from the time-synchronised spectrogram was utilised.

For measuring vowel formants, I referred to Peterson and Barney (1952) quoted in Kent and Read (1992) for English vowels and to Imaishi (1980) for Japanese vowels. A wide band filter with a nominal bandwidth of 300 Hz was employed to determine the first, second, and third formant frequencies at the mid-point of the vowel, displayed with LPC formant tracts superimposed on the spectrographic display. Diphthong formants were measured during the initial steady portion and at the termination of the diphthong. In most cases, LPC tracks indicated accurate detection of the first three formants. From the wide-band filter, consonantal characteristics were also examined.

A 50 Hz narrowband filter was used to study voice F0 contours. In addition a computer system calculated the fundamental frequency of phonation between points in the waveform. The calculated fundamental frequency of phonation was displayed on the computer.

3. RESULTS

At the beginning of data collection, it was found that the learners’ timing of the falling pitch was somewhat late, suggesting transfer of their Japanese timing (Q1). Compare, for example, the English native speakers’ data in Figure 2,

2. I have no good explanation as to why word initial [k] in Figure 2 (English native speaker data) and Figure 7 (MO’s data at month 12) show voicing.
where the falling pitch started at the beginning of the first syllable, with the Japanese native speakers’ data of the Japanese word kamera ‘camera’ (Figure 3) where the falling pitch started from the second syllable. All three learners transferred their Japanese timing; for example, NI and SK’s data at Month 0 for the production of calibre show that falling pitch started from the second syllable (Figures 4 and 5). This tendency was found for all of the test words.

Although learners’ prosodic features appeared to be distinctly non-native throughout the data collection, they did exhibit changes over time: after 5-7 months of exposure to native input, NI and MO’s timing of the falling pitch became target-like for English (Q2). For example, see Figure 6 [NI (Month 6) calibre] and Figure 7 [MO (Month 12) calibre]. SK continued to show transfer from her Japanese timing.

In addition, at the beginning of the data collection, all three transferred their L1 syllable structure and produced extra syllables by epenthesis. They also had a non-native like tendency to overemphasise unaccented syllables, i.e. producing syllables with amplitude and duration values more appropriate to accented than to unaccented syllables (Q3). For example, for the test word advocate, SK (Month 1) and MO (Month 0) inserted vowels after the coda consonants [d] and [t] and produced a form such as [a.d.o.vo.ker.ɪ.tu]. After 5-7 months of exposure to native input, NI and MO started to produce reduced vowels (Q4), e.g. NI (Month 7) and MO (Month 5): [æd.vɔ.kɛt]. However, SK, on the other hand, exhibited very little change: she never succeeded in producing reduced vowels, e.g. SK (Month 12): [ad.vɔ.kɛt]. To discuss the relevance of the acquisition of timing and the acquisition of reduced vowels, we first turn to the literature on L1 acquisition of prosodic features.

4. DISCUSSION
4.1. The L1 acquisition of prosodic features
It has been well documented in the L1 literature that children move from unmarked to more marked structures in the course of acquisition, guided by the Subset Principle (Berwick 1985). For the acquisition of syllable structure, Carreira (1991) and Lleó and Prinz (1996) suggested an acquisition order as fol-

---

3. In Figure 4, 5, 6, learners didn’t have aspiration in their production. This is clearly an interference from their Japanese which has only negligible aspiration in word initial voiceless obstruents.

4. Japanese has a fairly simple syllable structure. It does not allow any consonant clusters either in the onset or in the coda. The only coda consonants allowed in Japanese are /N/ (moraic nasal) and the first member of a geminate.
lows. Children start with CV syllables; the presence of CVC syllables implies the presence of CV. Acquisition of a coda consonant (CVC) (nasals>sonorants, and then others) precedes the acquisition of consonant clusters and coda clusters are acquired first (CVCC>CCVCC). Therefore, CV>CVC>CVCC>CCVCC. See also Demuth (1995) who captures the course of acquisition in terms of Prosodic Hierarchy (Selkirk 1984, Nespor and Vogel 1986) making use of Optimality Theory (Prince and Smolensky 1993).

As has been suggested by Fikkert (1995) among others, the development of stress (accent) goes hand-in-hand with the development of syllable structure. One of the first rhythmically important skills the child must learn in order to produce fluent English phrases is that of producing adult-like weak syllables (Allen and Hawkins 1980). Young children’s ability to reduce weak syllables adequately develops much later than their production of stressed (accented) syllables. Two year olds tend to use far fewer reduced syllables than do adults, so that their speech rhythm has fewer syllables per foot, or more beats per utterance: in short, it sounds more syllable-timed. Allen and Hawkins present data showing children acquire heavy syllable nuclei by the age of three, yet, the acquisition of light nuclei was delayed until the age of four or five.

4.2. The L2 acquisition of prosodic features

The L2 data in my study lend support to the idea that L2 acquisition is a process in which the same mechanisms as those found in L1 acquisition are at play. Certainly the ability to reduce weak syllables is essential in acquiring English rhythmic pattern, regardless of whether the learner is a child or an adult. SK failed to reduce vowels, which explains why she exhibited very little change overtime in accentuation. NI and MO succeeded in acquiring vowel reduction, which made it possible for them to proceed to get the timing correct for the target language. It should be noted that the data show that acquisition of reduced vowels preceded the acquisition of timing.

Was the L2 order for syllable structure like the L1 order? It was found that subjects started with CV syllables. The acquisition of CVC syllables followed. For example, si.no.pu.si.su [SK (Month 1)] > si.nop.sis [SK (Month 9)]. In addition, the acquisition of CVC preceded acquisition of CCVC. For example, NI data: a.bu.su.tein (Month 0) > a.bus.tein (Month 4) > ab.stein (Month 9). However, from the data examined in this paper, it was unclear if coda clusters were acquired before onset clusters (CVCC>CCVCC).

5. Transfer of their Japanese syllable structure, not necessarily due to the Subset Principle.
4.3. Concluding remarks

In this paper, adult Japanese L2 learners were studied as they continued to develop their L2 in a target language setting. The nature of phonological acquisition and the effect of such input to learners over time was examined.

Data show that although the learners’ prosodic features appeared to be distinctly non-native throughout the data collection, they did exhibit changes over time: two of the subjects started to produce reduced vowels, and their timing of the falling pitch became target-like.

Two main conclusions are reached in this paper. It was suggested that the acquisition of the reduced vowel is the key to the acquisition of prosodic features of English. It was also shown that L2 learners follow the developmental sequences observed in L1 acquisition.

For future research, it will be interesting to investigate whether differences exist in the way various learners perceptually process the phonology of an L2, and whether such differences can account for varying degrees of success in L2 speech production.

ACKNOWLEDGEMENTS

I am indebted to my supervisor, Martha Young-Scholten for her invaluable guidance and encouragement at every stage of the study. Special thanks to S.J. Hannahs for his expertise in acoustic phonetics. I gratefully acknowledge NI, MO and SK for their co-operation throughout the process of the data collection.

REFERENCES


APPENDIX: FIGURES

Figure 1.
English NS summit vs. Japanese NS kame "turtle"

![Graph showing frequency over time for English NS summit and Japanese NS kame.]

Figure 2.
English NS calibre

![Graph showing frequency over time for English NS calibre.]

Figure 3.
Japanese NS kamera "camera"

![Graph showing frequency over time for Japanese NS kamera.]

Figure 4.
SK (M0) calibre

![Graph showing frequency over time for SK (M0) calibre.]
