

TASK TYPE EFFECTS ON THE PRODUCTION OF VOICE ONSET TIME (VOT) BY JAPANESE LEARNERS OF ENGLISH

Keiko Sato

Waseda University, Tokyo, Japan

keiko_hotaru@fuji.waseda.jp

ABSTRACT

This research investigated effects of task types on the acquisition of English voice onset time (VOT) by Japanese university students with a short-term but intensive exposure to academic English. The data were collected from 11 Japanese learners of English as a foreign language (EFL) receiving English-medium instruction at a university level, 11 Japanese EFL learners enrolled in regular EFL courses at the same university, and 11 native English speakers in two different conditions: a sentence-reading task and a picture description task. Results show that the Japanese speakers who had received English medium-instruction for a few months produced voiceless stops with more English-like VOT values only in the sentence-reading task, but not in the picture description task. This may imply that although the short-term intensive exposure affects their VOT shift toward more English-like values, the task type differences may also play an important role in interpreting second language VOT acquisition.

Keywords: second language acquisition, VOT shift, task type differences, English-medium instruction, English as a foreign language

1. INTRODUCTION

Acquisition of second language (L2) speech sounds requires a considerable period with massive exposure to an L2; however, recent studies show that a segmental shift toward the target language could be possible with relatively short-term exposure. In addition, task type differences are reported to affect segmental production. This research investigated effects of task types on the production of VOT in English by Japanese speakers with short-term but intensive exposure in university-level English-medium instruction.

Cross-linguistic studies have revealed that VOT represents a language-specific feature. Lisker and Abramson [8] reported three modal categories along the VOT continuum: voiced, unaspirated voiceless, and aspirated voiceless. In English,

aspirated voiceless stops are realized as “voiceless” stops. The mean English VOT values of initial /p, t, k/ were reported to be 58ms, 70ms, and 80ms in isolation, and 28ms, 39ms, and 43ms in sentences, respectively [8]. Contrary to English, in Japanese, “voiceless” stops are represented as unaspirated stops. The mean VOT values for initial voiceless stops in Japanese are /p/ = 24ms, /t/ = 32ms, and /k/ = 45ms [5].

The language-specific feature also allows researchers to acoustically compare the production of VOT by bilingual and monolingual speakers. According to the Speech Learning Model (SLM) proposed by Flege [2], late L2 learners are more likely to create a new “merged” L2 category, which may deflect away from both L1 and L2 categories to maintain phonetic contrast between the two languages. Harada [4] showed that English-Japanese bilingual children in an immersion program produced Japanese voiceless stops with significantly longer VOT values than monolingual Japanese children did. Flege [1] reported that late L2 learners produced English /t/ with “compromise” VOT values (33ms) that corresponded to neither English (51ms) nor Spanish values (18ms).

Although many previous studies have examined acquisition of voiceless stops in L1 and L2, very few studies focus on the acoustic measurement of phonetic variations caused by different task types. Participants recruited in the studies above [1, 5, 7] were asked to read words from a list, either in isolation or in a carrier phrase. The task type differences have been overlooked in spite of the fact that non-acoustic studies have traditionally documented its effect. Sato [12], in her earlier study of the phonetic variation in the production of final consonants, observed significant task-related variation. Comparable findings were obtained by Flege, Takagi and Mann [3], who found that English /r/ and /l/ by Japanese speakers were rated significantly less correctly in the spontaneous speech task than in the definition and reading tasks. These results may imply that nonnative

speakers vary the degree of attention to their pronunciation of English, depending on the nature of tasks.

Although VOT acquisition requires a considerably long time, the shift of VOT has been reported to occur after “short-term” exposure to an L2. Harada [4] suggested that it occurred within an early stage of learning and little development was entailed at a later stage. Sancier and Fowler [11] also showed that a late learner of English had shifted her VOT values significantly in her English and Portuguese voiceless stops due to the changes in her language environment. These studies imply that L2 learners may be able to produce context-dependent VOT values significantly different from their native language’s.

The primary goal of the current study is to examine the production of English VOT by Japanese learners of English in an EFL context. Specifically, the study addresses the following research questions: 1) Does an intensive but short-term exposure in English-medium instruction at a university cause students’ VOT to become more English-like? 2) Do task differences affect the production of VOT?

2. METHOD

2.1. Participants

The participants were 11 Japanese EFL learners receiving English-medium instruction, 11 Japanese EFL learners in a regular university program, and 11 native English speakers, all of whom were students at a private university in the metropolitan area of Tokyo, Japan. The Japanese speakers had never spent any time abroad for more than 2 months, had never received any instruction of English phonetics and pronunciation, and had not had any additional exposure to English outside of school. The only differences between the two Japanese groups were the amount of exposure to English and the medium of the instruction for the last 19 weeks (216 hours on average in English for the English-medium instruction group vs. 41.6 hours on average in both English and Japanese for the other group in the traditional curriculum). The participants in the English-medium instruction group were both taught by both native English and nonnative English-speaking content teachers, while the participants in the EFL regular university program were taught by only Japanese EFL teachers.

