Perception of allophonic cues to English word boundaries by Japanese second language learners of English

Kikuyo Ito^{a)} and Winifred Strange

Ph.D. Program in Speech-Language-Hearing Sciences, City University of New York—Graduate School and University Center, 365 Fifth Avenue, New York, New York 10016-4309

(Received 30 April 2008; revised 10 December 2008; accepted 23 January 2009)

Perception of stop aspiration and glottal stop allophonic cues for word juncture in English by Japanese second language (L2) learners of English was examined, extending a study of Spanish L2 learners [Altenberg, E. P. (2005). *Second Lang. Res.* **21**, 325–358]. Thirty Japanese listeners ranging in length of residence (LOR) in the United States (2 weeks–12 years) were tested on 42 contrasting pairs (e.g., aspiration: *keeps talking* vs *keep stalking*, glottal stop: *a nice man* vs *an ice man*, and double cues: *grape in* vs *grey pin*). Phrases were presented in randomly ordered lists and subjects responded in a two-choice identification task followed by a phrase familiarity test. The Japanese listeners performed more poorly than an American English-speaking control group ($N = 10$), especially on aspiration pairs. Aspiration pairs were differentiated significantly less well (73% correct) by Japanese listeners than were glottal stop pairs (91% correct) and double cue pairs (94% correct); response biases predicted from relative familiarity of phrases were evident only for aspiration pairs. Performance correlated with LOR and suggested that aspiration cues take more immersion experience to learn than glottal stop cues. The patterns of errors were similar, but not identical, to Altenberg's Spanish data.

© 2009 Acoustical Society of America. [DOI: 10.1121/1.3082103]

PACS number(s): 43.71.Hw, 43.71.Es [PEI]

Pages: 2348–2360

I. INTRODUCTION

When processing running speech, listeners have to segment continuous streams of sound into recognizable units such as words and phrases. Identifying word/phrase boundaries is a necessary stage for auditory language processing because there are no reliable pauses between words in continuous speech. The process of word segmentation in speech seems to be so automatic and effortless for first language (L1) listeners that, more often than not, they are not even aware of doing it. On the other hand, everybody has experience of not being able to figure out where the words start and end when listening to unfamiliar foreign speech, and even with learning experience, correctly segmenting a stream of non-native speech into words and phrases often requires extensive exposure to the language and a considerable amount of practice.

Various kinds of information may be involved in the process of word segmentation: lexical/semantic cues (e.g., Gow and Gordon, 1995), syntactic/morphological cues (e.g., Sanders and Neville, 2000), phonotactic cues (e.g., McQueen, 1998), prosodic/rhythmic cues (e.g., Cutler and Norris, 1988), and allophonic acoustic-phonetic cues (e.g., Nakatani and Dukes, 1977). Native listeners are likely to utilize all these multiple cues available in natural speech for efficient word segmentation, and several cue-trading segmentation studies by native English listeners have been documented. For example, Gow and Gordon (1995) examined the interaction of acoustic-phonetic and lexical informa-

tion on word segmentation and found that lexical contexts induced simultaneous access to words other than those intended by speakers (e.g., accessing *tulips* in addition to *two lips* when listening to *two lips*). Mattys *et al.* (2005) investigated how word stress cues interacted with conflicting other cues and found that stress information became a strong segmentation cue when acoustic-phonetic, phonotactic, or lexical information was not available. Their results also indicated that acoustic-phonetic and phonotactic information were secondary to lexical knowledge, suggesting that higher-order information outweighs sublexical cues. Mattys and Melhorn (2007) also examined the effects of sentential context, lexical knowledge, and acoustic-phonetic cues in connected speech and found compensatory mechanisms between them: The more ambiguous the acoustic-phonetic cues, the larger the effect of both sentential and lexical context on segmentation when listening to near-homophonous phrases (e.g., /plʌmpaɪ/ for *plum pie* or *plump eye*). They also found that lexical effects reduced sensitivity to the acoustic-phonetic cues (e.g., favoring *scum pie* over **scump eye* even when it was acoustically more like **scump eye*), highlighting the graded, interactive, and flexible nature of multi-cue segmentation. The present study examined how non-native listeners process acoustic-phonetic segmentation cues present in L2 speech to perceive word boundaries when other types of cues are not available.

