The Development of Voice Onset Time (VOT) in a Content-Based Instruction University Program by Japanese Learners of English: A Longitudinal Study

Keiko Hanzawa

The Canadian Modern Language Review / La revue canadienne des langues vivantes, Volume 74, Number 4, November/novembre 2018, pp. 501-522 (Article)

Published by University of Toronto Press

The Development of Voice Onset Time (VOT) in a Content-Based Instruction University Program by Japanese Learners of English: A Longitudinal Study

Keiko Hanzawa

Résumé : Dans une étude longitudinale, l'auteur se demande si l'inscription à un programme de formation basé sur le contenu dans une université japonaise améliore le temps de voisement du /p/, du /t/ et du /k/ en anglais chez les apprenants d'une langue seconde. Les mesures du temps de voisement sont prises au début, au milieu et à la fin de l'année universitaire et soumises à une analyse acoustique. Des améliorations significatives du temps de voisement sont observées au cours de l'année durant laquelle les étudiants reçoivent environ 300 heures de contenu dans le cadre du programme. Toutefois, les résultats montrent aussi que les améliorations du temps de voisement ne sont pas influencées par le volume du contenu en langue seconde livré par des locuteurs natifs. Ces observations semblent indiquer que les apprenants peuvent devenir de meilleurs locuteurs de la langue seconde s'ils sont exposés à cette langue dans le cadre d'un programme de formation basé sur le contenu en phase initiale d'apprentissage, grâce à une exposition précoce et intensive à la langue seconde, peu importe la provenance de l'intrant.

Mots clés : cadre du programme de formation basé sur le contenu, contexte de langue étrangère, formation basée sur le contenu, perfectionnement en langue seconde, temps de voisement

Abstract: This longitudinal study examined whether and to what extent the production of voice onset time (VOT) for /p/, /t/, and /k/ in English was improved in second language (L2) learners by enrolment in a content-based instruction (CBI) program at a Japanese university. Measures of VOTs were elicited at the beginning, middle, and end of the academic year and acoustically analyzed. Significant improvements in VOT production were observed over the year in which students received approximately 300 hours of input in the CBI setting. However, the results also showed that improvements in VOT were not affected by the amount of L2 input from native speakers. These findings suggest that L2 speech learning can occur if learners receive exposure to L2 in a CBI environment in the early stage of L2 learning through early, intensive exposure to L2, regardless of the source of input.

© *The Canadian Modern Language Review/La Revue canadienne des langues vivantes*,
74, 4, (November / novembre), 501–522 doi:10.3138/cmlr.2018-0196

Keywords: CBI environment, content-based instruction, foreign language setting, L2 development, voice onset time

Many researchers have noted the role of experience with second language (L2) in L2 speech learning in naturalistic settings in which the target language is predominantly used on a daily basis (e.g., Munro & Derwing, 2008; Trofimovich & Baker, 2006). Some researchers (Harada, 2006, 2007; Netelenbos, Li, & Rosen, 2015; Simon & D'Hulster, 2012) have recently begun to investigate the extent to which the findings from naturalistic settings could be generalized to a specific foreign language setting in which L2 is used in a more naturalistic way as a medium of academic content instruction: content-based instruction (CBI). This current study reports the results of a longitudinal study investigating whether and to what extent L2 input in a university CBI program facilitated the improvement of phrase-initial voice onset time (VOT) for voiceless stops in Japanese learners of English over the course of an academic year.

Background

L2 speech learning in naturalistic settings

The crucial role of L2 exposure in L2 speech learning has been widely examined in the literature, such as in the Speech Learning Model (Flege, 1995), in which L2 speech learning takes place through experience with the L2. According to Flege (2009, p. 175), L2 input is defined as "all L2 vocal utterances the learner has heard and comprehended, including his own, regardless of whether these utterances have been produced correctly by L2 native speakers, or incorrectly by other non-native speakers of the L2." In line with this definition, L2 input can be described as an umbrella term that subsumes any exposure to the L2 that learners can receive in the course of their L2 learning, from other speakers as well as from themselves. Much work on L2 speech acquisition has demonstrated the benefits of experience for L2 speech learning of various speech features, including VOT for voiceless stops (the target phonetic feature in this study) (Baker & Trofimovich, 2006; Flege & Fletcher, 1992).

Voice onset time (VOT) is one of the primary articulatory characteristics of stop consonants. It refers to the interval of time between the release of stop closure and the onset of vocal-fold vibration (Lisker & Abramson, 1964). Although VOT can take any numerical value, there are – phonetically – three different qualitative ranges of VOT: prevoiced, voiceless unaspirated, and voiceless aspirated (Keating, 1984). Crucially, some languages that recognize a phonological distinction between

voiced and voiceless stops may differ as to where the two categories separate along the VOT continuum.[1] Languages such as English realize voiceless aspirated stops with longer VOT values as voiceless (/p, t, k/) and voiceless unaspirated stops with shorter VOT values as voiced (/b, d, g/) (Flege, Munro, & MacKay, 1995; Riney & Takagi, 1999). In contrast, languages such as Japanese and French realize voiceless unaspirated stops as voiceless, whereas they realize prevoiced stops with negative VOT values as voiced (Netelenbos et al., 2015; Riney, Takagi, Ota, & Uchida, 2007).

Because of the cross-linguistic variation in VOT, many researchers have examined how L2 learners produce and acquire L2 segmental production in terms of VOT values. Flege (1987), for example, conducted a cross-sectional study of late French learners of English with varied experience with English. He found that, while learners with the least experience in L2 produced L2 /t/ with the almost identical VOT values to those of the L1, the VOT values produced by more experienced learners were significantly closer to those of the L2. In other words, learners can learn to produce voiceless stops with more L2-like VOT values as they encounter increased amounts of L2 input. Interestingly, Flege also found that although the highly experienced learners approximated their L2 VOT values to those of native French speakers, they never reached native levels of accuracy, falling between the levels for monolingual speakers of both the L1 and the L2. Such intermediate or "compromised" VOT values (Flege, 1991, p. 395) have been observed in many late learners, that is, those who started learning L2 after puberty (Flege, 1991; Flege, Munro, & MacKay, 1995; Kang & Guion, 2006).

