PRODUCTION OF A NON-CONTRASTIVE SOUND IN A SECOND LANGUAGE

Keiichi Tajima[1], Mafuyu Kitahara[2], and Kiyoko Yoneyama[3]

[1]Hosei University, [2]Waseda University, [3]Daito Bunka University
tajima@hosei.ac.jp, kitahara@waseda.jp, yoneyama@ic.daito.ac.jp

ABSTRACT

A speech production experiment was conducted to investigate how well native Japanese speakers can produce a non-contrastive sound in American English, namely, intervocalic alveolar flaps, and how their production varies as a function of language experience. Native Japanese speakers who have prior experience living in the US produced English sentences containing potentially flappable segments. Results showed that Japanese speakers produced alveolar flaps to varying degrees. Acoustic analysis revealed that flaps produced by Japanese speakers show properties that have previously been associated with them, such as short duration and continuous voicing. Results also showed that Japanese speakers produced alveolar flaps within words, e.g. *letter*, and in certain types of phrases, e.g. *get on*. Finally, speakers who frequently produced alveolar flaps tended to be those who arrived in the US early in their life, who stayed in the US for a long time period, and who had high scores on TOEFL iBT.

Keywords: intervocalic alveolar flap, American English, language experience, Japanese learners

1. INTRODUCTION

Studies on second-language (L2) speech learning often focus on production and perception of sounds that signal lexical distinctions in L2. Much less attention has been given to sounds that do not signal lexical contrasts. However, learning to produce and perceive non-contrastive sounds could be important for L2 learners, particularly if they want to achieve native-level performance.

Intervocalic alveolar flapping, the realization of intervocalic /t/ and /d/ as an alveolar flap, is commonly observed in American English (AE), e.g. *letter*, *ladder*, *get on*. Alveolar flaps do not lexically contrast with other sounds in English, but flapping is known to occur highly regularly in certain dialects. For example, a brief analysis of the Buckeye Corpus which contains recordings of AE speakers from Ohio conversing freely with an interviewer [6] indicated that alveolar flaps were produced in 48.6% of words such as *letter* and 38.4% of two-word sequences such as *get on* that contained potentially flappable segments [4].

Production of non-contrastive sounds by L2 learners has not been extensively investigated. Regarding the production of alveolar flaps by Japanese learners of English (JE), analysis of existing speech corpora suggests that JE speakers do not produce alveolar flaps when speaking English. For example, an analysis of the English Read by Japanese (ERJ) corpus which contains recordings of 202 Japanese students from 20 universities across Japan reading lists of English words and sentences [5] showed that these speakers produced virtually no alveolar flaps. This suggests that Japanese university students typically do not produce alveolar flaps. However, it is possible that JE speakers, particularly those with some experience living in the US, might produce alveolar flaps.

The purpose of the present study is to investigate how often JE speakers with some experience living in the US produce alveolar flaps in English. The present study also examined how factors pertaining to the learners' language experience, such as age of arrival in the US, duration of stay, and performance on English proficiency tests, may be related to the production of alveolar flaps.

2. METHODS

2.1. Participants

A group of 40 JE speakers who have lived in the United States for varying lengths were recruited for paid participation in the experiment (26 females, 14 males; age mean = 22.7, age range = 18-37). Most participants were university students in the Tokyo metropolitan area. Age of first arrival in the US varied from age 0 (US-born) to 29 years (mean = 15.6 years, s.d. = 7.3). Duration of stay in the US varied from 0.5 month to 10.5 years (mean = 26.1 months, s.d. = 31.4). Actual or derived TOEFL iBT scores varied from 39 to 116 (mean = 79.2, s.d. = 18.2) [1].

2.2. Materials

Reading materials were sentences containing 65 words and phrases that contained potentially flappable alveolar stops. The flappable segments

always occurred after a vowel. The materials consisted of the following types of items.