In a series of studies on the use of multiple segmentation cues, Sanders and co-workers compared strategies in segmenting English input used by monolingual native speakers of English (Sanders and Neville, 2000, 2003a) and by late L2 learners of English (Sanders *et al.*, 2002; Sanders and Neville, 2003b). The studies used behavioral as well as event-

^{a)}Electronic mail: kikuyoito@hotmail.com

related potential (ERP) measurements and focused on the interaction of lexical, syntactic, and stress-pattern cues. The behavioral results for English listeners confirmed that they exploited these cues in combination (Sanders and Neville, 2000) whereas native Japanese and native Spanish late learners did not use syntactic information to the same extent as native listeners did, although they exploited lexical information to a similar extent (Sanders *et al.*, 2002). The results also indicated that Japanese late learners as well as Japanese near monolinguals used the typical English stress cues for segmentation better than Spanish late learners did, suggesting that the extent to which a cue was used depended on the characteristics of their native languages. In the ERP experiments, native English listeners showed strong word-onset effects, suggesting that neither semantic nor syntactic information is necessary for word-onset identification (Sanders and Neville, 2003a). Japanese late learners' ERP data, on the other hand, only showed semantic processing effects without showing the word-onset effects nor the syntactic processing effects (Sanders and Neville, 2003b). These results indicated that non-native listeners were not processing the acoustic differences in the same way as native listeners.

While the research on the use of multiple segmentation cues has increasingly drawn attention recently, studies on individual segmentation cues have not been as well documented to date. The bulk of segmentation research dealing with individual cues has concentrated on examining the use of rhythmic information, especially in the area of cross-language research. (e.g., Mehler *et al.*, 1981; Cutler and Norris, 1988; Cutler and Butterfield, 1992; Otake *et al.*, 1993; Cutler *et al.*, 1986, 1992; Cutler and Otake, 1994; Bradley *et al.*, 1993; Goettry and Kolinsky, 2000). Although the research on rhythmic cues has made a significant contribution to our understanding of native and non-native word segmentation, studies on other individual cues have not been well explored yet. The present study contributes by investigating the use of allophonic acoustic-phonetic segmentation cues (*allophonic cues*, hereafter) for word juncture in American English (AE, hereafter): (a) presence/absence of aspirated voiceless stop consonants (aspirated in word-initial stops) and (b) presence/absence of glottal stops as cues to word onset (present before word-initial vowels).¹

In studies on the use of allophonic segmentation cues by native listeners, researchers have found that the occurrence of allophones of the same phoneme can signal segmentation. Christie (1974) examined English listeners' perception of syllable boundaries by using synthetic stimuli and found that the aspiration of /t/ affected the perception of the boundary location whereas formant transitions had little or no effect on the boundary. Nakatani and Dukes (1977) examined English listeners' use of allophonic variations as segmentation cues, using spliced speech of near-homophonous phrases such as *no notion/known ocean* and *play taught/plate ought*. The authors found that glottal stops or laryngealization (creaky voice) at the onset of word-initial vowels, aspiration of voiceless stops, and syllable-initial vs -final /l/ and /r/ allophones were strong segmentation cues for AE listeners. They also found that duration differences of words caused by presence/absence of voiceless word-final consonants (e.g.,

plate in *plate ought* being longer than *play* in *play taught*) were not useful as a segmentation cue for native listeners. Church (1987) also argued from a computational linguistic viewpoint that the information provided by allophonic variations is useful for segmentation. He illustrated that the systematic distribution of allophonic variations, such as word-initial aspiration of stops, stop release, flapping, and palatalization in English, facilitates parsing speech segments by determining the location of syllables and stress.