However, the relationship between experience and segmental learning is not always so straightforward. The results of several studies on L2 speech learning have provided counter-evidence to experience-driven L2 speech learning (Baker & Trofimovich, 2005; Cebrian, 2006; Piske, MacKay, & Flege, 2001). These and other studies exploring the effects of experience with L2 on the development of L2 speech learning have revealed the following two findings. First, although the acquisition of more accurate and refined segmental production requires more than seven years of L2 input (Baker & Trofimovich, 2006), learners show the most obvious development in the early stages of L2 learning, typically within the first year of extensive exposure to L2 (Piske et al., 2001). This notion was first proposed by Flege (1988), based on his finding that no significant difference exists in foreign accent between learners with one and five years of residence in the United States. Recently, Munro and Derwing (2008) provided additional evidence for this by longitudinally examining vowel intelligibility by Mandarin and Slavic learners of

English in Canada. They found that the largest improvement occurred within the first six months of the learners' first intensive exposure to L2 (for similar results with English /ɹ/ by Japanese learners of English, see Saito & Munro, 2014).

Second, some researchers have claimed that not only the quantity but also the quality of input can play an important role in L2 speech learning (Moyer, 2009; Piske, Flege, MacKay, & Meador, 2002). The quality of input can be highly multifaceted in the field of L2 acquisition, such as is the case in interactions with native speakers or in meaningful communication (Piske & Young-Scholten, 2009), and is separate from the quantity of input. In the field of L2 speech learning, however, quality specifically involves L2 input related to native speakers of the target language. Researchers have often operationalized L2 input by combining quality with quantity. Therefore, when the importance of the quality of L2 input is emphasized in the L2 speech learning literature, it implies the amount of L2 input from native speakers of the target language (Jia & Aaronson, 2003, Moyer, 2004, 2011; Piske, 2007). For example, Flege and Liu (2001) compared two groups of Chinese learners of English in the United States and found that an improvement in L2 word identification was observed only in learners who were assumed to have received abundant input from native English speakers. This finding led them to speculate that L2 phonological development could occur to a measurable degree only if the learners received a substantial amount of input from native speakers. Furthermore, Moyer (2011) investigated the relationship between the degree of foreign accent and the quality of L2 input in correlational analyses, showing that the total number of weekly hours of L2 input from native speakers was significantly correlated with learners' foreign accent. Even though previous studies have used the term "quality of L2 input" to clarify the meaning, in this study the amount of L2 input from native English speakers, and not its quality, is used.

Content-based instruction and phonological development

Content-based instruction is defined as "an instructional approach in which non-linguistic curricular content such as geography or science is taught to students through the medium of a language that they are concurrently learning as an additional language" (Lyster & Ballinger, 2011, p. 279). It first appeared in the context of Canadian immersion education during the 1960s as a means to introduce subject-matter lessons to English-speaking children in L2 French (Lyster, 2007). The approach has since been adapted to European countries in such terms as Content and Language Integrated Learning (CLIL) in order to prepare young people for European integration and multilingualism

(Dalton-Puffer, 2011; Marsh & Frigols Martín, 2013), as well as in East Asian countries (Yang, 2014) where English is most likely to be used as an additional and foreign language (FL).

As noted by many researchers (e.g., Dalton-Puffer, 2011; Lyster, 2007), this approach has many variations driven by national contexts, institutional conditions, and educational levels. Met (1998) describes a range of CBI settings along a continuum of content and language integration, varying from a language-driven model to a content-driven model. At the language-driven end of the model, the focus is on developing the target language proficiency through the use of academic topics. Examples of this model are English for academic purposes (Airey, 2016) and theme-based instruction (Brinton, Snow, & Wesche, 2003), which select academic subjects that are not part of the formal curriculum to be taught with the aim of developing academic English skills. At the other end, in the content-driven model, learners receive a substantial amount of academic content instruction in L2 by teachers primarily trained in the content areas. Examples include total immersion (Lyster & Ballinger, 2011) and English-medium instruction (EMI) at the university level (Unterberger, 2012).

Although few in number, some attempts have been made to investigate the effectiveness of L2 input in CBI settings on L2 speech learning in line with the previous findings in naturalistic settings. One such attempt was that of Harada (2007), who examined English learners of Japanese in a Japanese CBI setting (i.e., immersion education) for their production of initial voiceless stops in first, third, and fifth grades. The results demonstrated that learners in the CBI setting produced L2 Japanese voiceless stops with significantly shorter VOT (thus more target-like values) than those of L1 English, indicating that the learners clearly distinguished between Japanese and English VOTs. However, he also found that the learners' L2 production did not change across grade levels, which suggests that the children did not improve their L2 VOT between the first and fifth grades (see also Netelenbos et al., 2015). Similarly, Simon and D'Hulster (2012) examined the role of CBI experience on vowel production by Dutch learners of English with different learning experiences in CBI. The results showed that the production of English /ɛ/ and /æ/ by learners with 0.5 and 3.5 years of CBI experience differed significantly from forms produced by learners with no CBI experience; however, differences in CBI experience (i.e., 0.5 vs. 3.5 years) did not correlate with differences in production in any significant way.

It is clear from the following two points that more research is required to further examine the role of input in driving L2 segmental

development in CBI contexts. First, the previous investigations of CBI settings with cross-sectional design examined the effects of relatively longer experience (from half a year to five years of CBI experience) on learners' production of L2 segments. As previously discussed, if experience effects are most evident in the early stage of L2 speech learning, which suggests that the greatest improvement occurs during the first year of intensive exposure to L2, then it is important to investigate learners' segmental development during the first year in a CBI setting. Second, previous studies did not take into account any differences in the quality of L2 input, that is, the amount of L2 input from native speakers of the target language. In contrast to the relative homogeneity of L2 input in naturalistic settings (Bohn & Bundgaard-Nielsen, 2009), students in CBI receive L2 input from speakers with various linguistic backgrounds (i.e., native and non-native speakers of the target language). If the amount of L2 input from native speakers plays a specific role in L2 speech learning (in that successful learning depends on whether learners receive a substantial amount of L2 input from native speakers), it can be reasonable to assume that the differing amount of L2 input from native speakers of the target language in a CBI setting can specifically contribute to differences in learners' development.

In response to a call for longitudinal work in the field of SLA (Ortega & Iberri-Shea, 2005) and L2 speech (Piske et al., 2001), the current study aims to investigate the effect of L2 in a CBI program on L2 VOT development by means of a longitudinal investigation of 50 first-year Japanese university students enrolled in either a CBI program or a non-CBI program. Their word-initial voiceless stops were elicited on three separate occasions (the beginning of the first semester [T1], the end of the first semester [T2], and the end of the second semester [T3]) and submitted for acoustic analysis. The following research questions were addressed:

1. Can students in a university CBI program improve their L2 VOT over the course of an academic year?
2. To what extent can they approximate target-like VOT production?
3. To what extent does the amount of L2 input from native speakers in the CBI program contribute to VOT development?

