Concentric: Studies in Linguistics
37.2 (July 2011):155-178

Acoustic Analysis of Voice-Onset Time in Taiwan Mandarin and Japanese*

Naomi Ogasawara
National Taiwan Normal University

The purpose of this study is to present a thorough acoustic investigation of VOT for stops in connected speech in Taiwan Mandarin (TM) and Japanese. Aspirated and unaspirated stops in TM and voiced and voiceless stops in Japanese not only in isolated syllables but also in word-initial/-medial (Japanese) or phrase-initial/-medial (TM) positions were involved. The study found that TM aspirated stops definitely belong to the long-lag category, and unaspirated stops belong to the short-lag category. Conversely, VOT for Japanese voiceless stops in the initial position falls somewhere between short-lag and long-lag; while, VOT in the medial position becomes extremely short, almost as short as voiced stops with positive VOT. Moreover, the occurrence of prevoicing for voiced stops is random. These results show that Japanese stops are acoustically quite different from TM stops in terms of VOT, which further implies that Japanese stops might be difficult for TM speakers to distinguish correctly.

Keywords: VOT, Taiwan Mandarin, Japanese

1. Introduction

Mastering phonological contrasts in a second language (L2) is not an easy task for learners whose native language (L1) does not have the same phonological contrasts. A well-known example of that is discriminating English /l/ and /r/ by Japanese listeners. The task is hard for Japanese listeners who have had little exposure to spoken English because there is no liquid contrast, and the flap /ɾ/ substitutes for the alveolar liquids in Japanese (Aoyama et al. 2004, Eckman and Iverson 1993, Flege et al. 1995, 1996). However, such difficulty cannot be attributed only to the lack of L2 phonological contrasts in the L1. The difficulty also comes from phonetic differences between L1 and L2 sounds. The Perceptual Assimilation Model (PAM) incorporates phonetic-articulatory similarity/dissimilarity of L2 sounds to L1 sounds in the discrimination process (Best et al. 2003, Hallé et al. 1999). For example, Hallé, Best, and Levitt (1999) investigated the discrimination of American English (AE) approximant contrasts (/w, j, l, r/) by French listeners. The discrimination performance of /w/-/j/ by French listeners was as good as that of AE listeners; while, /r/-/l/ discrimination was more difficult than the /w/-/j/ contrast, and /w/-/r/ was the most difficult pair for

* This research was supported by a Grant for New Researchers, National Science Council, Taiwan (NSC 98-2410-H-003-035). I would like to thank Dr. S.T. Bischoff and two anonymous reviewers for their helpful comments on the earlier version of the paper. I also would like to thank Professor Chun-Yin Chen and Ms. Ya-Hui Lin for their advice and help in recruiting subjects, and Ms. Shang-Chen Chiu, Ms. Hsiang-Ying Lo, and Ms. De-Rong Liao for their assistance in conducting the experiments.

French listeners. These results confirmed that phonetic-articulatory differences between L1 and L2 are substantive for discrimination of non-native phonological contrasts. French has the /r/-/l/ contrast, and the French /l/ is phonetically similar to the AE /l/, but the French /r/ is not similar to the AE /r/. It is often a fricative-like uvular approximant or trill rather than an alveolar; in addition, the lip-rounding involved in the AE /r/ makes /r/ more similar to /w/ in French. Due to these phonetic differences, the AE /r/ is a poor exemplar of the French /r/, and it is perceptually confused with the French /w/, which makes /w/-/r/ discrimination extremely difficult for French listeners.

Learners' difficulties in mastering L2 phonological contrasts can also be attributed to their inability to employ the acoustic cues which native listeners use to identify phonemes. Flege (1989) tested the discrimination performance of the word-final English /t/-/d/ by native English adults and children and by adult L2 listeners whose native language was Mandarin, Taiwanese, Shanghainese, or Hakka. Native English listeners, both adults and children, successfully identified the two phonemes (always more than 95% of the time) including stimuli without the prominent acoustic cues of prevoicing and stop burst. On the contrary, the performance of the Chinese listeners was significantly worse when they listened to the modified stimuli without the prominent cues (58% to 88% of the time). Chinese listeners relied heavily on the prominent cues for perception of the stops, and did not use secondary acoustic cues, such as the duration of the adjacent vowel or formant transition of the vowel offset, which also can characterize /t/ and /d/.

Similarly, Taiwan Mandarin (TM) speakers seem to have difficulty in discrimination of voicing Japanese stops in daily conversation or in a classroom setting.[1] Some examples are /udɑ/ for /utɑ/ 'song', /nego/ for /neko/ 'cat', /wɑdɑʃi/ for /wɑtɑʃi/ 'I', and /hidoɾi/ for /hitoɾi/ 'alone'. Such difficulty is seen both in perception and production. Voiceless stops produced by native Japanese speakers are often misheard as voiced stops, and the words are repeated with the incorrect stop. There are notable features observed in these kinds of errors: 1) voiceless stops are frequently substituted by voiced stops, but the opposite case is less often observed; and 2) the errors occur more frequently in the word-medial position than in the word-initial position. It seems that Japanese voiceless stops are difficult to hear, especially intervocalically, for TM speakers. Previous literature has shown that phonological contrast between voiceless and voiced stops is closely related to VOT, which is a prominent phonetic feature of stops (Cho and Ladefoged 1999, Lisker and Abramson

[1] For example, the author has been teaching Japanese to students in Taiwan and half of the students had this problem on a pronunciation test.

1964, Keating 1984, Kent and Read 1992, among others). This leads to the hypothesis that VOT for distinction of stop contrast is not identical between Mandarin and Japanese, and that even for the same stop, VOT values are inconsistent based on the position in a word or a phrase where the stop occurs.

The most apparent articulatory characteristic for stop consonants is the occlusion of air made by two articulators coming into contact with each other and a burst of air when the articulators part. The interval of air occlusion is called stop closure, and the burst of air is called release (Kent and Read 1992). Phonologically, stops are categorized into two types: voiced and voiceless. Phonetically, voiced stops have slight vocal fold vibrations during closure, which is called "prevoicing" or "voicing lead" (Lisker and Abramson 1964, Kent and Read 1992). Air pressure builds in supraglottis, but with the effect of enlargement of the space by lowering the larynx, advancing tongue root, etc., it can be delayed to some extent, which makes transglottal air pressure enough to maintain vocal fold vibrations (van Alphen 2007). In waveform presentation, this glottal activity is observed as periodical waves with low amplitude during closure. Contrastively, the glottal activity is absent during closure for voiceless stops. However, this criterion cannot always apply when categorizing voiced and voiceless stops because in some languages, prevoicing is absent for voiced stops. For instance, prevoicing is not seen for English stops in the initial position. Based on the criterion, those stops are actually "voiceless". In addition to the presence or absence of prevoicing, there is another criterion for categorizing stops, whether a stop is aspirated or unaspirated. Aspiration is a breathy noise of the glottal fricative /h/ occurring after the release burst until the onset of voicing, which is called "voicing lag". Aspirated stops have a longer lag, and unaspirated stops have either no lag or a very short lag. However, this criterion is applied to voiceless stops only.

