The effect of VOT on the intelligibility of English voiceless stops produced by native speakers of Japanese

Akiyo Joto, Prefectural University of Hiroshima

Yoshiki Nagase, University of Yamanashi

Seiya Funatsu, Prefectural University of Hiroshima

1 Introduction The phonetic or articulatory difference impedes the perception of certain English sounds produced by native Japanese speakers; English voiceless stops fall under this category of sounds. English voiceless stops are characterized by aspiration in articulation, whereas their Japanese counterparts are known to be unaspirated. As Cruttenden (2001) and Jenkins (2000) point out to learners of English, aspiration is a crucial cue for English voiceless stops. Then, how intelligible are Japanese English (henceforth, JE) voiceless stops, and how acoustically different are intelligible and unintelligible stops?

This paper examined the intelligibility of the English voiceless stops /p/, /t/, and /k/ produced by native Japanese speakers and the relationship between the different levels of intelligibility and VOT. Two experiments were performed to investigate these: one involved an evaluation of the intelligibility of the JE voiceless stops perceived by native English speakers, and the other performed an acoustic analysis of the JE productions to measure their VOTs.

2 Experiment 1 (Evaluation of the intelligibility of the JE stops)

2.1: Materials The stimulus corpus for the evaluation comprised 39 different English words, each of which was read in isolation by twenty native Japanese speakers, including “pet,” “Ted,” and “keck” to be examined for this study. The 39 words consisted of 36 monosyllabic words with 23 different initial consonants and 7 different final consonants and 3 disyllabic ones with 3 different medial consonants, /Z, dZ, N/. Three of the 39 words were open-syllabic. Twenty-two monosyllabic words contained the vowel /E/ and the others, /I/, /ĩ/, /U/, /ũ/, or /□/. Each of the 780 productions (39 words x 20 speakers) recorded on DAT tapes was sampled at 44.1 kHz, randomly ordered, and burned on 4 CDs.

2.2: Speakers The speakers were 20 Japanese learners of English, 9 males and 11 females. They were all university students in Japan, between 18 and 23 years of age. They had received standard English education in junior and senior high school in Japan; none of them were educated in a school where English was the sole medium of instruction.

2.3: Evaluators Five native English speakers, aged between 23 and 50, participated in the evaluation. They comprised two Americans, a male and a female, living in the US and three British people, two males and one female, living in the UK. None of the

evaluators were accustomed to the English spoken by native Japanese speakers. Two of them had received phonetic training at institutions of higher education.

2.4: Procedure Each evaluator was given evaluation sheets and the CDs containing the total of 780 English words pronounced by the native Japanese speakers. Each word had two sheets, (1) and (2); one side of sheet (1) was folded to ensure that the evaluators could not see the word written on it. Each evaluator performed the evaluation task at his/her own pace by following the written instructions.

First, they listened to the words one by one, and then spelled out each sound they heard, on sheet (1). If they were completely unable to understand the word, they could mark a cross (x). They then opened the fold on sheet (1) to view the actual word pronounced by the Japanese speaker written on sheet (2), compared it with the word they had written down on sheet (1), and rated the intelligibility level of the word and the particular consonant spelled in capital letters in the word on sheet (2) on a three-point scale: 3 = Excellent/Good, 2 = Fair, or 1 = Poor. They could write comments, if any, on the JE pronunciation in the specified place on sheet (2). The intelligibility level 1 indicated that a native English speaker recognized a strong foreign accent and found difficulty in understanding the word or a particular consonant. Level 2 meant native speakers recognized a foreign accent which was not strong, and that they anticipated JE productions might be misunderstood or not understandable under some conditions or contexts. Level 3 represented a high level of intelligibility at which native speakers recognized no foreign accent or a very slight accent which had little effect on intelligibility.

2.5: Results The scores assigned by five different evaluators to each sound produced by a native Japanese speaker were averaged. Table 1 shows the number of speakers assigned to each of the three intelligibility levels and the mean scores for each initial stop consonant.

ave.: average

Table 1. The mean scores and the number of Japanese speakers at the three intelligibility levels
for the English stop consonants.

Table 1 indicates that /t/ had the lowest intelligibility (mean score: 2.49) and that the intelligibility of /k/ (mean score: 2.80) was the highest. The intelligibility of /p/ was closer to that of /t/. The words written on sheet (1) revealed that English native speakers heard

most JE productions of the voiceless stops /p/, /t/, and /k/ as the voiced /b/, /d/, and /g/ when they assigned scores of 1 or 2.

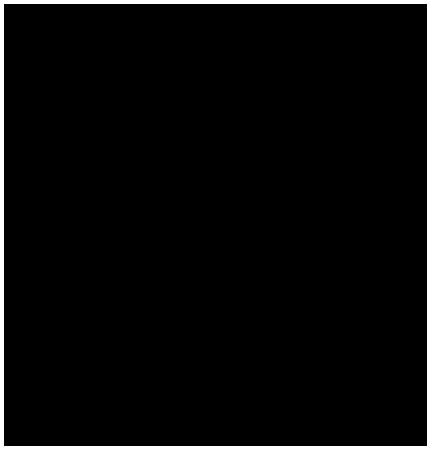
3 Experiment 2 (Acoustic analyses of the JE, English, and Japanese stops)

3.1: Speech materials The JE productions were the same as those used in Experiment 1. The sounds produced by native English speakers (hereafter, AE) were chosen from the "English Learners' Speech Database" (Vols. 9 and 10) built in 2003 for the Grant-in-Aid founded by the Ministry of Education, Culture, Sports, Science and Technology. All the speakers, three males and three females, spoke General American.