Methods

Participants

In total, 50 first-year Japanese students at a university in Japan participated in the study (age range: 18–19 years). Given that they began studying English in secondary school (at the age of 12–13), the participants

were identified as late learners, thereby eliminating any possible confusion between age effect and language experience (see, e.g., Flege, 2009). The participants had never been abroad for more than one month and had not previously received explicit pronunciation instruction (i.e., in English phonetics or pronunciation). None had received CBI in the past, during either junior or senior high school. All of them, therefore, had their first intensive exposure to spoken English after entering the program.

Participants were further subdivided into two groups according to the instruction they received at the university. There were 30 students in a CBI program (16 female and 14 male), and 20 students in a non-CBI program (11 female and nine male).[2] In the CBI program, the students were required to take 5–6 hours of academic content courses and could choose from a variety of topics (e.g., psychology, economy, business, and philosophy). All courses were conducted in English and were taught by native English speakers (e.g., from the United States, Britain, or Canada) as well as by speakers of other languages with learning and/or teaching experience in an English-speaking country. In addition to academic content courses, students were also required to complete four to five hours of classes in English for academic purposes (EAP), which were taught by native English speakers. The EAP classes focused on strategic academic language skills (e.g., note-taking, argumentative writing, debate, and public speech), with no intentional focus on pronunciation, vocabulary, or grammar. A questionnaire using a 7-point Likert scale was administrated to participants to measure their levels of participation in academic and EAP classes (wherein 1 = I do not speak at all, and 7 = I always speak up). In academic content classes, participants generally adopted receptive roles in their learning, during both the first and the second semesters (M = 2.4 and M = 2.9, respectively). The fact that students' output was limited is consistent with previous research involving CBI at Japanese universities (Moriyoshi, 2010). On the other hand, participants spoke more frequently in the EAP classes (M = 6.2 and M = 6.3 in the first and second semesters, respectively).

In the non-CBI program, the students were majoring in business, marketing, and economics (11 females, 9 males). While taking academic subject classes in Japanese, the students registered for three hours of English lessons taught by native English-speaking and Japanese teachers during the time of the project. As the standard syllabus of the department stated, the primary goal of the English lessons was to develop receptive skills by doing comprehension-based practice, such as extensive and intensive reading and listening activities. Accordingly, they received L2 exposure for about 65.2 hours in the first semester and 50 hours in the second semester. According to the questionnaire

regarding subjects' participation in their English classes, participants generally adopted relatively active roles in their learning, during both the first and the second semesters (M = 4.9 and M = 5.0, respectively).

Ten native English speakers (4 females, 6 males) from Canada with an average age of 27 years were recruited to examine how closely learners' VOTs approximated target-like VOT values over the year. They were graduate students at a university in Montreal, born and raised in English-speaking families. It is crucial to acknowledge that the students in the CBI program were exposed to a wide variety of English native speakers and not only to Canadian English. Their VOT productions of Canadian English were, therefore, considered as one of the varieties that the learners in the program were exposed to.

Amount of input from native and non-native English speakers in the CBI context

To determine the amount and source of L2 input that participants received during the CBI program, students were asked questions about (a) their exposure to and interaction with native and non-native English speaking lecturers or instructors in class, and (b) the frequency of their interactions with native and non-native English speakers outside of class. The first item was calculated by multiplying the number of courses students took during the first and second semesters by the length of each CBI lesson (i.e., 1.5 hours). The second was based on participants' accounts of how many minutes they conversed with native and non-native English interlocutors each week throughout the experiment.

In a CBI group, the amount of L2 input from native and non-native English-speaking teachers or lecturers varied between the first semester (M = 166.9, 11.0 weekly, range: 120–187 hours) and the second (M = 133.4, 8.8 weekly, range 82–178 hours). As with other CBI contexts (Harada, 2007), these results show that students had limited exposure to English outside of class (M = 12.86 hours yearly, and approximately 25.74 minutes weekly). Instead, the exposure to English that students received was provided almost exclusively during class by teachers or lecturers.

Materials and procedure

Research has repeatedly demonstrated that VOT is sensitive to factors such as place of articulation and subsequent vowel height (Nearey & Rochet, 1994; Yaneş & Wildermuth, 2006), as well as prosody (Sundberg & Lacerda, 1999) and the number of syllables (Kessinger & Blumstein, 1997). To minimize potential compounding effects caused by these factors, this study focused on the effect of place of articulation (i.e., the voiceless stops /p/, /t/, and /k/). Therefore, target words were chosen

according to the following criteria: (a) vowel quality ([æ] was selected), (b) word-initial voiceless stops, (c) stress on the first syllable, and (d) an identical number of tokens for /p/, /t/, and /k/. Eighteen disyllabic words were identified, evenly divided among the three voiceless stops in word-initial position (see Table 1).

To elicit the production of voiceless stops, participants were asked to read each word three times in a carrier phrase (i.e., "say ___ once again") in a randomized fashion, and to do so at a constant speed in order to minimize the effect of speech rate. Recordings were made individually in the university's laboratories using digital audio recorders (Marantz PMD 660 for CBI students, and Roland-05 for English native speakers) at 16-bit and 44.1 kHz sampling rates. Directions were given to both groups in their native language to ensure that the procedures were clear.

Data analysis

Each experimental word occurred three times on every list. In total, there were 8,640 tokens elicited for the voiceless stops (18 words × 3 repetitions × 50 participants × 3 times for Japanese students + 18 words × 3 repetitions × 10 participants for native English speakers). Praat (version 5.3.82) (Boersma & Weenik, 2011) was used for data analysis. The VOT start and end points for each token were determined manually by using Praat's annotation feature. The start point was defined as the onset of a stop consonant burst, and the end point as the beginning of voicing energy in the F2 formants corresponding to the onset of a subsequent vowel. When the onset of the F2 formants was unclear, the release burst onset for the first positive peak in the periodic portion of a waveform was considered to be the start point. A Praat script was then used to automatically measure the VOTs for each token. Intra-rater reliability was assessed by randomly selecting 10% of the tokens for re-measurement a few months after the initial measurement. A strong correlation was found between the first and second sets of measurements (r = 0.94, mean difference = 1.3 ms).