1. Simple monomorphemic words (N=12) with a flappable /t/ or /d/, e.g. *letter*, *party*.
2. Suffixed words with a stem-final flappable /t/ (N=10), e.g. *writing*, *greater*, or /d/ (N=8), e.g. *riding*, *grader*.
3. Phrases with a flappable /t/ or /d/, either in intervocalic position ("easy" phrases, N=20), e.g. *get it*, *set up*, or in intervocalic position after reduction/deletion of the following consonant ("hard" phrases, N=10), e.g. *get her*, *hit them*.
4. Non-flapping words (N=5) that normally do not show alveolar flapping in American English, e.g. *thirteen*, *return*.

The 65 target items was each embedded in the carrier sentence *Say ___ now*.

2.3. Procedure

Participants were recorded individually in a quiet room. Each participant read aloud two different randomized lists of the 65 sentences. The materials were presented on a laptop computer screen using a sentence presenter program implemented on Praat [1,3]. Recordings were made using a head-mounted or desktop microphone and a digital recording device at 44.1-kHz sampling frequency and 16-bit resolution, which were later saved as audio files.

2.4. Analysis

A total of 5,200 utterances (40 speakers × 65 items × 2 repetitions) were analyzed. For each utterance, judgment was made as to whether or not the target segment was produced as an alveolar flap using two methods: (1) phonetic transcription by one of the three co-authors based on auditory inspection of each utterance, (2) acoustic analysis of each utterance. Acoustic analysis was based on the following three parameters that are often said to be associated with alveolar flaps [2]. A score from 1 to 3 was assigned for each parameter according to the criteria below, with a larger score indicating a more flap-like segment.

1. Closure duration. If duration of the closure portion of the target segment is between 20ms and 40ms, then assign a score of 3. If duration is less than 10ms or greater than 50ms, then assign a score of 1. For intermediate values, assign a score of 2.
2. Release burst. If release of the target segment is not accompanied by a visible noise burst, then assign a score of 3. If it is accompanied by a visible burst, then assign a score of 1. For intermediate cases, assign a score of 2.
3. Voicing. If vocal fold vibration is observed throughout the closure, then assign a score of 3. If it is absent, then assign a score of 1. For intermediate cases, assign a score of 2.

If the sum of these three scores, which theoretically ranges from 3 to 9, was 7 or greater, then the target segment was judged to be a flap.

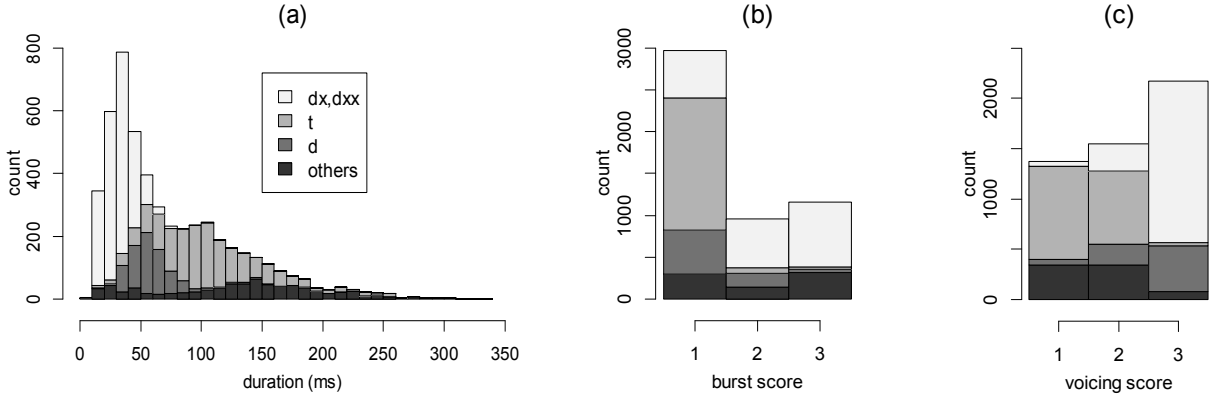
3. RESULTS

3.1. Acoustic analysis

Of the 5,200 utterances recorded, 107 contained pronunciation errors or disfluencies, and were excluded from analysis. Of the remaining 5,093 tokens, 35.5% were judged as alveolar flaps according to the acoustic analysis as described above.