Cross-language research on word segmentation dealing with allophonic cues is very limited. Altenberg (2005) investigated whether Spanish L2 learners of English could use English aspiration cues and glottal stop cues for segmentation in the same way as native listeners did, using an adapted version of the design by Nakatani and Dukes (1977). Participants were presented speech stimuli, such as *keeps talking* and *a nice man*, and were asked to indicate whether they heard *keep stalking* or *keeps talking*, and *a nice man* or *an ice man*. In these examples, the presence of aspiration in /t/ in *keeps talking*, which is absent in *stalking*, and the presence of glottal stop or creaky voice preceding *ice* in *an ice man*, which is absent in *nice*, provide strong segmentation cues for native speakers of English (Nakatani and Dukes, 1977; Altenberg, 2005). Altenberg (2005) found that Spanish listeners perceived the word boundaries significantly less accurately than English listeners did. She also found that Spanish listeners' performance on the aspiration pairs was significantly worse than that on the glottal stop pairs, while no differences were seen in the performance of native English speakers. Altenberg (2005) concluded that Spanish L2 learners of English were unable to fully use these allophonic cues available to native English listeners and suggested that various factors, including L1 transfer and markedness, may be responsible for greater ease or difficulty of utilizing particular L2 word segmentation cues. Further research is needed to see if the observed relative difficulties generalize to listeners of other languages and to explore the factors that may cause ease or difficulty in processing L2 allophonic segmentation cues.

The present study is an extension of the Altenberg study (2005) of aspiration and glottal stop segmentation cues. Japanese L2 learners of English served as experimental subjects with native English speakers as a control group. The distribution and the extent of aspiration of voiceless stops in Japanese are different from those in English and also different from those in Spanish. Japanese voiceless stops /p/, /t/, and /k/ are weakly aspirated in word-initial position but unaspirated elsewhere (Hattori, 1951; Vance, 1987). According to Homma (1980), the voice onset time (VOT) of Japanese word-initial /t/ is approximately 25 ms. Recently, Riney *et al.* (2007) measured VOT values of Japanese /p/, /t/, and /k/ produced in several consonant + vowel (CV) contexts and compared them with those of English /p/, /t/, and /k/ produced in equivalent conditions by monolingual AE speakers. The mean values of Japanese /p/, /t/, and /k/ were 30.0, 28.5, and 56.7 ms, respectively, whereas the corresponding AE VOT values were 78.2, 94.8, and 102.8 ms. As for the VOT values of Spanish initial voiceless stops, the Latin American Spanish data by Williams (1977) seem to be the closest approximation of the native VOT values for Alten-

berg's (2005) Spanish participants, who were primarily Dominicans, Mexicans, and Panamanians. In Williams' data (1977), the average VOT values of /p/, /t/, and /k/ in Guatemalan, Venezuelan, and Peruvian Spanish ranged from 9.8 to 15.2 ms, 10.3 to 20.6 ms, and 25.7 to 32.6 ms, respectively. Thus, since aspiration of voiceless stops in initial position appears to be longest in AE, shortest in Spanish, and of intermediate duration in Japanese, we can assume that allophonic differences between initial and non-initial voiceless stops are greatest for AE, least for Spanish, and intermediate for Japanese, assuming that values for noninitial stops are near zero VOT for all three languages.

Glottal stops in Japanese frequently occur in utterance-final position in emphatic speech (Bloch, 1950; Vance, 1987), and some researchers further argue that any utterance-final short vowel in Japanese is followed by a glottal stop (Kawakami, 1977; Vance, 1987). A glottal stop in Japanese generally appears before an utterance-initial vowel in emphatic speech (Vance, 1987) as well and it may appear even before a word-initial or phrase-initial vowel (Kawakami, 1977). Thus, in Japanese, glottal stops appear to assume a role as an utterance-, phrase-, or word-boundary marker. A glottal stop in Japanese can be realized as an allophonic variation of the moraic obstruent /Q/ (i.e., the first half of geminate consonants) as well. However, it cannot be confusable with the boundary-marking glottal stops because the moraic obstruent /Q/ cannot occur utterance finally nor immediately before a vowel. In Spanish, as with Japanese, a glottal stop can be inserted before a word-initial vowel, and after a word-final vowel in emphatic speech (Stockwell, *et al.*, 1956).