Lisker and Abramson (1964) suggested VOT as a unified measure to precisely categorize stops instead of having the two discrete criteria (presence or absence of prevoicing and aspiration) separately. VOT is defined as the interval between the release of the stop closure and voicing onset. In the case of prevoiced stops, VOT is measured between voicing during closure and the stop release (therefore a negative value for VOT). Lisker and Abramson (1964) measured VOT for stops in eleven languages. Table 1 shows a summary of the VOT results from several languages with stops of two categories (voiced and voiceless). Some English voiced stops had prevoicing; thus, both positive and negative numbers are given.[2] English data in

[2] One speaker produced stops with prevoicing in 95% of the data.

isolation show that VOT falls in the range between +1 and +80 ms and between -102 and -88ms (prevoicing). The data from other languages show similar ranges, -100 to +10 ms VOT for stops in isolation. Compared with the stops in isolation, it is notable that VOT becomes shorter in connected speech especially in the non-initial position.

Table 1. Average VOT (ms) in two-category languages (Lisker and Abramson 1964:391-411)

		Labial		Coronal		Dorsal	
		b	p	d	t	g	k
Dutch	In isolation	-85	10	-80	15	N/A	25
	Initial (in sentence)	-41	11	-51	16	N/A	34
	Non-initial (in sentence)	N/A	9	-40	20	N/A	33
PR Spanish[3]	In isolation	-138	4	-110	9	-108	29
	Initial	-110	4	-109	7	-92	25
	Non-initial	-90	4	N/A	8	N/A	20
Hungarian	In isolation	-90	2	-87	16	-58	29
	Initial	-55	0	-70	20	-61	28
	Non-initial	N/A	4	N/A	24	N/A	34
Tamil	In isolation	-74	12	-78	8	-62	24
	Initial	-61	12	-64	10	-56	27
	Non-initial	N/A	6	N/A	6	N/A	10
English	In isolation	1 -101	58	5 -102	70	21 -88	80
	Initial	7 -65	28	9 -56	39	17 -45	43
	Non-initial	4 -63	34	7	37	16	49

Generally, stops are categorized phonetically in three ways based on the values of VOT: voiceless unaspirated stops (short-lag — 0-25 ms), aspirated stops (long-lag — 60-100 ms), and prevoiced stops with negative VOT (Lisker and Abramson 1964, Keating 1984, Riney et al. 2007). Figures 1A, B, and C display the waveforms of stops in the three categories. Cho and Ladefoged (1999) examined VOT values in eighteen languages, and they suggest four categories for voiceless stops: unaspirated (0 ms up to around 30 ms), slightly aspirated (around 50 ms), aspirated (around 90 ms), and highly aspirated (above 90 ms).

[3] PR stands for Puerto Rican.

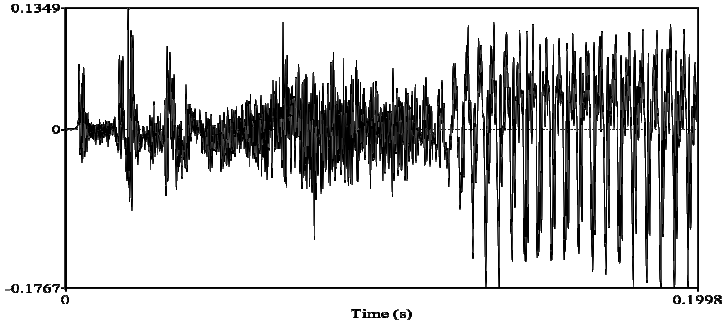

Figure 1A. Waveform of TM aspirated [k^h] (VOT = 116.57 ms)

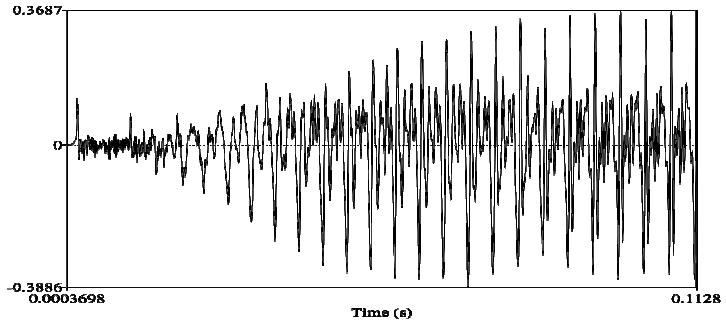

Figure 1B. Waveform of TM unaspirated [k] (VOT = 17.17 ms)

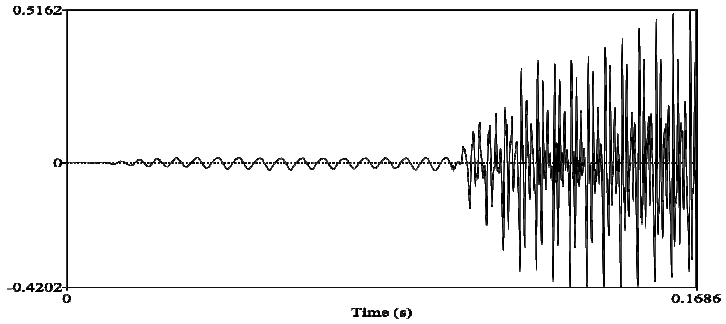

Figure 1C. Waveform of Japanese voiced [b] (VOT = -91.39 ms)

In Mandarin, there are six stops with three different places: bilabial aspirated /p^h/, unaspirated /p/, dental aspirated /t^h/, unaspirated /t/, velar aspirated /k^h/, and unaspirated /k/ (Duanmu 2000, Lin 2007), and they occur only in the syllable-initial position. VOT for Taiwan Mandarin stops has been investigated in previous studies. Data from two studies are summarized in Table 2. VOT for /p^h/, /t^h/, and /k^h/ ranges 75 to 92 ms, which indicates TM aspirated stops belong to the long-lag category, and VOT for unaspirated /p/, /t/, and /k/ ranges 14 to 27 ms, which belongs to the short-lag category. Chen, Chao, and Peng (2007) point out that TM aspirated stops have higher VOT values than English voiceless stops, but the difference of VOT values between labials and coronals is smaller in Mandarin than in English.

Table 2. Average VOT (ms) for TM stops (Chao and Chen 2008: 223; Chen, Chao, and Peng 2007:8, 10)

	Bilabial		Dental		Velar	
	p^h	p	t^h	t	k^h	k
Chao et al. 2008	82	14	81	16	92	27
Chen et al. 2007	77.8	13.9	75.5	15.3	85.7	27.4

In Japanese, there are also six stops with three places of articulation (bilabial /p, b/, (lamino-)alveolar /t, d/, velar /k, g/), and stops occur only in the syllable-initial[4] position. Unlike Mandarin stops, which are clearly divided in the short-lag and long-lag categories, VOT for Japanese stops show some complication. Vance (2008) has described Japanese voiceless stops as having aspirated and unaspirated allophones, but the distribution of aspirated stops is limited to phonetically salient positions such as word-initial or in an accented syllable. However, those word-initial stops are not prototypical aspirated stops having a long-lag as in other languages; instead, their VOT values fall somewhere between short-lag and long-lag. Homma (1980) reported a mean VOT of 25 ms for word-initial /t/ in *tada*. In other studies, it was 24 ms for /p/, 32 ms for /t/, and 45 ms for /k/ (Homma 1981), or 29 ms for /t/ in /ta/ (Shimizu, 1996). Harada (2003) reported that VOT for voiceless stops produced by monolingual Japanese adults was 24 ms for /p/, 26 ms for /t/, and 42 ms for /k/. Riney, Takagi, Ota, and Uchida (2007) conducted a thorough investigation of VOT for monomoraic word-initial voiceless stops with four vowels (/i, e, o, ɑ/) in connected speech. They obtained a mean VOT 30.0 ms for /p/, 28.5 ms for /t/, and 56.7 ms for /k/, which do not fall either in the short-lag or long-lag categories.