Fig. 1 shows histograms of acoustic measurements and scores for the three criteria above. Fig. 1(a) shows the distribution of closure duration of the target segment. The panel shows that the lightest gray bars, which indicate tokens that were transcribed as a flap (dx) or weak flap (dxx) by the labellers, have relatively short durations, while the darker gray bars, which indicate tokens transcribed as /t/, /d/, or other segments, have longer durations. In fact, 61.3% of

Figure 1: Histogram of (a) flap closure duration, (b) burst score, and (c) voicing score. Different grayscale shadings are used for tokens transcribed as a flap or weak flap, /t/, /d/, and other sounds (see legend in panel (a).



tokens transcribed as a flap or weak flap fell within the 20-40ms range (score=3). An additional 31.8% fell just below (10-20ms) or above (40-50ms) this range (score=2). In contrast, only 2.9%, 12.8%, and 8.8% of tokens transcribed as /t/, /d/, and others, respectively, were within the 20-40ms range.

Fig. 1(b) shows the distribution of scores for the burst criterion. Among the tokens transcribed as a flap or weak flap, 40.1% were produced without a visible release burst (score=3), while 29.6% were produced with a visible burst (score=1). Only a few tokens transcribed as /t/ (2.0%) or /d/ (4.9%) were produced without a visible burst. Tokens transcribed as others included glottal stops, /h/ (e.g. *get him* pronounced with a deleted /t/), and /t/ followed by another consonant (e.g., *get them*). Among these, 42.1% were produced without a visible burst.

Fig. 1(c) shows the distribution of scores for the voicing criterion. Of the tokens produced as a flap or weak flap, 83.5% showed vocal fold vibration throughout the closure (score=3). For tokens transcribed as /d/, this percentage was also high (63.1%). In contrast, this percentage was low for tokens transcribed as /t/ (2.0%) and others (10.8%).

Put together, these results suggest that when alveolar flaps are produced by JE speakers, they show characteristics that have previously been associated with alveolar stops, such as having a closure duration in the 20-40ms range, showing no release burst, and showing vocal fold vibration throughout the closure.

3.2. Flap rate by item type

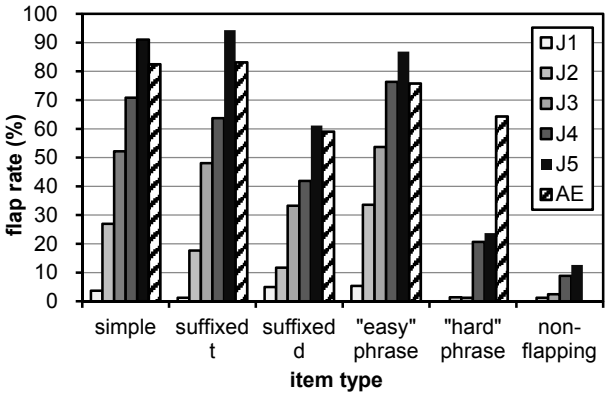
When flap rate was calculated by taking the number of tokens transcribed by the labellers as a flap or weak flap and dividing it by the total number of tokens produced, JE speakers were found to produce alveolar flaps in 37.6% of the tokens. Flap rate was found to vary considerably across speakers, from 1.6% to 78.1% depending on the speaker. Flap rate was also found to vary across item types, from 5.0% in non-flapping words to 49.1% in simple words.

In order to examine how flap rate for different item types varied as a function of speakers' overall flap rate, the JE speakers were divided into five equal subgroups according to their overall flap rate, with eight speakers in each subgroup. The highest 20% of the speakers was designated J5, the next 20% was designated J4, and so on, down to the lowest 20% which was designated J1.

Fig. 2 shows flap rates for the five subgroups of JE speakers as a function of item type. Also shown are flap rates of native AE speakers, obtained by searching for the 65 target items in the Buckeye Corpus and calculating the rate at which these items were flapped in the corpus. The AE speakers' data should only be taken as a rough estimate of flap rates by native AE speakers because the JE and AE data are based on different speech styles (lab speech vs. conversational speech).