Table 1: The words elicited in the study

Place of articulation		
/p/	**/t/**	**/k/**
panda	talent	camel
parrot	tadpole	cancer
package	tablet	campus
panic	taxi	camper
passion	taffy	candy
pallet	tanker	carrot

As shown in several previous studies (Nearey & Rochet, 1994; Stölten, Abrahamsson, & Hyltenstam, 2015; Yaneş & Wildermuth, 2006), speaking rate plays an important role in the duration of VOT, and "the faster the speech, the shorter the VOT" (Johnson & Wilson, 2002, p. 278). A further measurement, word duration time, followed the VOT analysis. Since the word-final segment varied in the tokens, two criteria were used to determine whether the final segment was plosive. When the final segment was plosive, word length was determined by measuring the time interval between the release burst (i.e., the same as VOT) and the closure of the following stops. When the final segments were not plosives (i.e., vowel, nasal, liquid, or fricative), word length was measured as the time interval between the onset of the release burst and the last visible voicing in F2 or fricative offset in the word final segment. The word duration times at three different testing times (T1, T2, and T3) in two groups (CBI and non-CBI) were submitted to a repeated-measures ANOVA with place of articulation and testing time as within-group factors and group as a between-group factor. No significant interaction effects were found for either testing time and group, $F(2, 96) = 0.005$, $p = 0.995$, or for time, group, and place of articulation, $F(4, 192) = 0.990$, $p = 0.414$. These findings indicated that students' word durations in both groups involved statistically similar speech at all three times, suggesting that it might have negligible effects on learners' VOT productions. The raw VOT values, therefore, were used in further statistical procedures to examine the development of VOT over the course of the academic year.

Results

VOT improvement over the course of the academic year

Figure 1 shows the mean VOT values for English voiceless stops among the Japanese students in two different instructional groups (CBI and non-CBI), categorized by place of articulation at T1, T2, and T3. As can be seen, whereas the CBI students produced longer VOTs as the year progressed, non-CBI students did not (for descriptive statistics of participants' mean VOT values, see the Appendix). A repeated measures ANOVA was conducted using the mean VOT values with testing time (i.e., T1, T2, and T3) and place of articulation (i.e., /p, t, k/) as within-group factors and group (CBI and non-CBI students) as the between-group factor. The ANOVA found overall interaction effects for testing time and group, $F(2, 96) = 6.579$, $p = 0.001$, but not for testing time, group, and place of articulation, $F(4, 192) = 6.020$, $p = 0.937$. According to Bonferroni post-hoc tests, although the VOT values in both groups were not significantly different at T1 ($p = 0.226$), the VOT values of the

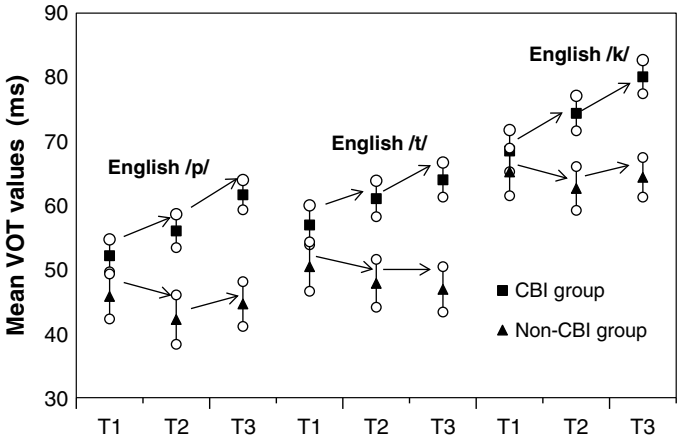

Figure 1: Mean VOT values for English voiceless stops produced by CBI and non-CBI students over time.

Note: The error bars enclose ± one standard error. CBI = content-based instruction; T1 = time at the beginning of the first semester; T2 = time at the end of the first semester; T3 = time at the end of the second semester.

CBI group were significantly longer than those of the non-CBI group at T2 ($p = 0.006$) and T3 ($p < 0.001$). Besides, CBI students exhibited a significant change over time. According to the Bonferroni post-hoc tests ($\alpha = 0.016$), VOTs at T3 were significantly longer than those at T1, with small to large effect sizes for /p/ ($p = 0.001$, $d = 0.60$), /t/ ($p = 0.007$, $d = 0.48$), and /k/ ($p < 0.001$, $d = 0.85$).

Comparison with native English speakers

To examine how closely the CBI students approximated target English VOTs, their values at T3 were compared to those of native English speakers. As expected, native English speakers produced voiceless stops with longer mean VOT values than CBI students (/p/ = 86.52 ms, /t/ = 86.52 ms, and /k/ = 92.58 ms). Due to unequal native and CBI groups, non-parametric Mann-Whitney U test statistical analysis was chosen to examine the difference between the two groups. According to the results, the VOT values produced by CBI students at T3 were significantly shorter than those of native English speakers, regardless of place of articulation, for /p/ ($p < 0.001$), /t/ ($p < 0.001$), and /k/ ($p = 0.002$). An additional analysis was conducted to determine whether the mean values of students' VOTs fell within the range of native English speakers' VOT values (Birdsong, 2007; Bongaert, 1999); that is, if the CBI students' VOTs were greater than or equal to the native speakers' mean

VOT minus one standard deviation, their performance was deemed native-like. For /p/, /t/, and /k/, the criteria were 73.00 ms, 81.81 ms, and 88.83 ms, respectively. The results showed that very few students (five out of 30) could produce English voiceless stops with VOT values that fell within the native speakers' range.

Impact of L2 input from native English speakers

As Table 1 shows, the amount of L2 input the CBI students received, whether from native or non-native English speakers, differed substantially among participants throughout the year. The aim of the next analyses was to examine the extent to which the different sources of L2 exposure could be related to CBI students' VOT development over time. First, a correlation analysis was conducted with the VOT development over the year as a dependent variable and the total hours of L2 exposure from native English speakers as an independent variable. The developmental scores were calculated by subtracting the T1 VOTs from the T3 VOTs. For example, if the VOT value of a participant was 40 ms at T1, and 60 ms at T3, then his or her developmental score would be 20 ms. The alpha value was set to 0.05 for this analysis. The results showed that the input factor was not significantly predictive of the participants' VOT developments for /p/ ($r = 0.156$, $p = 0.410$), /t/ ($r = 0.052$, $p = 0.786$), or /k/ ($r = 0.017$, $p = 0.927$).