The English native speakers were exchange students who had recently come to Japan, and reported that the language they predominantly used in Japan was English.

2.2. Procedures

VOT data were elicited under two different conditions: the picture description task and the sentence-reading task.

First, the participants were asked to make a narrative story based on a cartoon. After completing the picture description task, the participants engaged in the sentence-reading task. The speakers were asked to read sentences in which the target words were embedded into a carrier phrase.

The same procedures were applied to all the participants except for the language used for the instruction; Japanese for the Japanese speakers and English for the English speakers.

2.3. Materials

In selecting the target words, I followed the same protocol as in Harada [4]. The words selected were all disyllabic with the vowel of the same quality /æ/ following the word-initial voiceless stops and the stress on the first syllable.

In the picture description task, the cartoon story the participants described had been designed to elicit three words with word-initial voiceless stops (*panda*, *taxi*, *candy*). The sentence-reading task had been designed for the participants to read each word three times, inserting it in a carrier phrase, *say _____ once again*. The words chosen in the study were as follows:

/p/	/t/	/k/
panda	tablet	candy
parrot	taxi	carrot

2.4. Data measurements

Each word was recorded twice in the picture description task and three times in the sentence-reading task for each participant. The total number of tokens was 198 for the picture description task (3 words x 2 repetitions x 33 participants). For the sentence-reading task, there were 594 tokens (6 words x 3 repetitions x 33 participants).

VOT in initial stops was measured to the nearest millisecond from the beginning of the occlusion burst to the onset of voicing energy in F2 formant. The secondary information was the waveform; VOT was measured from the beginning

of the release burst to the first positive peak in the periodic portion of the following vowel.

In the picture description, the speech rate varied significantly according to groups, which was reported to affect their VOT production [9]. Therefore, only for the picture description task, estimated VOT values obtained from ANCOVA, in which the speech rate was a covariate, were applied to conduct the following statistical procedures.

3. RESULTS

The mean VOT values produced by the Japanese learners of English who received English-medium instruction (EMI), the Japanese learners of English in the regular university English program (EFL), and the native English speakers (NE) in the sentence-reading task and the picture description task are shown in Figure 1 and Figure 2, respectively.

Figure 1: The mean VOT values of English voiceless stops by NE, EMI, and EFL in the sentence reading task. The error bars enclose +/- one standard error.

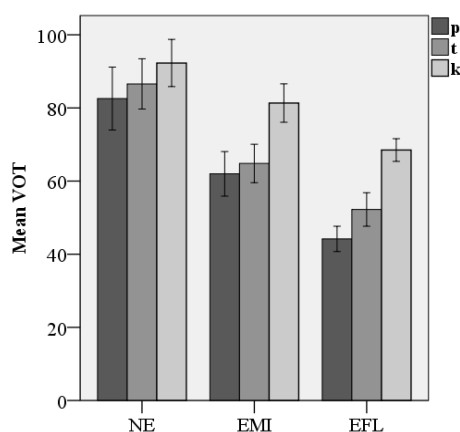
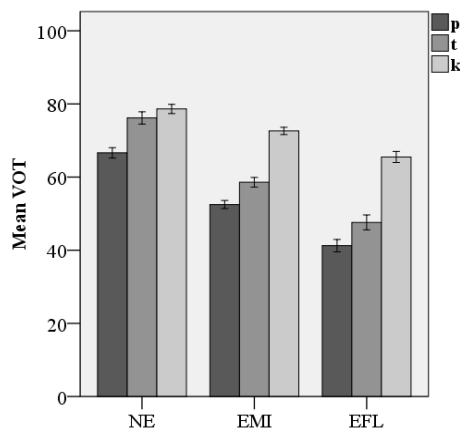


Figure 2: The mean VOT values of English voiceless stops by NE, EMI, and EFL in the picture description task. The error bars enclose +/- one standard error.



The mean VOT values in both tasks were submitted to a (3) Group, (3) Place of Articulation and (2) Task Type repeated measures ANOVA, which yielded the significant main effects for Group [$F(2, 60) = 6.57, p = .01$], Task [$F(1, 60) = 6.57, p = .013$], but no significant Group x Task interaction [$F(2, 60) = 0.759, p = .473$]. Bonferroni post hoc tests revealed that for the sentence-reading task, the NE produced voiceless stops with significantly longer values than the EMI, and the EMI's VOT values were significantly longer than the EFL's. This shows that, compared with the EFL group, the EMI group has reached the more English-like VOT norm.

For the picture description, however, Bonferroni post hoc tests indicated that the NE's mean VOT values were significantly longer than those of both the Japanese groups ($p < .05$), but there was no significant difference in the mean VOT values between the EMI and EFL groups ($p = .088$). This suggests that in the picture description as opposed to the sentence-reading task neither the EMI group nor the EFL group showed any development.