Fig. 2 indicates that flap rate showed a staircase-like increase from J1 to J5 for most item types, i.e. simple words, suffixed words with /t/ and /d/, and easy phrases. This staircase-like pattern was to be expected from the way the subgroups were defined, and demonstrates that flap rate in fact varied substantially across speakers. Flap rates for suffixed words with /d/ were lower than those for suffixed words with /t/; this trend was observed also for AE speakers. Mean flap rates for some of the higher subgroups, e.g. J4 and J5, appear to be comparable to that for AE speakers. Flap rates for "easy" phrases were similar to those for words for all five subgroups. In contrast, flap rates for "hard" phrases were much lower, with a mean of 0-2% for groups J1-J3 and 20-25% for groups J4-J5. For non-flapping words, mean flap rate was 0-3% for groups J1-J3 and AE, but was roughly 10% groups J4-J5.

Figure 3: Flap rate of JE speaker groups J1-J5 and AE speakers as a function of item type



A two-way analysis of variance with subgroup (5 levels: J1-J5) as a between-subjects factor, item type (6 levels) as a within-subject factor, and flap rate as the dependent variable showed significant main effects of subgroup [$F(4,35)=115.4$, $p < .001$] and item type [$F(5,175)=91.6$, $p < .001$], and a significant interaction between subgroup and item type [$F(20,175)=8.5$, $p < .001$]. Further analysis of the subgroup-by-item-type interaction indicated the following pattern of significant differences at the $p < .05$ level for each item type: simple words: J1 < J2 < J3 < J4 < J5; suffixed words with /t/: J1 < J2 < J3 = J4 < J5; suffixed words with /d/: J1 = J2 < J3 = J4 < J5; "easy" phrases: J1 < J2 < J3 < J4 = J5; "hard" phrases: J1 = J2 = J3 < J4 = J5. Non-flapping words did not show significant differences among J1-J5.

In short, flap rate varied substantially across JE speakers. Flap rates for "easy" phrases were similar

to those for single words for all subgroups, but flap rates for "hard" phrases were lower than those for words and easy phrases for all subgroups. Flap rates for hard phrases were very low (0-2%) for groups J1-J3, but were higher (20-25%) for groups J4-J5.

3.3. Correlations with speaker-related factors

In order to examine how strongly JE speakers' flap rate was related to their language experience, Table 1 shows pairwise correlations among five variables: age of first arrival in the US (AOA), total duration of stay in the US (DOS), TOEFL iBT score, flap rate based on the labellers' transcriptions, and flap rate based on the acoustic criteria.

Table 1: Correlations among age of first arrival in the US (AOA), duration of stay in the US (DOS), TOEFL iBT score, flap rate based on the transcribers' transcription (transcr.), and flap rate based on acoustic criteria (acoustic). Asterisks indicate statistically significant correlations.

	AOA	DOS	TOEFL	transcr.	acoustic
AOA	1.000				
DOS	-0.851***	1.000			
TOEFL	-0.548***	0.564***	1.000		
transcr.	-0.394*	0.432**	0.512***	1.000	
acoustic	-0.339*	0.355*	0.395*	0.974***	1.000

*$p < .05$, **$p < .01$, ***$p < .001$

Results show that both transcription-based and acoustic-based flap rates were significantly negatively correlated with AOA, indicating that speakers who arrived in the US at a younger age (including US-born speakers) tended to have higher flap rates. Both flap rates were also significantly positively correlated with DOS and TOEFL iBT score, indicating that speakers who lived in the US for a longer time period and speakers who scored higher on TOEFL iBT tended to have higher flap rates. All of these correlations are in the expected direction. Among the three speaker-related factors, TOEFL iBT score showed the highest correlation with both flap rates, suggesting that standardized test scores that reflect learners' overall English proficiency was a better predictor of flap rate than simple measures of English language experience such as age of arrival and duration of stay. However, it should be pointed out that these three speaker-related factors were all highly correlated with one another, as shown in Table 1. Thus, it would be misleading to attempt to identify a single factor that explains the variability in flap rate.

4. DISCUSSION AND CONCLUSION

The present study investigated how often JE speakers with some experience living in the US produced alveolar flaps in English, and how factors such as age of arrival in the US, duration of stay, and performance on English proficiency tests, are related to the production of alveolar flaps.