Thus, if ease or difficulty in using non-native segmentation cues is based on their use in the native language (i.e., L1 transfer), Japanese listeners in the present study would be expected to show better perception of aspiration cues than Spanish listeners (Altenberg, 2005), while showing similar performance on glottal stop cues. By examining the perception of Japanese listeners using the same lexical items as the Altenberg study (2005), the present study contributes new information on the use of allophonic cues for L2 word segmentation.

In this study a new native speaker of AE produced the materials, using the same instructions as were used in the original experiment. A control group of AE listeners was included to verify that the materials contained sufficient segmentation cues for natives to perform with near ceiling as did Altenberg's control group (2005). Detailed acoustic analysis was also conducted to document the multiple parameters that distinguished the stimulus pairs. Two additional testing procedures were added: (a) listeners' confidence ratings of their response choices, and (b) a follow-up familiarity rating test. The latter was used to establish whether the familiarity of members of the pairs differed systematically with respect to the presence/absence of aspiration/glottal stop cues to segmentation and to help in the interpretation of the results.²

The research questions in the present study were as follows.

- (1) Is Japanese L2 listeners' use of English allophonic cues for word segmentation poorer than that of AE native listeners? It was expected that AE listeners would perform at ceiling, whereas Japanese listeners would perform significantly worse on both kinds of segmentation cues. However, in addition to group differences, establishing AE listeners' performance levels on the same materials allowed us to compare individual Japanese participants' performance against a native-speaker level of accuracy as a function of immersion experience.
- (2) Are aspiration cues more difficult for Japanese listeners to utilize than glottal stop cues? Given the distribution of such cues in Japanese, we expected that performance on aspiration cues would be significantly poorer than performance on glottal stop cues. In addition, performance on presence vs absence of the cues was compared, given Altenberg's finding (2005) that the presence of aspiration and glottal stop cues provided more effective information than their absence.
- (3) Does Japanese listeners' use of English allophonic cues improve with more experience in an English-speaking environment? Correlational analyses of immersion experience and performance on each type of segmentation cues were also employed to explore whether some cues take more immersion experience to learn than others.

II. METHODS

A. Participants

Thirty adult native speakers of Japanese living in the United States and a control group of ten adult native speakers of AE participated in the experiment. All participants passed a hearing screening (ANSI standards 25 dB HL at 500, 1 000, 2 000, and 4 000 Hz). Information about the Japanese participants, such as age, gender, length of residence (LOR) in English-speaking countries, which includes stays in the United States, Canada, United Kingdom, or Australia,³ age of arrival (AOA) in English-speaking countries, and educational experience in English-speaking countries (immersion experience), was collected by means of a questionnaire. Since none of the Japanese participants came to English-speaking countries earlier than 15 years of age (AOA: 15–45 years of age), all Japanese participants were considered late learners of English. The information for the AE participants was also collected via a questionnaire. All AE participants had had experience in learning at least one of the following languages at school: Spanish, French, Italian, or Latin. None of them, however, had learned Japanese or any Asian languages in their lives. The age range and gender proportion of the AE group were matched to those of the Japanese group as closely as possible. Table I shows the information about the Japanese participants and the AE participants.

B. Stimulus materials

Eighty-four naturally produced stimuli that constitute forty-two stimulus pairs were recorded. The stimuli were identical to those used in the Altenberg (2005) study but were produced by a different speaker: a phonetically trained

TABLE I. Biographical information for AE participants and Japanese participants (age, gender, LOR, AOA, and immersion experience).