While the previous literature has focused on stops in the initial position, little attention has been paid to stops in the medial position. The present study aims to provide an acoustic investigation of VOT for stops in both positions in order to make a comparison between TM and Japanese.

[4] Obstruent geminates occur in the syllable-final position, but they are not considered in this paper.

2. Experiment 1: VOT for Mandarin stops

2.1 Method

2.1.1 Materials

Six stops (/p^h, p, t^h, t, k^h, k/) in Taiwan Mandarin were used for the experiment. These stops were tested in three environments: 1) in a monosyllable; 2) in a disyllabic real phrase; and 3) in a disyllabic nonsense phrase.[5] In the monosyllable environment, stops were placed in the initial position followed by the vowel /a/ with tone 1. The low vowel was chosen because syllables of stop + /a/ are the most prevalent in the lexicon. Hence, six syllables were selected (/p^ha/, /pa/, /t^ha/, /ta/, /k^ha/, /ka/). All items were written with the Zhu-yin phonographic system. In the real phrase and nonsense phrase environments, stops were placed in the initial position of either the first syllable (phrase-initial) or the second syllable (phrase-medial). The vowel /a/ with tone 1 followed the target stop. The adjacent syllable had either tone 1, 2, 3, or 4, which created 48 combinations (6 stops x 2 positions (initial or medial) x 4 adjacent tones). Disyllabic real phrases consisting of the first syllable with tone 2, 3, or 4 plus the second syllable /ga/ with tone 1 were not found in the Mandarin lexicon. A phrase for /k^ha/ plus the second syllable with tone 1 could not be found, either. As a result, forty-four items were used in the real phrase environment. Each item was placed in a frame sentence, "這是 _____ 吧。" 'This is probably _____.' Whole sentences were written with Chinese characters. See Appendix A for the lists of the experimental items. In each environment, items were randomized and two lists with a different item order were created.

2.1.2 Participants

Ten university students (five males and five females) were recruited in Taipei. Their first language is Taiwan Mandarin, and they cannot speak Taiwan Southern Min fluently. None of them reported speech or hearing disorders.

[5] The nonsense phrase condition was included in order to see if there was any difference between the real phrase and the nonsense phrase. If so, this would be a confronting factor.

2.1.3 Procedure

Each participant was recorded individually in a sound-attenuated booth in the phonetics lab at National Taiwan Normal University. Participants were seated in front of a microphone connected to a recording device. Each participant was randomly assigned one of the two item lists for each environment. Participants were asked to look through the lists before recording to make sure there were no ambiguous items and that they were able to pronounce all the items. They were asked to read each list out loud at a normal speed. They put a pause with a natural length between items. After completing the list one time, they read through the list out aloud for a second round. Their speech was recorded at a 44.1 KHz sampling rate. Recordings were made for the monosyllable list, the real phrase list, and the nonsense phrase list for each participant. Oral instruction was given to each participant by a female native speaker of Taiwan Mandarin. Upon the completion of the experiment, participants were asked to fill out a questionnaire about their own and their parents' linguistic background.

2.1.4 VOT measurement

A total of 1960 tokens ((6 monosyllables + 44 real phrases + 48 nonsense phrases) x 10 participants x 2 repetition) were recorded. VOT for the target stop in each token was measured from release to the onset of voicing using PRAAT software (Boersma and Weenink 2011). Release was recognized as a wave with higher intensity in the waveform and a vertical striation in the spectrogram. The onset of voicing was recognized as the beginning of periodic waves at the zero crossing with much higher intensity and the beginning of the first formant in the spectrogram.

2.2 Results and discussion

Among 1960 tokens, 35 tokens were discarded from the analyses due to mispronunciation of the target stops. Since participants produced each item twice, mean VOT was calculated for each item and for each participant. Table 3 shows mean VOT in each condition.

Table 3. Average VOT (ms) for Mandarin stops

	Bilabial		Dental		Velar	
	p^h	p	t^h	t	k^h	k
Monosyllable	69.8	9.19	69.85	9.63	86.42	18.95
	(20.16)	(2.54)	(16.58)	(2.57)	(16.07)	(3.64)
Real phrase initial	62.87	9.25	62.98	11.40	76.80	20.82
	(19.52)	(3.93)	(24.04)	(2.50)	(15.91)	(3.93)
Real phrase medial	52.34	9.87	49.59	11.68	68.68	25.84
	(22.89)	(3.90)	(26.27)	(3.46)	(19.47)	(6.00)
Nonsense phrase initial	64.60	8.94	58.74	11.09	77.78	20.87
	(21.92)	(2.70)	(27.90)	(3.06)	(22.56)	(4.57)
Nonsense phrase medial	55.24	9.32	48.00	10.99	65.83	20.75
	(27.13)	(3.53)	(25.00)	(3.01)	(21.31)	(4.14)

Note* Numbers in parentheses show the standard deviation (SD).

The descriptive statistics reveal some notable points. First, overall, mean VOT in the present study is slightly shorter than those in the previous studies in Table 2. A possible reason for this is alternation in methodology. In the present study, except for the monosyllable condition, stops were placed in connected speech in which segments were temporally compressed and less fully articulated (Lisker and Abramson 1964). Another reason might be the following vowel. In general, VOT is longer when stops are followed by high vowels /i, u/ than when followed by the low vowel /a/ (Chao and Chen 2008, Port and Rotunno 1979). Only the low back vowel /a/ was chosen for the current experiment, which could possibly lead to an overall shorter VOT. Second, by comparing standard deviation between aspirated and unaspirated stops, it is obvious that the VOT for aspirated stops is more widely distributed than those for unaspirated stops along the scale. VOT ranges 8.55 ~ 110.16 ms for /p^h/, 15.55 ~ 113.63 ms for /t^h/, and 28.02 ~ 123.72 ms for /k^h/. Conversely, much smaller ranges are observed for unaspirated stops (1.25 ~ 19.51 ms for [p], 5.66 ~ 22.06 ms for /t/, 12.53 ~ 38.23 ms for /k/). These results are compatible with Chao and Chen's study (2008) for TM stops. According to Lisker and Abramson (1964), widely distributed VOT is also observed in other languages which have an aspirated-unaspirated distinction (Cantonese, Eastern Armenian, Thai). This may be a cross-linguistic acoustic feature for aspirated stops.

Next, VOT values were tested statistically. Beforehand, it was ensured that no correlation (r_{pb} = -.016, p > .5) was found between VOT values and phrase type (real phrase vs. nonsense phrase). In addition, multiple regression with VOT variables and adjacent tone as a predictor did not reach significance (F = .329, p > .5). Hence, the two factors, phrase type and adjacent tone, were excluded from the following analyses. Mean VOT values were obtained for each stop and for each subject by collapsing

conditions (monosyllabic, real phrase, and nonsense phrase) and position of a stop in a phrase (initial and medial): /p^h/ (*M* = 59.1, *SE* = 20.04); /t^h/ (*M* = 55.33, *SE* = 22.73); /k^h/ (*M* = 72.83, *SE* = 17.2); /p/ (*M* = 9.36, *SE* = 3.02); /t/ (*M* = 11.2, *SE* = 2.42); /k/ (*M* = 21.25, *SE* = 3.28).