To further clarify the role of L2 exposure from native English speakers in the development of VOT values over the year, CBI participants were rearranged based on the amount of native English input they had received (in hours) during the year; 20 students from the top and bottom were selected. While 10 students (the LE group) received a lower amount of L2 input from native speakers ($M = 146.2$), the other 10 students (the ME group) received a larger amount of L2 input from native speakers ($M = 242.1$). An independent t-test revealed that two groups differed significantly in terms of the amount of L2 input from native speakers, $t(18) = 10.46$, $p < 0.001$. As summarized in Table 2, both groups differing in the total amount of L2 exposure from native English speakers increased their VOT values over the year, regardless of the place of articulation. A repeated-measures ANOVA was conducted using the VOT productions at T1, T2, and T3 as dependent variables, with place of articulation and groups as independent variables. Significant main effects were found for time, $F(2, 36) = 11.168$, $p < 0.001$, and group, $F(1, 28) = 6.616$, $p = 0.019$; however, there were no interaction effects for time and group, $F(2, 36) = 0.358$, $p = 0.701$, or time, group, and place of articulation, $F(4, 72) = 0.869$, $p = 0.487$. The Bonferroni post-hoc test ($\alpha = 0.016$) further demonstrated that participants in

Table 2: VOT developments over one academic year between ME and LE groups (in milliseconds)

Mean VOT values	***M (SD)***		
	T1	**T2**	**T3**
ME group			
/p/	43.70 (7.56)	49.10 (10.61)	55.37 (9.15)
/t/	47.44 (8.75)	56.38 (11.31)	57.85 (10.10)
/k/	61.30 (10.19)	67.38 (10.47)	74.38 (11.41)
LE group			
/p/	59.60 (16.73)	62.64 (23.01)	72.80 (24.87)
/t/	66.24 (16.50)	68.64 (18.71)	72.86 (19.28)
/k/	76.90 (12.46)	80.21 (16.39)	87.21 (16.82)

Note: T1 = time at the beginning of the first semester; T2 = time at the end of the first semester; T3 = time at the end of the second semester; VOT = voice onset time.

the ME group significantly increased their VOT values between T1 and T2 (mean difference [*M*] = 6.809, $p = 0.013$), and T1 and T3 ($M = 11.72$, $p = 0.005$), but not T1 and T2 ($M = 4.91$, $p = 0.178$). Significant changes were also found in the LE group for T1–T3 ($M = 10.04$, $p = 0.014$), but not T1–T2 ($M = 2.91$, $p = 0.252$) or T2–T3 ($M = 7.124$, $p = 0.057$). These results indicate that both groups did make a significant improvement in English VOT values over the course of the year, regardless of the differing amount of L2 exposure from native English speakers. Another repeated-measures ANOVA was then run using the developmental scores as a dependent variable, with place of articulation and group as independent variables. The results showed that neither the main effect of group, $F(1, 18) = 0.104$, $p = 0.750$, nor the interaction of place of articulation and group, $F(2, 36) = 1.480$, $p = 0.241$, reached significance, which indicated that the degree of the development did not differ between the two groups.

Discussion and conclusion

A number of pieces of empirical evidence in naturalistic settings showed that increased experience with L2 facilitated more accurate L2 production of segmental and suprasegmental features, and these developments could be affected by L2 input quality. In light of growing interest in the generalizability of these findings to new educational settings (such as CBI), where L2 exposure is maximized in FL learning conditions by providing academic content in the L2, the current study focused on 30 Japanese university-level students by longitudinally examining their L2 VOT development and the relationship between their development

and the L2 input they received over the course of one academic year in a CBI program. Overall, two crucial findings emerged. First, Japanese learners were found to significantly increase their L2 VOT regardless of place of articulation, that is, they attained more target-like L2 VOT production from receiving approximately 300 hours' exposure to English in the CBI setting. Second, L2 VOT development did not correlate with amounts of L2 input from native speakers.

In general, the finding that 300 hours of exposure to L2 in the CBI setting facilitated L2 VOT development supports the claim that in naturalistic settings, the L2 experience effect is most evident in the early stages of L2 learning, typically within the first year of the first intensive exposure to L2 (Flege, 1988; Piske et al., 2001). This suggests that, as in a naturalistic setting, intensive exposure to L2 at the early stages of L2 learning seems crucial, at least for L2 VOT development in the CBI setting. Yet it should be noted for the second research question that despite learners' sensitivity to VOT values, the present findings also indicated that few students reached target-like VOT values over the course of one academic year. Instead, most participants' English VOT fell between those of monolingual speakers in the L1 and the L2 – in other words, intermediate or "compromised" VOT values (Flege, 1991, p.395). Since similar results were found in L2 learners with more L2 experience in another CBI setting (Harada, 2007), the present findings suggest that although mere exposure to L2 in a CBI setting could be sufficient to trigger the first improvement in the early stages of L2 learning, it is not enough to allow learners to attain target-like VOT values.

Another interesting finding, for the third research question, is that input source (i.e., native or non-native L2 speakers) did not significantly affect VOT development. In this study, neither statistical analysis showed that the differing amounts of L2 input from native speakers contributed significantly to L2 VOT development among the learners in the CBI setting. Although this finding was inconsistent with some studies (e.g., Flege & Liu, 2001) that reported that L2 speech learning can occur if learners receive high-quality L2 input from native speakers, the present findings are consistent with the results of Levis, Sonsaat, Link, & Barriuso (2016), who found that students improved their comprehensibility regardless of the source (native speaker or non-native speaker) of their input. One possible explanation for the different findings could involve the timing of the first intensive exposure to the L2 in L2 speech learning. Apart from the learning contexts and methodologies used, the current study differs significantly from previous studies in the amount of L2 input the learners had received at the time of the experiment. For example, the learners in Flege and Liu (2001) had spent at least

1.7 years in the United States, which suggests that all the learners in the study had already passed the early stages of L2 learning during which experience-driven improvement is the most evident (Piske et al., 2001). The learners in the current study, in contrast, had their first intensive exposure to the L2 after entering the CBI program, which suggests that they were still in the early stages of L2 learning during the time of the experiments. That is, these differences in the timing of the experiment suggest that native speakers' input contributes significantly to additional L2 speech learning only when learners have had L2 input for more than around seven years (Trofimovich & Baker, 2006), but not to the early stages of L2 learning, as in the current study.