The post hoc tests also revealed that the NE's VOT values were significantly longer in the sentence-reading task than those in the picture description task. On the other hand, the Japanese speakers' VOT values (EMI and EFL) showed no significant differences in both tasks.

4. DISCUSSION

Despite a small number of the participants, this study suggests that the Japanese learners of English who had received English-medium instruction for a relatively short time at the university level showed some development toward more English-like in the production of VOT, compared to the Japanese learners of English in regular EFL courses. However, their development was limited to the sentence-reading task; in the picture description task, their VOT values ended up being less different than the EFL group's.

The EMI group's success in producing more target-like VOT indicates that the short-term but intensive exposure to English caused a VOT shift in their voiceless stops. This findings support the studies on the acquisition of VOT in Japanese [5], English, and Portuguese [11]. However, it is worth noticing that the VOT shift only occurred in the sentence-reading task, and not in the picture description task. In other words, the short-term but intensive exposure to a second language may lead

to the VOT shift; however, the shift can only be observed in a relatively simple task such as sentence reading. When learners with a substantial amount of exposure to a second language perform a cognitively demanding task such as picture description, they have to distribute their attention not only to the phonological domain, but also to other domains such as semantic domains and end up producing voiceless stops with the VOT values similar to the ones in their native language. This is consistent with findings that the nature of the task influenced nonnative speakers' attention to their pronunciation in Flege, Takagi and Mann [3], although some research indicates that the foreign accent ratings were affected by word frequency and listening context in Levi, Winter and Pisoni [7], and speech rate in Munro and Derwing [10].

The conflicting result that the VOT shift was only detectable in the sentence-reading task implies that L2 researchers need to reconsider the common speech elicitation technique used by previous studies. The findings in the study suggest that there is a different VOT shift pattern depending on task types in which L2 learners are engaged. Most research on the acquisition of VOT has focused exclusively on sentence reading to elicit data [1, 4]; however, as Tarone [13] argued, in addition to the data elicited by sentence reading, the performance in which an L2 learner's focus is more on accomplishing a communicative task is also necessary to completely describe an L2 learner's interlanguage phonology. Flege's [2] hypotheses in the SLM, which he built by using the data mainly from sentence reading tasks, might need to be revised from this point of view. However, it will require additional studies to use the data elicited under a condition in which more focus is on communicative tasks because spontaneous speech as in a picture description task involves such factors as prosodic domains [6]. In sum, the VOT shift resulting from the short-term but intensive exposure to English is a predictable outcome, but the results should be cautiously interpreted due to the effect of task type differences.

5. REFERENCES

- [1] Flege, J.E. 1991. Age of learning affects the authenticity of voice-onset-time (VOT) in stop consonants produced in a second language. *Journal of the Acoustical Society of America* 89(1), 395-411.
- [2] Flege, J.E. 1995. Second-language speech learning: Theory, findings, and problems. In Strange, W. (ed.), *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*. Timonium MD, York Press, 233-277.
- [3] Flege, J.E., Takagi, N., Mann, V. 1995. Japanese adults can learn to produce English /ɹ/ and /l/ accurately. *Language and Speech* 38(1), 25-55.
- [4] Harada, T. 2007. The production of voice onset time (VOT) by English-speaking children in a Japanese immersion program. *International Review of Applied Linguistics in Language Teaching (IRAL)* 45(4), 353-378.
- [5] Homma, Y. 1981. Durational relationship between Japanese stops and vowels. *Journal of Phonetics* 9, 273-281.
- [6] de Jong K.J. 1994. The supraglottal articulation of prominence in English: Linguistic stress as localized hyperarticulation. *Journal of the Acoustical Society of America* 97(1), 491-504.
- [7] Levi, S.V., Winters, S.J., Pisoni, D.B. 2007. Speaker-independent factors affecting the perception of foreign accent in a second language. *Journal of the Acoustical Society of America* 121, 2327-2338.
- [8] Lisker, L., Abramson, A. 1964. A cross-language study of voicing in initial stops: Acoustical measurements. *Word* 20, 395-411.
- [9] Magloire, J., Green, K.P. 1999. A cross-language comparison of speaking rate effects on the production of voice onset time in English and Spanish. *Phonetica* 56, 158-185.
- [10] Munro, M.J., Derwing, T.M. 1998. The effects of speaking rate on listener evaluations of native and foreign-accented speech. *Language Learning* 48, 159-182.
- [11] Sancier, M.L., Fowler, C.A. 1997. Gestural drift in a bilingual speaker of Brazilian Portuguese and English. *Journal of Phonetics* 25, 421-436.
- [12] Sato, C. 1985. Task variation in interlanguage phonology. In S. Gass, C. Madden (eds.), *Input in second language acquisition*. Rowley, MA: Newbury House, 181-196.
- [13] Tarone, E. 1982. Systemicity and attention in interlanguage. *Language Learning* 32 (1), 69-84.