Two-way analysis of variance (ANOVA) was conducted with consonant (bilabial, alveolar, velar) and position (phrase-initial and -medial) as within-subject factors and counterbalanced group as a between-subject factor. For aspirated stops, the main effect of the consonant was significant ($F(2, 16) = 23.89$, $p < .001$), and the main effect of the position in the phrase was also significant ($F(1, 8) = 34.96$, $p < .001$), but the interaction between the two factors was insignificant ($F(2, 16) = .31$, $p > .5$). Planned comparison revealed that among aspirated stops, the VOT difference between /p^h/ and /t^h/ was not significant (initial position: $F(1, 8) = .58$, $p > .1$; medial position: $F(1, 8) = 5.04$, $p > .05$); while it reached significance between /p^h/ and /k^h/ (initial position: $F(1, 8) = 17.29$, $p = .003$; medial position: $F(1, 8) = 35.71$, $p < .001$) and between /t^h/ and /k^h/ (initial position: $F(1, 8) = 18.42$, $p = .003$; medial position: $F(1, 8) = 30.56$, $p = .001$). For unaspirated stops, the main effect of the consonant was significant ($F(2, 16) = 104.36$, $p < .001$), but neither the main effect of position in the phrase ($F(1, 8) = 3.37$, $p > .1$) nor the interaction between the two factors ($F(2, 16) = .79$, $p > .4$) was significant. All paired comparisons reached significance: between /p/ and /t/ (initial position: $F(1, 8) = 81.84$, $p < .001$; medial position: $F(1, 8) = 6.81$, $p < .04$); between /p/ and /k/ (initial position: $F(1, 8) = 8.83$, $p < .02$; medial position: $F(1, 8) = 95.62$, $p < .001$); and between /t/ and /k/ (initial position: $F(1, 8) = 188.24$, $p < .001$; medial position: $F(1, 8) = 107.48$, $p < .001$). The statistical results support that VOT for /k^h/ is significantly longer than for /p^h/ or /t^h/, but VOT for /p^h/ and /t^h/ is almost the same. In contrast, VOT for unaspirated stops follows the cross-linguistic pattern: VOT length positively correlates with the backness of the closure for stops (Cho and Ladefoged 1999, Fischer-Jørgensen 1954, Peterson and Lehiste 1960).

Finally, the effect of position in a phrase was tested for each stop. In the real phrase and nonsense phrase environments, stops were placed either in the initial of the first syllable (phrase-initial) or of the second syllable (phrase-medial). Mean VOT values were obtained for each stop, for each position, and for each subject by collapsing the environment factor: phrase-initial /p^h/(*M* = 64.41, *SE* = 17.97); phrase-medial /p^h/(*M* = 53.78, *SE* = 21.49); phrase-initial /t^h/ (*M* = 61.86, *SE* = 22.56); phrase-medial /t^h/ (*M* = 48.80, *SE* = 22.07); phrase-initial /k^h/ (*M* = 78.4, *SE* = 15.86); phrase-medial /k^h/ (*M* = 67.25, *SE* = 17.44); phrase-initial /p/(*M* = 9.1, *SE* = 2.79); phrase-medial /p/(*M* = 9.61, *SE* = 3.36); phrase-initial /t/ (*M* = 11.06, *SE* = 2.07); phrase-medial /t/ (*M* = 11.34, *SE* = 2.84); phrase-initial /k/ (*M* = 20.65, *SE* = 3.2);

phrase-medial /k/ (M = 21.85, SE = 3.41). Planned comparison revealed that VOT for aspirated stops was significantly longer in the initial position than in the medial position: /p^h/ ($F(1, 8) = 13.33$, $p < .01$); /t^h/ ($F(1, 8) = 27.38$, $p = .001$); /k^h/ ($F(1, 8) = 16.52$, $p < .005$). On the contrary, VOT for unaspirated stops was consistent regardless of position: /p/ ($F(1, 8) = 4.39$, $p > .05$); /t/ ($F(1, 8) = .38$, $p > .5$); /k/ ($F(1, 8) = 2.01$, $p > .1$).

In sum, Experiment 1 revealed some important acoustic characteristics of VOT for TM stops: 1) VOT for aspirated stops is widely distributed on the continuum; 2) among aspirated stops, only the velar stop is longer than the other stops; while among unaspirated stops, VOT increases as the place for stop closure moves further back; 3) regarding the effect of position, VOT for initial aspirated stops is longer than medial stops; while, VOT values for unaspirated stops are consistent and not affected by the position.

3. Experiment 2: VOT for Japanese stops

3.1 Method

3.1.1 Materials

Six Japanese stops (/p, b, t, d, k, g/) were used. As in Experiment 1, stops were tested in three different environments: 1) in a monosyllable, 2) in a disyllabic real word, and 3) in a disyllabic nonsense word. In the monosyllable environment, stops were placed in the initial position followed by a vowel /ɑ/, /e/, or /o/, which created eighteen syllables (/pɑ, bɑ, tɑ, dɑ, kɑ, gɑ, pe, be, te, de, ke, ge, po, bo, to, do, ko, go/). The high vowels /i/ and /ɯ/ were excluded because the alveolar stops change to affricates or a fricative /z/ before the high vowels, such as /tʃi/, /tsɯ/, /dʒi/, and /zɯ/ (Vance 2008). All items were written with the *hiragana* phonographic system. In the real word environment, only alveolar and velar stops were used because the Japanese lexicon does not contain Japanese-origin words starting with the voiceless bilabial stop. Also, the mid front vowel /e/ was excluded due to the small number of entries in the lexicon. Hence, eight syllables (/tɑ, dɑ, kɑ, gɑ, to, do, ko, go/) were selected. Sixteen disyllabic real words, in which stops were placed in the initial of the first syllable (word-initial) or the second syllable (word-medial), were chosen. In order to avoid the effect of pitch accent, all words had the identical High-Low (HL) pitch accent in the Tokyo dialect of Japanese. Each item was placed in a frame sentence, “Kore wa _____ desu.” ‘This is _____.’ The frame sentence was written in *hiragana*,

and the experimental words were written in both *kanji* (Chinese letters used in Japanese) and *hiragana*. In the nonsense word environment, all eighteen syllables were used. Thirty-six disyllabic nonsense words were created. Stops appeared in the word-initial or -medial position. As in the real word environment, experimental items were assigned only the HL pitch accent. Each item was placed in the same frame sentence written in *hiragana* and only the experimental words were written in *katakana*. See Appendix B for the lists of experimental items. For each environment, items were randomized and two lists with a different item order were created.

3.1.2 Participants

Ten native speakers of the Tokyo dialect of Japanese (five males and five females) were recruited in Taipei. The average length of living in Taiwan was 2.9 months at the time of recording. Their proficiency of Mandarin was at the beginning or intermediate level. None of them reported speech or hearing disorders.