	American participants	Japanese participants
Sample size	10 (9 females, 1 male)	30 (26 females, 4 males)
Mean age (range)	31.2 years (26–38 yrs)	30.7 years (20–46 years)
Mean LOR ^a (range)		4.0 years (2 weeks–12 years)
Mean AOA ^b (range)		25.0 years (15–45 years)
Mean immersion experience ^c (range)		2.5 years (0–10 years)

^aLOR=length of residence in English-speaking countries.

^bAOA=age of arrival in English-speaking countries.

^cImmersion experience=educational experience in English-speaking countries (includes experience at language school).

female native speaker of AE with a New York dialect. The stimuli were produced in the carrier phrase “Say__again.” Following the procedure of the [Altenberg study \(2005\)](#), the speaker was instructed to speak naturally but at a slightly slower rate than usual to minimize the likelihood of listeners’ performance errors (p. 336). [Altenberg \(2005\)](#) noted that care was taken to minimize the influence of lexical familiarity. The stimuli were selected from a larger set of 168 phrases (84 pairs) based on naturalness ratings by 7 native speakers of English “such that the naturalness rating of one member of the set (e.g., *keep stalking*) would be as close as possible to the naturalness rating of the other member of the set (e.g., *keeps talking*)” (p. 336). All stimuli were digitally recorded in a sound-attenuated room at 22.050 kHz, monaural, with 16 bit resolution, using a microphone (SHURE SM 48) fed through a preamplifier (Earthworks LAB 101) into the computer analog to digital sound card using SOUND FORGE 4.5 software.

The stimuli are subdivided into three types: 18 Aspiration stimulus pairs (36 phrases) in which presence or absence of aspiration of a voiceless-stop at a word boundary signals the boundary, 18 Glottal stop stimulus pairs (36 phrases) in which presence or absence of glottal stop or creaky voice signals the boundary, and 6 Double cue stimulus pairs (12 phrases) in which both presence or absence of aspiration and of glottal stop can be used as cues for segmentation.

Aspiration stimuli are further broken down into three groups: the stimuli that have a vowel before the /s/ at the juncture followed by a voiceless stop (VsC group; e.g., *loose tops/Lou stops*), the stimuli that have a consonant before the /s/ followed by a voiceless stop (CsC group; e.g., *keeps talking/keep stalking*), and the stimuli that have a consonant before the /s/ followed by a voiceless stop plus another consonant (CsCC group; e.g., *cook’s prints/cook sprints*). Glottal stop stimuli are also broken down into three groups: the stimuli that have a nasal consonant at the juncture (Nasal group; e.g., *an itch/a niche*⁴), the stimuli that have an obstruent at the juncture (Obstruent group; e.g., *wife ill/why fill*), and the stimuli that have a liquid at the juncture (Liquid group; e.g., *I’ll earn/I learn*). The stimuli with an aspiration cue are referred to as *Positive Aspiration stimuli* and those with a glottal stop cue as *Positive Glottal stop stimuli*. The stimuli without aspiration and those without glottal stop are referred to as *Negative Aspiration stimuli* and *Negative Glottal stop stimuli*, respectively. Double cue stimuli do not have

subcategories. The items of each pair are either one of the following two types: *Aspiration-negative/Glottal-positive* (hereafter *Asp-/Glott+*) stimuli, which contain the glottal stop cue but no aspiration cue (e.g., *grape in*); or *Aspiration-positive/Glottal-negative* (hereafter *Asp+/Glott-*) stimuli, which contain the aspiration cue but no glottal stop cue (e.g., *grey pin*).

In addition to the 42 stimulus pairs, 6 extra stimulus pairs (12 stimuli) for familiarization trials were recorded by the same speaker. These stimulus pairs were taken from those used in [Nakatani and Dukes, 1977](#), except for one stimulus pair (*ice cream/I scream*). All stimulus pairs, including those used in the familiarization trials, are presented in the Appendix.