Another possible explanation for native-speaker input having no significant impact on learners' VOT development may be related to the different role played by native input in the acquisition of L2 phonological development in different learning conditions. Although the present findings contradict the assertion made for naturalistic settings (Flege 2009; Flege & Liu 2001), they are consistent with FL research in which native input did not have a significant impact on L2 phonological development, particularly among individuals who began learning an L2 at a relatively older age. Larson-Hall (2008) found that whereas students who began learning English between the ages of four and six tended to perform better on phonetic discrimination tasks if they had received input from native-speaking teachers, this effect diminished after the age of eight. Likewise, in a study examining the effects of age on perceptual ability in Japanese learners of English, Harada (2013) found that individuals who started learning English at the age of five were unaffected by input source (i.e., native or non-native English-speaking teachers). The difference between older and younger learners is hypothesized as being due to the robustness of an L1 phonetic system when they begin learning the L2 (Best & Tyler, 2007; Flege, 1995). The earlier learners start learning the L2, the more sensitive they may be to the phonetic difference between L1 and L2 sounds, even with relatively restricted exposure to the L2. In contrast, because learners at a relatively older age have robustly established the L1 phonetic system when they start learning the L2, the robust L1 phonetic may make it difficult for the learners to become attuned to the phonetic information that might be relevant to the L2 but not to the L1. However, being exposed to L2 input for a longer time can eventually enable these older learners to be attuned to the information, which promotes L2 pronunciation development (Trofimovich & Baker, 2006). Taken together, these findings suggest that in FL settings where native L2 input is restricted to classrooms, how long the students receive L2 input (i.e., more than

300 hours) rather than from whom they receive L2 input may be more important for L2 pronunciation development, particularly in students who began learning an L2 from age five to eight onward. Given that the students recruited in the current study were those who began learning English at the age of 12–13, when they entered secondary school, it may be reasonable to assume that input source (i.e., the total amount of L2 input from native speakers) did not significantly affect their VOT development over the course of the academic year.

Given this study's exploratory nature, further research is necessary to reach any definitive conclusions with respect to the effectiveness of L2 input in CBI settings on L2 speech learning. First, owing to the considerable variation in how university CBI programs are implemented, attempts should be made to replicate the present study's results. Second, it must be emphasized that the findings were based exclusively on a single linguistic feature. As a number of learners with extensive L2 experience in naturalistic settings struggle to accurately produce linguistic features such as vowels (e.g., Munro & Derwing, 2008) and consonants (e.g., Larson-Hall, 2008), future research should explore a wide range of features for a better understanding of the overall development of L2 pronunciation in university CBI programs. Third, the current study focused only on the quality and quantity of the input that the learners in the CBI program received as a potential factor accounting for their (potentially) different learning trajectories. There are, however, other factors that could affect L2 learners' speech learning. One of the factors that has garnered a good deal of attention in recent L2 phonological research is vocabulary size. Bundgaard-Nielsen and her colleagues reported that L2 vocabulary size can play a crucial role in restructuring the L2 phonological system (Bundgaard-Nielsen, Best, Kroos, & Tyler, 2012; Bundgaard-Nielsen, Best, & Tyler, 2011a, 2011b). To obtain a better understanding of L2 phonological development in CBI contexts, future studies should explore the role of vocabulary size.

This study adopted a longitudinal approach to examine the development of L2 VOT by Japanese students in a university-level CBI program. It shows that just as in naturalistic settings, CBI university students approximated their L2 VOT values to more target-like VOT values over the course of one academic year. However, in contrast to naturalistic settings, such improvement in a CBI setting was not influenced by the source of the input (i.e., whether it was from native or non-native speakers of the target language). Therefore, for those who start learning the L2 after puberty in FL settings (including CBI contexts), the amount of L2 input may be more important than the source of that input for promoting their L2 speech learning.

Correspondence should be addressed to **Dr. Keiko Hanzawa**, School of International Liberal Studies, Waseda University, 1-6-1 Nishi Waseda, Shinjuku, Tokyo, Japan 169-8050; e-mail: keiko_hotaru@fuji.waseda.jp.

Notes

1 It should be noted that not all languages across the world have a two-way contrast between voiced and voiceless stops; some language such as Thai and Korean have a three-way contrast (voiced, voiceless unaspirated, and voiceless stops; Lisker & Abramson, 1964) and others do no have a voicing contrast such as Fijian and Maori (Maddieson, 2013).

2 As an anonymous reviewer of this article pointed out, VOT for English stops has been demonstrated to vary systematically with gender – that is, females are likely to produce English stops with longer VOT values (Ryalls, Zipprer, & Baldauff, 1997). To that end, the learners' VOT values at T1, T2, and T3 were subjected to a repeated-measures ANOVA, but no significant effects were found either for gender, $F(1, 48) = 0.352$, $p = 0.556$, or for interaction with time, place of articulation, and gender, $F(4, 192) = 0.774$, $p = 0.543$. A repeated-measures ANOVA examining gender difference among native English speakers also showed no significant effects for gender, $F(1, 8) = 0.001$, $p = 0.970$, or gender and place of articulation, $F(2, 16) = 0.515$, $p = 0.607$. Since the finding indicated that gender was not likely to be a significant factor to influence VOT values (Morris, McCrea & Herring, 2008), the following statistical analyses were conducted by collapsing across gender.

References

Airey, J. (2016). EAP, EMI or CLIL? In K. Hyland & P. Shaw (Eds.), *The Routledge handbook of English for Academic Purposes* (pp. 71–83). New York, NY: Routledge.

Baker, W., & Trofimovich, P. (2005). Interaction of native- and second-language vowel system(s) in early and late bilinguals. *Language and Speech*, *48*(1), 1–27. https://doi.org/10.1177/00238309050480010101

Baker, W., & Trofimovich, P. (2006). Perceptual paths to accurate production of L2 vowels: The role of individual differences. *International Review of Applied Linguistics in Language Teaching*, *44*(3), 231–250. https://doi.org/10.1515/iral.2006.010

Best, C.T., & Tyler, M.D. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In O.-S. Bohn & M.J. Munro (Eds.), *Language experience in second language speech learning: In honor of James Emil Flege* (pp. 13–34). Amsterdam, Netherlands: John Benjamins.

Birdsong, D. (2007). Nativelike pronunciation among late learners of French as a second language. In O.-S. Bohn & M.J. Munro (Eds.), *Language experience in second language speech learning: In honor of James Emil Flege* (pp. 99–116). Amsterdam, Netherlands: John Benjamins.

Boersma, P., & Weenik, D. (2011). Praat: Doing phonetics by computer. Retrieved from http://www.praat.org

Bohn, O.-S., & Bundgaard-Nielsen, R.L. (2009). Second language speech learning with diverse inputs. In T. Piske & M. Young-Scholten (Eds.), *Input matters in SLA* (pp. 207–218). Bristol, England: Multilingual Matters.

Bongaert, T. (1999). Ultimate attainment in L2 pronunciation: The case of very advanced late L2 learners. In D. Birdsong (Ed.), *Second language acquisition and the critical period hypothesis* (pp. 133–159). Mahwah, NJ: Lawrence Erlbaum Associates.