3.1.3 Procedure

A similar procedure as in Experiment 1 was employed in this experiment except the oral instruction was given to participants by a female native speaker of the Tokyo dialect of Japanese.

3.1.4 VOT measurement

A total of 1400 tokens ((18 monosyllables + 16 real words + 36 nonsense words) x 10 participants x 2 repetition) were recorded. VOT for the target stop for each token was measured from release to the onset of voicing using PRAAT software (Boersma and Weenink 2011) in the same fashion as in Experiment 1. For prevoicing, VOT was measured from the stop burst to the offset of the preceding vowel at zero crossing with the end of vertical striation in the spectrogram and waves with low intensity (see Figure 1C).

3.2 Results and discussion

Among the 1400 tokens, 80 tokens were discarded from the analyses due to mispronunciation of the target syllables or an unclear stop burst (mostly in voiced stops with prevoicing). Participants produced the items twice, and mean VOT was

calculated for each item and for each participant for the voiceless stops. On the other hand, mean negative VOT and positive VOT were calculated separately for voiced stops. For some subjects and some items, both negative and positive VOT appeared in two readings; therefore, mean VOT did not reflect the real data and they were excluded from the analyses.

3.2.1 VOT for Japanese voiced stops

Table 4 shows mean VOT for voiced stops in five conditions. From the descriptive analysis of the data, the following points about Japanese voiced stops become clear. First, VOT is scattered on the continuum from negative to positive. In this experiment, 51.7 % of voiced stop tokens had negative VOT (prevoicing), and 48.3 % (voicing lag) had positive VOT. Mean VOT ranged from 4.94 ~ 22.62 ms / -122.02 ~ -28.92 ms for /b/, 4.97 ~ 23.69 ms / -131.76 ~ -22.74 ms for /d/, and 1.30 ~ 34.38 ms / -89.30 ~ -19.55 ms for /g/. Lisker and Abramson (1964) found both positive and negative VOT for English voiced stops. Among their four subjects, only one subject constantly produced stops with prevoicing. Unlike the English case, all the Japanese participants produced stops with or without prevoicing inconsistently. Based on the data, the frequency of prevoicing seems to be relatively higher in connected speech, especially in the word-medial position.[6] Since voiced stops followed a vowel in the sentence, voicing tended to continue through the stop closure. Second, positive VOT values for /b, d, g/ are similar to ones for TM unaspirated stops. By looking through Table 4, positive VOT in the initial position including the monosyllable condition seems to follow the general pattern: velar stop /g/ tends to have a longer VOT than the other stops (/b/ vs. /g/: $t(8) = -6.045$, $p < .001$, $r = .91$; /d/ vs. /g/: $t(8) = -5.911$, $p < .001$, $r = .90$; /b/ vs. /d/: $p > .1$). However, the amount of prevoicing (negative VOT) seems random. It does not show any pattern in relation to the place of articulation.

[6] The percentage of negative VOT was as follows: in a syllable (27.4%); real word-initial (45.6%); real word-medial (75.0%); nonsense word-initial (47.2%); nonsense word-medial (83.5%).

Table 4. Average VOT (ms) for Japanese voiced stops

		Bilabial b	Alveolar d	Velar g
Monosyllable	ɑ	11.22 (5.23)	12.62 (6.21) -78.20 (29.83)	19.64 (5.51) -84.34 (.00)
	e	9.41 (3.34) -98.40 (.00)	13.02 (5.37) -81.32 (34.88)	19.55 (5.61)
	o	12.49 (2.75) -47.23 (.00)	11.52 (3.52) -67.28 (26.16)	21.81 (7.17) -50.11 (.19)
Real word initial	ɑ	N/A	9.79 (3.16) -61.10 (14.67)	16.41 (5.64) -70.79 (9.04)
	o	N/A	12.84 (5.21) -60.10 (11.56)	21.52 (7.78) -57.39 (18.71)
Real word medial	ɑ	N/A	14.59 (7.08) -39.75 (8.47)	2.95 (.00) -40.81(22.24)
	o	N/A	15.27 (.00) -36.92 (8.48)	12.56 (8.43) -36.08 (11.35)
Nonsense word initial	ɑ	11.64 (3.18) -66.76 (11.71)	11.83 (3.43) -68.60 (11.77)	18.02 (5.71)
	e	14.25 (.88) -78.69 (22.61)	14.18 (1.17) -68.82 (23.34)	18.27 (4.86) -75.61 (19.37)
	o	5.05 (.15) -73.52 (13.96)	13.49 (.50) -74.08 (14.12)	18.96 (7.54) -72.84 (.00)
Nonsense word medial	ɑ	6.06 (.00) -52.27 (6.74)	-32.29 (6.71)	-29.78 (4.98)
	e	6.96 (.00) -44.80 (11.72)	-39.38 (10.82)	-33.33 (6.98)
	o	6.10 (.00) -46.54 (8.39)	-31.92 (6.72)	15.66 (4.32) -37.76 (13.57)

Note* The average positive and negative VOT values are given separately. Numbers in the parentheses show the standard deviation.

3.2.2 VOT for Japanese voiceless stops

Table 5 shows the mean VOTs for voiceless stops in five conditions. Average VOT in the initial position seems to be close to the VOT obtained in the study by Riney, Takagi, Ota, and Uchida (2007). In their study, values were collapsed across vowels and the obtained mean VOT was: 30.0 ms for /p/, 28.5 ms for /t/, and 56.7 ms for /k/. In the current study, no correlation (r_{pb} = .025, p > .5) was found between VOT values and word type (real word and nonsense word), and multiple regression with VOT variables and vowel type as a predictor did not reach significance (F = .621, p > .5). Thus, VOT was calculated by collapsing these two factors. The obtained mean VOT in the monosyllable condition was 33.07 ms (SD=10.31) for /p/; 31.99 ms (SD=10.37) for /t/; and 51.93 ms (SD=12.31) for /k/, and these VOT values indicate

that Japanese voiceless stops belong to neither short-lag nor long-lag categories; it is an "intermediate degree", which is compatible with the data that Riney, Takagi, Ota, and Uchida (2007) found in their study.

Table 5. Average VOT (ms) for Japanese voiceless stops

		Bilabial p	Alveolar t	Velar k
Monosyllable	ɑ	30.09 (17.80)	33.05 (10.07)	50.05 (15.88)
	e	29.10 (9.89)	34.20 (11.43)	53.27 (12.33)
	o	40.03 (10.72)	28.72 (15.45)	53.38 (14.32)
Real word initial	ɑ	N/A	22.56 (5.65)	40.10 (11.19)
	o	N/A	24.72 (7.47)	41.17 (8.69)
Real word medial	ɑ	N/A	13.37 (4.23)	21.30 (8.98)
	o	N/A	12.50 (4.21)	21.23 (5.90)
Nonsense word initial	ɑ	25.82 (10.87)	32.74 (13.73)	40.05 (11.44)
	e	24.58 (8.91)	31.59 (15.27)	42.47 (12.85)
	o	38.07 (11.80)	32.62 (11.99)	39.50 (10.19)
Nonsense word medial	ɑ	10.42 (1.94)	13.81 (5.19)	23.36 (6.62)
	e	10.49 (3.87)	16.55 (3.71)	20.46 (8.80)
	o	14.04 (4.95)	14.48 (5.34)	21.90 (9.06)

Note* Numbers in the parentheses show the standard deviation.