The Japanese materials for analysis comprised disyllabic/trimoraic meaningful and nonsense words considered phonemically equivalent to the English materials: they were /petto/, /teddo/, and /kekku/. The Japanese words were produced by three males and three females chosen from among the speakers of the English materials.

3.2: Methods The VOT was measured for each speaker's production of JE, AE, and Japanese voiceless stops. We measured the VOT from consonant release to the onset of glottal vibration on wideband spectrograms.

3.3: ResultsTable 2 shows that the higher the intelligibility scores, the longer were the VOTs for the stop consonants. For all the stops, the intelligibility level was low when the VOT was shorter than 20 ms, and it was high when the VOT was longer than 55 ms.



JE: English stops produced by Japanese speakers
AE: English stops produced by English speakers
JJ: Japanese stops

Table 2. The mean scores and VOTs for the English and Japanese voiceless stop consonants.

The VOTs varied with the stop consonants especially when the intelligibility was "Fair" or "Excellent/Good." When intelligibility was the highest (3), the VOT was more than about 30 ms for /p/, around 50 ms or longer for /t/, and around 55 ms or more for /k/. When the value of VOT ranged from 20 ms to the above-mentioned values for each stop consonant, native English speakers sometimes misheard them as voiced ones. Table 2 also indicated that the mean VOTs for JE and Japanese stops were closer to each

other and considerably shorter than those for the English stops. The difference between the VOTs for Japanese stops and for the most intelligible JE stops varied with the consonants. The difference for /t/ was the largest (20.3 ms): that for /k/was 2.7 ms, and for /p/, 10.5 ms. Therefore, Japanese speakers would find it most difficult to pronounce /t/ with enough aspiration to be intelligible.

4 Discussion As reported in a number of previous studies, Experiments 1 and 2 showed that the JE voiceless stops with lower intelligibility had a VOT shorter than 20 ms, which was less aspirated, and that the native English speakers perceived them as voiced. Then, for the JE stops to be more intelligible, how do learners need to pronounce English stops?

In our study, as well as the information based on the identification task (spelling out the sounds), rating the intelligibility of the JE stops on a 3-point scale was considered to be useful from the pedagogical point of view. In most cases of the evaluation of all the voiceless stops at intelligibility level 2, native English speakers identified each consonant rightly (8 out of 11 cases for /t/, 16 out of 18 for /p/, and 3 out of 8 for /k/), but recognized a foreign accent which might cause misunderstanding in some conditions or contexts. They wrote the comments that these stops were dental /t/, or /p/ and /k/ with weak aspiration. It indicated that there would be a stronger possibility of the JE stops with a VOT shorter than 30 ms for /p/, 50 ms for /t/, or 55 ms for /k/ being misunderstood, and thus, it would be suggested that Japanese learners need to practice pronouncing the English stops to have a VOT longer than the above values. For the JE stops with the highest intelligibility (level 3), the VOTs were not necessarily the same as those for the AE stops. The VOTs for the AE stops were considerably greater than those for the JE stops; nevertheless, the native English speakers perceived them correctly if the VOT was greater than the above values: longer than 30 ms for /p/, 50 ms for /t/, or 55 ms for /k/.

A comparison between the VOTs for the JE and those for the Japanese stops demonstrated that the difficulty for Japanese learners in producing intelligible English voiceless stops hinged on the difference in VOTs between the most intelligible JE and Japanese stops, and thus it revealed negative L1 transfer in the articulation of the English stops. In order to apply our findings in the field of education in Japan, we suggest that on the whole, a VOT of more than 55 ms would be sufficient for JE voiceless stops to be intelligible and that learners need more practice of pronouncing /t/ because /t/ was the most difficult consonant for Japanese. The findings of this study could also provide useful information for English pronunciation practice by using speech sound analysis software, which has been done at some institutions in Japan and Korea.

5 Conclusions The intelligibility of JE voiceless stop consonants and the VOTs for was investigated and the VOTs for JE voiceless stops were measured at the different intelligibility levels. The VOT for the JE stops with lower intelligibility was found to be shorter than that for those with higher intelligibility. When the VOT was less than 20 ms, the possibility of the JE voiceless stops being heard as voiced would be stronger.

6 Acknowledgements This study was supported by a Grant-in-Aid for Scientific Research promoted by the Japan Society for the Promotion of Science (Grant No. 18520387).

7 References

Cruttenden, A. (2001) *Gimson's Pronunciation of English*. 6th ed. London: Arnold.
Jenkins, J. (2000) *The Phonology of English as an International Language.* Oxford: Oxford University Press.
Kent, R. D. & Read, C. (1992) *The Acoustic Analysis of Speech.* San Diego: Singular Publishing Group, Inc.
Ladefoged, P. (2006) *A Course in Phonetics.* 5th ed. Boston: Thomson Wadsworth.