Brinton, D.M., Snow, M.A., & Wesche, M.B. (2003). *Content-based second language instruction*. Ann Arbor, MI: University of Michigan Press.

Bundgaard-Nielsen, R.L., Best, C.T., Kroos, C., & Tyler, M.D. (2012). Second language learners' vocabulary expansion is associated with improved second language vowel intelligibility. *Applied Psycholinguistics*, *33*(3), 643–664. https://doi.org/10.1017/S0142716411000518

Bundgaard-Nielsen, R.L., Best, C.T., & Tyler, M.D. (2011a). Vocabulary size is associated with second-language vowel perception performance in adult learners. *Studies in Second Language Acquisition*, *33*(3), 433–461. https://doi.org/10.1017/s0272263111000040

Bundgaard-Nielsen, R.L., Best, C.T., & Tyler, M.D. (2011b). Vocabulary size matters: The assimilation of second-language Australian English vowels to first-language Japanese vowel categories. *Applied Psycholinguistics*, *32*(1), 51–67. https://doi.org/10.1017/s0142716410000287

Cebrian, J. (2006). Experience and the use of non-native duration in L2 vowel categorization. *Journal of Phonetics*, *34*(3), 372–387. https://doi.org/10.1016/j.wocn.2005.08.003

Dalton-Puffer, C. (2011). Content-and-language integrated learning: From practice to principles? *Annual Review of Applied Linguistics*, *31*, 182–204. https://doi.org/10.1017/s0267190511000092

Flege, J.E. (1987). The production of "new" and "similar" phones in a foreign language: Evidence for the effect of equivalence classification. *Journal of Phonetics*, *15*, 47–65.

Flege, J.E. (1988). Factors affecting degree of perceived foreign accent in English sentence. *Journal of the Acoustical Society of America*, *84*, 70–79. https://doi.org/10.1121/1.396876

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*, 395–411. https://doi.org/10.1121/1.400473

Flege, J.E. (1995). Second language speech learning: Theory, findings, and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 233–277). Timonium, MD: York Press.

Flege, J.E. (2009). Give input a chance! In T. Piske & M. Young-Scholten (Eds.), *Input matters in SLA* (pp. 175–190). Bristol, England: Multilingual Matters.

Flege, J.E., & Fletcher, L. (1992). Talker and listener effects on degree of perceived foreign accent. *Journal of the Acoustical Society of America*, *91*, 370–389. https://doi.org/10.1121/1.402780

Flege, J.E., & Liu, S. (2001). The effect of experience on adults' acquisition of a second language. *Studies in Second Language Acquisition*, *23*(4), 527–552.

Flege, J.E., Munro, M.J., & MacKay, I.R.A. (1995). Effects of age of second-language learning on the production of English consonants. *Speech Communication*, *16*(1), 1–26. https://doi.org/10.1016/0167-6393(94)00044-B

Harada, T. (2006). The acquisition of single and geminate stops by English-speaking children in a Japanese immersion program. *Studies in Second Language Acquisition*, *28*(4), 601–632. https://doi.org/10.1017/s0272263106060281

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*, *45*(4). https://doi.org/10.1515/iral.2007.015

Harada, T. (2013). Effects of early language learning on speech perception: From an ELF perspective. Waseda working papers in ELF, *2*, 111–122.

Jia, G., & Aaronson, D. (2003). A longitudinal study of Chinese children and adolescents learning English in the United States. *Applied Psycholinguistics*, *24*(1), 131–161. https://doi.org/10.1017/s0142716403000079

Johnson, C.E., & Wilson, I.L. (2002). Phonetic evidence for early language differentiation: Research issues and some preliminary data. *International Journal of Bilingualism*, *6*(3), 271–289. https://doi.org/10.1177/13670069020060030401

Kang, K.-H., & Guion, S.G. (2006). Phonological systems in bilinguals: Age of learning effects on the stop consonant systems of Korean-English bilinguals. *Journal of the Acoustical Society of America*, *119*(3), 1672–1683. https://doi.org/10.1121/1.2166607

Keating, P.A. (1984). Phonetic and phonological representation of stop consonant voicing. *Language*, *60*(2), 286–319. https://doi.org/10.2307/413642

Kessinger, R.H., & Blumstein, S.E. (1997). Effects of speaking rate on voice-onset time in Thai, French, and English. *Journal of Phonetics*, *25*(2), 143–168. https://doi.org/10.1006/jpho.1996.0039

Larson-Hall, J. (2008). Weighing the benefits of studying a foreign language at a younger starting age in a minimal input situation. *Second Language Research*, *24*(1), 35–63. https://doi.org/10.1177/0267658307082981

Levis, J.M., Sonsaat, S., Link, S., & Barriuso, T.A. (2016). Native and nonnative teachers of L2 pronunciation: Effects on learner performance. *TESOL Quarterly*, *50*(4), 894–931. https://doi.org/10.1002/tesq.272

Lisker, L., & Abramson, A.S. (1964). A cross-language study of voicing in initial stops: Acoustical measurements. *Words*, *20*(3), 384–422. https://doi.org/10.1080/00437956.1964.11659830

Lyster, R. (2007). *Learning and teaching languages through content: A counterbalanced approach*. Amsterdam, Netherlands: John Benjamins.

Lyster, R., & Ballinger, S. (2011). Content-based language teaching: Convergent concerns across divergent contexts. *Language Teaching Research*, *15*(3), 279–288. https://doi.org/10.1177/1362168811401150

Maddieson, I. (2013). Voicing in plosives and fricatives. In M.S. Dryer & M. Haspelmath (Eds.). *The world atlas of language structures online*. Leipzig,

Germany: Max Planck Institute for Evolutionary Anthropology. Retrieved from http://wals.info/chapter/4

Marsh, D., & Frigols Martín, M.J. (2013). Content and language integrated learning. In C.A. Chapelle (Ed.), *The encyclopedia of applied linguistics* (pp. 1–10). Chichester, England: Wiley-Blackwell.

Met, M. (1998). Curriculum decision-making in content-based language teaching. In J. Cenoz & F. Genesee (Eds.), *Beyond bilingualism: Multilingualism and multilingual education* (pp. 35–63). Clevedon, England: Multilingual Matters.