Figure 2 shows VOT for three voiceless stops in each position (initial and medial). Two-way ANOVA was conducted with consonant (bilabial, alveolar, velar) and position (word-initial and –medial) as within-subject factors and counterbalanced group as a between-subject factor. The main effect of consonant ($F(2, 16) = 68.63$, $p < .001$), the main effect of position in word ($F(1, 8) = 74.61$, $p < .001$), and the interaction between the two factors ($F(2, 16) = 6.7$, $p < .01$) were all significant. The simple effect of consonant was further tested by splitting across the position factor. The consonant factor was significant both in the word-initial position ($F(2, 16) = 40.58$, $p < .001$) and word-medial position ($F(2, 16) = 48.83$, $p < .001$). Planned comparison revealed that the VOT difference between /p/ and /t/ was not significant in the initial position ($F(1, 8) = .26$, $p > .5$), but it was significant in the medial position ($F(1, 8) = 70.1$, $p < .001$). The VOT difference also reached significance between /p/ and /k/ (initial position: $F(1, 8) = 82.68$, $p < .001$; medial position: $F(1, 8) = 14.19$, $p = .005$), and between /t/ and /k/ (initial position: $F(1, 8) = 124.53$, $p < .001$; medial position: $F(1, 8) = 37.46$, $p < .001$). In the word-initial position, /k/ is significantly longer than /p/ or /t/, but in the word-medial position, VOT becomes longer as the position of stop closure moves further back.

Next, the effect of the position on VOT was examined. Average VOT for word-initial stops was: 31.28 ms (SD=9.23) for /p/, 30.04 ms (SD=7.57) for /t/, and 44.94 ms (SD=7.89) for /k/. Mean VOT for word-medial stops was: 11.65 ms (SD=3.13) for /p/, 13.86 ms (SD=3.99) for /t/, and 21.65 ms (SD=6.17) for /k/. Planned comparison revealed that VOT was significantly shorter in the medial position than in the initial position: /p/ ($F(1, 8) = 46.53$, $p < .001$); /t/ ($F(1, 8) = 69.13$, $p < .001$); /k/ ($F(1, 8) = 74.24$, $p < .001$). Moreover, VOT values in the word-medial position definitely fall in the short-lag category, and the amount of voicing lag is almost identical with voiced stops. This suggests that medial voiceless stops can be confused with voiced stops if listeners heavily rely on VOT cues.

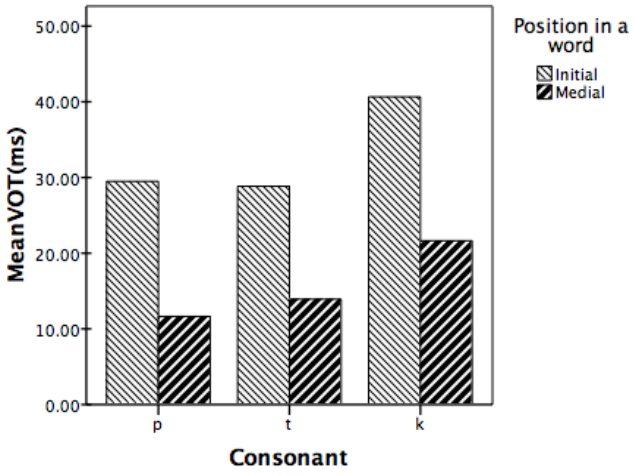

Figure 2. Mean VOT (ms) for word-initial and word-medial voiceless stops in Japanese

In sum, Experiment 2 found some acoustic characteristics of VOT for Japanese stops. Voiced stops have both negative and positive VOT, and the occurrence of prevoicing is unpredictable. Among the voiced stops with positive VOT, the velar stop has significantly longer VOT than the bilabial and alveolar stops do. Regarding voiceless stops, their VOT values belong neither to long-lag nor short-lag categories; they are an intermediate degree of VOT. Most interestingly, VOT for word-medial voiceless stops was consistently shorter than for word-initial stops. Voicing lag in this position shifts to the short-lag category. This phonetic change may make word-medial stops poor exemplars of voiceless stops, which would lead some less-experienced L2 learners to confuse those voiceless stops with voiced ones.

4. General discussion

The two production experiments in this study examined VOT of Taiwan Mandarin and Japanese stops. Figure 3 below is a scatter plot of VOT range for the two languages. VOT for TM aspirated stops ranges from 60-80 ms and for unaspirated stops the range of 9-20 ms is much shorter. These results confirm that TM stops definitely belong to either the long-lag (aspirated) or short-lag (unaspirated) category, and the distinction is clear. Although the phrase-medial position has some influence of decreasing VOT on TM aspirated stops, they do not overlap with unaspirated stops on the continuum. Hence, the phonemic contrast in the medial position still holds phonetically.

Conversely, Japanese stops show some complication. First, VOT for voiced stops are dispersed from negative to positive on the continuum, and prevoicing is totally unpredictable; it could occur across speakers, phonemes, following vowels, or positions in a word. Second, VOT for voiceless stops in the initial position ranges between 20-50 ms, which falls neither in the short-lag nor long-lag categories. Instead, its status is intermediate. Moreover, the positional effect on VOT found in this study is more complicated. VOT for voiceless stops in the medial position drastically decreases to 10-20 ms, which is almost the identical value for voiced stops with positive VOT.

It is possible that these complications of VOT in Japanese stops may cause some difficulty in production and perception of those stops for learners whose native language has prototypical stops like TM. If TM learners rely heavily on VOT cues to establish phonemic categories of Japanese stops, inexperienced learners might refer to the VOT for stops in their native language. They might link Japanese voiceless stops in the initial position with TM aspirated stops, and link voiceless stops in the medial position with unaspirated stops since the VOT range for medial Japanese voiceless stops partially (not completely) overlaps with TM unaspirated stops as Figure 3 shows. This implies that Japanese voiceless stops might be grouped in two different categories based on the position, and medial voiceless stops might be recognized as its voiced counterpart.