Results of the production experiment indicated, first, that JE speakers with some experience living in the US did produce alveolar flaps, contrary to other data that suggest that JE speakers rarely produce alveolar flaps [5]. Acoustic analysis suggests that alveolar flaps produced by JE speakers show acoustic properties that have previously been associated with alveolar flaps, such as closure duration of 20-40 ms, vocal fold vibration throughout the closure, and absence of release burst. Release burst was absent in many productions, but there were also many other tokens with a clearly visible release burst (Fig. 1(b)). Further research is needed to determine whether this characteristic is observed only in JE speakers' productions or is seen also in AE speakers' productions, and whether this characteristic might also be seen in productions of liquid consonants in Japanese which are phonetically similar to AE alveolar stops.

Results also showed that JE speakers produced alveolar flaps in "easy" phrases such as *get on* as often as they did within single words such as *letter*. This suggests that producing alveolar flaps across word boundaries is not necessarily harder than producing them within words. However, JE speakers produced alveolar flaps less frequently in "hard" phrases such as *get her*, presumably because it entails an additional phonological process of consonant reduction or deletion. Some JE speakers, particularly those with high overall flap rates, even over-generalized and produced alveolar flaps in words that AE speakers typically do not produce flaps in, e.g. *thirteen*.

Finally, correlation analysis among flap rates and speaker-related factors suggested that the age of first arrival in the US, duration of stay, and TOEFL iBT score were all moderately correlated with flap rate. That is, flap rate tended to be higher for JE speakers who arrived in the US early in their life (including US-born speakers), who stayed in the US for a long time period, and who had high scores on TOEFL iBT.

Taken together, the present study suggests that living or studying in an L2-speaking community may have a significant impact on L2 learners' development of pronunciation skills in L2, including production of non-contrasting L2 sounds.

5. ACKNOWLEDGMENTS

This research was supported in part by JSPS Grant-in-Aid for Scientific Research, grant numbers 25370443, 26370508, and 15K02492.

REFERENCES

[1] Boersma, P., Weenink, D. 2014. Praat: doing phonetics by computer, University of Amsterdam. http://www.fon.hum.uva.nl/ praat/

[2] de Jong, K. 1998. Stress-related variation in the articulation of coda alveolar stops: Flapping revisited. *J. Phonetics* 26, 283-310.

[3] Dellwo, V. 2014. Praat plugins, UZH Phonetics Laboratory. University of Zurich. http://www.pholab.uzh.ch/static/volker/software.html

[4] Kitahara, M., Tajima, K., Yoneyama, K. 2014. Production of a non-phonemic variant in a second language: Acoustic analysis of Japanese speakers' production of American English flap. *J. Acoust. Soc. Am.* 136, 2146.

[5] Minematsu, N., Tomiyama, Y., Yoshimoto, K., Shimizu, K., Nakagawa, S., Dantsuji, M., Makino, S. 2002. English speech database read by Japanese learners for CALL system development. *Proc. International Conference on Language Resources and Evaluation*, Canary Islands, Spain, 896-903.

[6] Pitt, M., Dilley, L., Johnson, K., Kiesling, S., Raymond, W., Hume, E., Fosler-Lussier, E. 2007. Buckeye Corpus of Conversational Speech, (2nd release), Dept. of Psychology, Ohio State University (Distributor). http://www. buckeyecorpus.osu.edu

[1] TOEFL iBT scores were not available for all 40 of the JE speakers in the experiment. In such cases, conversion formulas were used to convert scores of other standardized tests to TOEFL iBT scores. For TOEIC scores, the following formula was used to first convert them to TOEFL PBT scores: TOEFL PBT = TOEIC × 0.348 + 296. Then, TOEFL PBT scores were converted to TOEFL iBT scores based on the table in the following site: http://www.conversation.jp/faq/faqenglish/TOEIC-TOEFL.html. For EIKEN scores, they were converted to TOEFL iBT scores based on the table in the following site: http://ieltsnavi.com/score_conversion.html.