Moriyoshi, N. (2010). *Content-based instruction (CBI) in Japanese college classrooms: Focusing on language, content, or both?* (Masters thesis). Retrieved from http://digitool.Library.McGill.CA:80/R/-?func = dbin-jump-full&object_id = 97218&silo_library = GEN01

Morris, R.J., McCrea, C.R., & Herring, K.D. (2008). Voice onset time differences between adult males and females: Isolated syllables. *Journal of Phonetics*, *36*(2), 308–317. https://doi.org/10.1016/j.wocn.2007.06.003

Moyer, A. (2004). Accounting for context and experience in German (L2) language acquisition: A critical review of the research. *Journal of Multilingual and Multicultural Development*, *25*(1), 41–61. https://doi.org/10.1080/01434630408666519

Moyer, A. (2009). Input as a critical means to an end: Quantity and quality of experience in L2 phonological attainment. In T. Piske & M. Young-Scholten (Eds.), *Input matters in SLA* (pp. 159–174). Bristol, England: Multilingual Matters.

Moyer, A. (2011). An investigation of experience in L2 phonology: Does quality matter more than quantity? *The Canadian Modern Language Review*, *67*(2), 191–216. https://doi.org/10.3138/cmlr.67.2.191

Munro, M.J., & Derwing, T.M. (2008). Segmental acquisition in adult ESL learners: A longitudinal study of vowel production. *Language Learning*, *58*(3), 479–502. https://doi.org/10.1111/j.1467-9922.2008.00448.x

Nearey, T.M., & Rochet, B.L. (1994). Effects of place of articulation and vowel context on VOT production and perception for French and English stops. *Journal of the International Phonetic Association*, *24*(1), 1–18. https://doi.org/10.1017/s0025100300004965

Netelenbos, N., Li, F., & Rosen, N. (2015). Stop consonant production of French immersion students in Western Canada: A study of voice onset time. *International Journal of Bilingualism*, *20*(3), 346–357. https://doi.org/10.1177/1367006914564566

Ortega, L., & Iberri-Shea, G. (2005). Longitudinal research in second language acquisition: Recent trends and future directions. *Annual Review of Applied Linguistics*, *25*, 26–45. https://doi.org/10.1017/s0267190505000024

Piske, T. (2007). Implications of James E. Flege's research for the foreign language classroom. In O.-S. Bohn & M.J. Munro (Eds.), *Language experience in second language speech learning: In honor of James Emil Flege* (pp. 301–314). Amsterdam, Netherlands: John Benjamins.

Piske, T., Flege, J.E., MacKay, I.R.A., & Meador, D. (2002). The production of English vowels by fluent early and late Italian-English bilinguals. *Phonetica*, *59*(1), 49–71. https://doi.org/10.1159/000056205

Piske, T., MacKay, I.R.A., & Flege, J.E. (2001). Factors affecting degree of foreign accent in an L2: A review. *Journal of Phonetics*, *29*(2), 191–215. https://doi.org/10.1006/jpho.2001.0134

Piske, T., & Young-Scholten, M. (2009). Introduction. In T. Piske & M. Young-Scholten (Eds.), *Input matters in SLA* (pp. 1–26). Bristol, England: Multilingual Matters.

Riney, T.J., & Takagi, N. (1999). Global foreign accent and voice onset time among Japanese EFL speakers. *Language Learning*, *49*(2), 275–302. https://doi.org/10.1111/0023-8333.00089

Riney, T.J., Takagi, N., Ota, K., & Uchida, Y. (2007). The intermediate degree of VOT in Japanese initial voiceless stops. *Journal of Phonetics*, *35*(3), 439–443. https://doi.org/10.1016/j.wocn.2006.01.002

Ryalls, J., Zipprer, A., & Baldauff, P. (1997). A preliminary investigation of the effects of gender and race on voice onset time. *Journal of Speech, Language, and Hearing Research*, *40*(3), 642–645. https://doi.org/10.1044/jslhr.4003.642

Saito, K., & Munro, M.J. (2014). The early phase of /ɹ/ production development in adult Japanese learners of English. *Language and Speech*, *57*(4), 451–469. https://doi.org/10.1177/0023830913513206

Simon, E., & D'Hulster, T. (2012). The effect of experience on the acquisition of a non-native vowel contrast. *Language Sciences*, *34*(3), 269–283. https://doi.org/10.1016/j.langsci.2011.10.002

Stölten, K., Abrahamsson, N., & Hyltenstam, K. (2015). Effects of age and speaking rate on voice onset time: The production of voiceless stops by near-native L2 speakers. *Studies in Second Language Acquisition*, *37*(1), 71–100. https://doi.org/10.1017/s0272263114000151

Sundberg, U., & Lacerda, F. (1999). Voice onset time in speech to infants and adults. *Phonetica*, *56*(3–4), 186–199. https://doi.org/10.1159/000028450

Trofimovich, P., & Baker, W. (2006). Learning second language suprasegmentals: Effect of L2 experience on prosody and fluency characteristics of L2 speech. *Studies in Second Language Acquisition*, *28*(1), 1–30. https://doi.org/10.1017/s0272263106060013

Unterberger, B. (2012). English-medium programmes at Austrian business faculties: A status quo survey on national trends and a case study on programme design and delivery. *AILA Review*, 25, 80–100. https://doi.org/10.1075/aila.25.06unt

Yaneş, M., & Wildermuth, R. (2006). The effects of place of articulation and vowel height in the acquisition of English aspirated stops by Spanish speakers. *International Review of Applied Linguistics*, *44*(3), 251–263. https://doi.org/10.1515/IRAL.2006.011

Yang, W. (2014). Content and language integrated learning next in Asia: Evidence of learners' achievement in CLIL education from a Taiwan tertiary degree programme. *International Journal of Bilingual Education and Bilingualism*, *18*(4), 361–382. https://doi.org/10.1080/13670050.2014.904840

Appendix: Mean VOT values for English voiceless stops produced by CBI and non-CBI at T1, T2, and T3 (in milliseconds)

Mean VOT values	*M (SD)*		
	T1	**T2**	**T3**
CBI students			
/p/	52.1 (13.8)	56.0 (16.8)	61.6 (17.7)
/t/	56.9 (14.2)	61.0 (15.3)	63.9 (14.9)
/k/	68.4 (12.7)	74.3 (14.7)	80.0 (14.3)
Non-CBI students			
/p/	45.7 (16.8)	42.1 (17.4)	44.6 (15.63)
/t/	50.4 (19.9)	47.8 (17.8)	46.9 (15.4)
/k/	65.2 (18.2)	62.6 (16.4)	64.7 (15.5)

Note: CBI = content-based instruction; T1 = time at the beginning of the first semester; T2 = time at the end of the first semester; T3 = time at the end of the second semester; VOT = voice onset time.