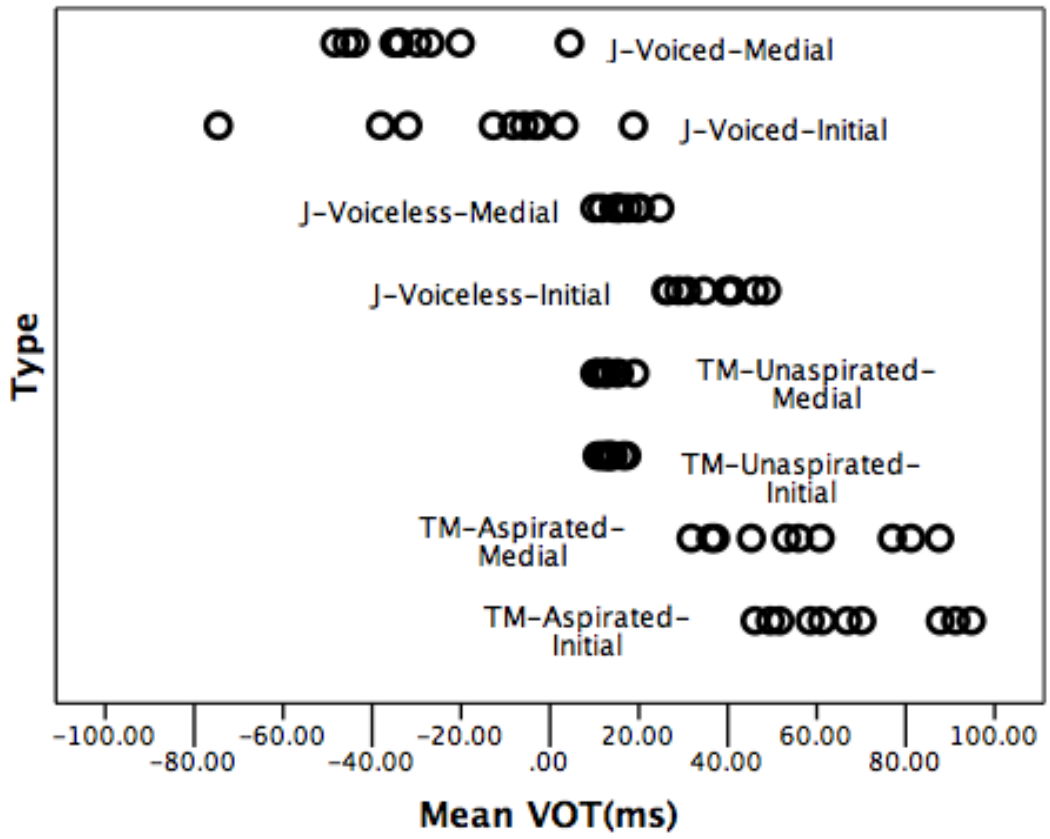

Figure 3. Range of mean VOT (ms) in Taiwan Mandarin and Japanese

5. Conclusions

The aim of this study was to provide an acoustic investigation on VOT for stops in Taiwan Mandarin and Japanese. It was revealed that the place of articulation of stops affects VOT. VOT for velar stops is significantly longer than that for bilabial or alveolar stops, which was commonly observed between TM aspirated stops and Japanese voiceless stops. On the other hand, TM unaspirated and Japanese voiced stops follow the cross-linguistic pattern observed in many other languages that VOT length positively correlates with the backness of articulatory constriction. Although TM and Japanese share some common characteristics of VOT for stops, a significant difference was also found. VOT for TM stops clearly falls in either long-lag or short-lag categories; on the contrary, Japanese voiceless stops in the initial position belong to neither the long-lag nor short-lag category; instead, they have an intermediate state of VOT.

The methodology for the current study employed VOT for not only initial stops, but also medial stops in order to present evidence for the effect of position in a word or phrase on VOT variance. VOT for medial stops becomes significantly shorter than

for initial stops in both languages; however, VOT for TM medial aspirated stops never crosses the border of unaspirated stops. On the other hand, VOT for medial Japanese voiceless stops becomes almost identical with voiced stops, which would lead non-native listeners, whose native language has prototypical two-way categories like TM, to confusion of Japanese stops. This further brings up an important question in regards to how non-native listeners perceive Japanese initial and medial voiceless stops. Further research including a perception test is necessary to answer this question.

Appendix A. Experimental materials for experiment 1 (Taiwan Mandarin)

		Real word		
	Adjacent tone	p^h	t^h	k^h
Word-initial	1	啪啦 p^hala ‘hand-clap’	他殺 $t^haʂa$ ‘homicide’	咖啡 $k^hafeɪ$ ‘coffee’
	2	趴平 $p^hap^hiŋ$ ‘lie down’	他人 $t^hazə̯n$ ‘other person’	咖博 k^hap^wo ‘coffee exhibition’
	3	趴板 p^hapan ‘surfboard’	他者 $t^hatʂə$ ‘other’	
	4	趴下 $p^haɕia$ ‘get down’	塌下 $t^haɕia$ ‘collapse’	喀掉 $k^hatiaʊ$ ‘cut off, eat all’
Word-medial	1	嗨趴 $haɪp^ha$ ‘pass with a high score’	坍塌 t^hant^ha ‘collapse’	嗨咖 $haɪk^ha$ ‘excited person’
	2	奇葩 $tɕ^hip^ha$ ‘masterpiece’	其他 $tɕ^hit^ha$ ‘others’	牌咖 $p^haɪk^ha$ ‘card game player’
	3	打趴 tap^ha ‘beat up’	倒塌 $taʊt^ha$ ‘collapse’	網咖 $waŋk^ha$ ‘Internet café’
	4	累趴 $leɪp^ha$ ‘very tired’	揍他 $tsoʊt^ha$ ‘punch him’	怪咖 $k^waɪk^ha$ ‘weirdo’
	Adjacent tone	P	t	k
Word-initial	1	巴西 paɕi ‘Brazil’	搭車 $tatʂ^hɤ$ ‘get on’	嘎空 $kak^{hw}oŋ$ ‘stock market jargon’
	2	疤痕 pahən ‘scar’	搭乘 $tatʂ^həŋ$ ‘get on’	旮旯 kala ‘corner’
	3	芭比 papi ‘Barbie doll’	搭起 $tatɕ^hi$ ‘build up’	咖哩 kali ‘curry’

	4	巴士 paʂɯ ‘bus’	搭配 tapʰeɪ ‘match’	嘎票 kapʰjaʊ ‘stock market jargon’
Word-medial	1	三八 sanpa ‘silly’	穿搭 tʂʰʷanta ‘outfit’	嘰嘎 tɕika ‘a kind of sound’
	2	淋巴 limpa ‘lymph’	白搭 paɪta ‘useless’	
	3	尾巴 weɪpa ‘tail’	混搭 hʷənta ‘mix and match’	
	4	大巴 tapa ‘big bus’	拒搭 tɕta ‘refuse to get in’	

Nonsense word

	Adjacent tone	pʰ	tʰ	kʰ
Word-initial	1	趴天 pʰatʰjɛn	它央 tʰajaŋ	咖風 kʰafʷoŋ
	2	趴額 pʰaɤ	他鳴 tʰamiŋ	咖陪 kʰapʰeɪ
	3	趴剪 pʰatɕjɛn	塌本 tʰapən	咖滾 kʰakun
	4	趴妙 pʰamjaʊ	塌怨 tʰaɥɛn	咖要 kʰajaʊ
Word-medial	1	刀趴 taʊpʰa	篇他 pʰiɛntʰa	君咖 tɕnkʰa
	2	芸趴 ɥynpʰa	情他 tɕʰiŋtʰa	翔咖 ɕjaŋkʰa
	3	影趴 jiŋpʰa	膽塌 tantʰa	井咖 tɕiŋkʰa
	4	定趴 tiŋpʰa	垢塌 koʊtʰa	業咖 jɛkʰa

	Adjacent tone	p	t	k
Word-initial	1	巴喵 pamjaʊ	搭淵 taɥɛn	嘎冰 kapiŋ
	2	八停 patʰiŋ	搭零 taliŋ	嘎毛 kamaʊ
	3	巴軌 pakʷeɪ	搭跑 tapʰaʊ	嘎餅 kapiŋ
	4	巴欠 patɕʰjɛn	搭樂 talə	嘎店 katjɛn
Word-medial	1	暈八 ɥynpa	冰搭 piŋta	鬆嘎 soŋka
	2	泉巴 tɕʰɥɛnpa	言搭 jɛnta	瓶嘎 pʰiŋka
	3	野巴 jɛpa	海搭 haɪta	雅嘎 jaka
	4	智巴 tʂɯpa	餓搭 ɤta	徹嘎 tʂʰɤka

Appendix B. Experimental materials for experiment 2 (Japanese)

Real word

	t	d	K	g
Word-initial	tɑko 'octopus'	dɑɾe 'who'	kɑʧi 'value'	gɑʃi 'starvation'
	tono 'lord'	doʤi 'blob'	koto 'Japanese harp'	gosɑ 'error'
Word-medial	kɑtɑ 'shoulders'	hɑdɑ 'skin'	sɑgɑ 'Saga Prefecture'	wɑkɑ 'Japanese poetry'
	koto 'old city'	mɑdo 'window'	igo 'game of go'	kɑko 'past'

Nonsense word

	p	b	t	d	k	g
Word-initial	pɑse	bɑse	tɑju	dɑju	kɑɸɯ	gɑɸɯ
	pese	bese	teju	deju	keɸɯ	geɸɯ
	pose	bose	toju	doju	koɸɯ	goɸɯ
Word-medial	ɾipɑ	ɾibɑ	nitɑ	nidɑ	sekɑ	segɑ
	ɾipe	ɾibe	nite	nide	seke	sege
	ɾipo	ɾibo	nito	nido	seko	sego

References

van Alphen, Petra Martine. 2007. Prevoicing in Dutch initial plosives: Production, perception, and word recognition. *Voicing in Dutch: (De)voicing—Phonology, Phonetics, and Psycholinguistics*, ed. By Jeroen van de Weijer and Eric Jan van der Torre, 99-124. Amsterdam: John Benjamins Publishing Company.

Aoyama, Katsura, James Emil Flege, Susan. G. Guion, Reiko Akahane-Yamada, and Tsuneo Yamada. 2004. Perceived phonetic dissimilarity and L2 speech learning: The case of Japanese /r/ and English /l/ and /r/. *Journal of Phonetics* 32:233-250.

Best, Catherine T., Pierre Hallé, Ocke-Schwen Bohn, and Alice Faber. 2003. Cross-language perception of nonnative vowels: Phonological and phonetic effects of listeners' native languages. *Proceedings of the 15th International Congress of Phonetic Sciences*, 2889-2892. Barcelona: N.p.

Boersma, Paul, and David Weenink. 2011. Praat: Doing Phonetics by Computer, Computer program, Version 5.2.17. Retrieved March 02, 2011, from http://www.praat.org/.

Chao, Kuan-Yi, and Li-mei Chen. 2008. A cross-linguistic study of voice onset time in stop consonant productions. *Computational Linguistics and Chinese Language Processing* 13:215-232.

Chen, Li-mei, Kuan-Yi Chao, and Jui-Feng Peng. 2007. VOT productions of word-initial stops in Mandarin and English: A cross-language study. *Proceedings of the 19th Conference on Computational Linguistics and Speech Processing*, 303-317. N.p.: N.p.

Cho, Taehong, and Peter Ladefoged. 1999. Variation and universals in VOT: Evidence from 18 languages. *Journal of Phonetics* 27:207-229.

Duanmu, San. 2000. *The Phonology of Standard Chinese*. Oxford: Oxford University Press.

Eckman, Fred R., and Gregory K. Iverson. 1993. Sonority and markedness among onset clusters in the interlanguage of ESL learners. *Second Language Research* 9:234-252.

Fischer-Jørgensen, Eli. 1954. Acoustic analysis of stop consonants. *Miscellanea Phonetica* 2:42-59.

Flege, James Emil. 1989. Chinese subjects' perception of the word-final English /t/-/d/ contrast: Performance before and after training. *Journal of the Acoustical Society of America* 86:1684-1697.

Flege, James Emil, Naoyuki Takagi, and Virginia Mann. 1995. Japanese adults can learn to produce English /ɹ/ and /l/ accurately. *Language and Speech* 38:25-55.

Flege, James Emil, Naoyuki Takagi, and Virginia Mann. 1996. Lexical familiarity and English-language experience affect Japanese adults' perception of /ɹ/ and /l/. *Journal of the Acoustical Society of America* 99:1161-1173.

Hallé, Pierre, Catherine T. Best, and Andrea Levitt. 1999. Phonetic vs. phonological influences on French listeners' perception of American English approximants. *Journal of Phonetics* 27:281-306.

Harada, Tetsuo. 2003. L2 influence on L1 speech in the production of VOT. *Proceedings of the 15th International Congress of Phonetic Sciences*, 1085-1088. Barcelona: N.p.

Homma, Yayoi. 1980. Voice onset time in Japanese stops. *Onsei Gakkai Kaihō* 163:7-9.

Homma, Yayoi. 1981. Durational relationships between Japanese stops and vowels. *Journal of Phonetics* 9:273-281.

Keating, Patricia A. 1984. Phonetic and phonological representation of stop consonant voicing. *Language* 60:286-319.

Kent, Ray D., and Charles Read. 1992. *The Acoustic Analysis of Speech*. London: Singular Publishing Group, Inc.

Lin, Yen-Hwei. 2007. *The Sounds of Chinese*. Cambridge: Cambridge University Press.

Lisker, Leigh, and Arthur S. Abramson. 1964. A cross-language study of voicing in initial stops: Acoustical measurements. *Word* 20:384-422.

Peterson, Gordon E., and Ilse Lehiste. 1960. Duration of syllable nuclei in English. *Journal of the Acoustical Society of America* 32:693-703.

Port, Robert F., and Rosemarie Rotunno. 1979. Relation between voice-onset time and vowel duration. *Journal of the Acoustical Society of America* 66:654-662.

Riney, Timothy James, Naoyuki Takagi, Kaori Ota, and Yoko Uchida. 2007. The intermediate degree of VOT in Japanese initial voiceless stops. *Journal of Phonetics* 35:439-443.

Shimizu, Katsumasa. 1996. *A Cross-Language Study of Voicing Contrasts of Stop Consonants in Six Asian Languages*. Tokyo: Seibido.

Vance, Timothy J. 2008. *The Sounds of Japanese*. Cambridge: Cambridge University Press.

[Received 1 March 2011; revised 29 April 2011; accepted 30 June 2011]

Department of English
National Taiwan Normal University
Taipei City, TAIWAN
Naomi Ogasawara: naomi703@ntnu.edu.tw

國語和日語聲帶振動起始時間的聲學分析

小笠原奈保美
國立臺灣師範大學

本研究旨在呈現國語和日語連續話語中，塞音聲帶振動起始時間的完整聲學調查。除了觀察國語送氣與不送氣塞音和日語有聲與無聲塞音在單獨音節的表現外，也觀察他們在字首／中（日語）跟詞首／中（國語）的表現。發現國語的送氣塞音屬長延遲類無疑、不送氣塞音屬短延遲類。而日語的無聲塞音在首位則介於短延遲和長延遲間，在中位則極短趨近於有聲塞音的正聲帶振動起始時間。至於有聲塞音的前發聲則是隨機的。這樣的結果顯示，就聲帶振動起始時間而言，日語和國語的塞音聲學上是相當不同的；也暗示著，說國語者正確聽辨日語塞音可能有困難。

關鍵詞：聲帶振動起始時間、國語、日語