C. Acoustic analysis

An acoustic analysis was carried out for all 84 items to detect any consistent acoustic cues that signal the word juncture in the stimuli, using the computer software MULTISPEECH. The critical area of the target phrase, including pre-boundary and post-boundary segments, was selected and displayed on the screen in waveforms and wideband spectrograms that were temporally aligned. The selected areas contained four to five phonemes (e.g., /ɪpstɔ/ in *keeps talking* and *keep stalking*). The analysis size of the spectrogram was set at 100 points (323.00 Hz) with a 0–5000 Hz display in Hamming window. The durations of distinct acoustic segments, such as silence period, frication, VOT, stop burst, and nasal murmur, were measured by placing cursors in reference to both waveform and spectrogram. Where successive segments were difficult to divide further, as in a sequence of an approximant and a vowel, it was treated as one unit to avoid inaccurate segmentation.

1. Aspiration stimuli

The left panel of Fig. 1 shows the durations of acoustic segments observed in Aspiration stimuli that were averaged across the subcategorized groups with positive and negative distinctions shown separately. The primary aspiration cue was quantified by measuring the duration of VOT, which was seen as noise between the onset of the stop burst and the first striation of the following vowel in the spectrogram. As predicted, the VOTs were distinctively longer in Positive (range 40–114 ms) stimuli than in Negative (range=0–53 ms)

FIG. 1. Mean durations of acoustic segments: Aspiration stimuli (left panel) and Glottal stop stimuli (right panel). *In the positive stimuli of the Liquid group, the liquid segment was not distinguishable from the preceding segment. For this subgroup, preceding segment includes both the preceding segment as well as the liquid segment.

stimuli in all three subcategories, and the positive items were always longer than the negative counterparts (differences ranging from 22 to 95 ms). This difference was highly significant by a paired sample t -test ($t[17]=12.17$, $p<0.001$). In addition, the /s/ frication durations were noticeably shorter in the Positive than in the Negative stimuli ($t[17]=5.92$, $p<0.001$) and the preceding phonemes were slightly shorter in Positive stimuli than in Negative stimuli in all three subcategories ($t[17]=2.46$, $0.01<p<0.05$).

2. Glottal stop stimuli

The right panel of Fig. 1 shows the mean durations of acoustic segments in Glottal stop stimuli categorized in the same way as those in Aspiration stimuli. The most distinctive acoustic cue of Glottal stop stimuli, identified as glottal stop or creaky voice, was characterized by the silent gap or aperiodic sound preceding the word-initial vowel. None of the negative stimuli contained these cues. In Positive stimuli, there were two tokens, *an ice man* and *save amps*, that did not contain the glottal stop/creaky voice. They sounded potentially confusable with the negative counterparts, *a nice man* and *say vamps*, to the first author who was a native speaker of Japanese. However, two native speakers of AE judged them as the correct phrases and as sounding natural. Thus, these two tokens were kept as stimuli to see if native listeners were still able to correctly perceive them. The performance on these two tokens by the Japanese group and by the AE group is discussed in Sec. III B. Durational differences between positive and negative differences, including the two aberrant tokens, were statistically significant ($t[17]=8.60$, $p<0.001$).

It was also observed that the nasal murmurs in the Nasal group and the /f/ or /v/ frication in the Obstruent group were significantly shorter in Positive stimuli than in Negative stimuli ($t[11]=3.15$, $p<0.01$).⁵ Furthermore, in the Liquid group, there was an obvious difference between Positive and Negative stimuli. While the liquid portion and the preceding vocalic part were distinguishable in spectrograms in negative stimuli (e.g., /l/ in *I learn*), they were not separable in the positive group (e.g., /l/ in *I'll earn*), due to their pronunciation as postvocalic allophones that are continuous with preceding vowel segments (vocalic [ɤ̣] and velarized [ɤ̣] in the

above example). As a result, the “preceding segment” of positive stimuli in the Liquid group in the right panel of Fig. 1 includes the liquid part as well.

3. Double cue stimuli

As mentioned earlier, in Double cue stimuli, one type of item is characterized by the absence of aspiration and presence of glottal stop (*Asp-/Glott+* stimuli), and the other by the presence of aspiration and absence of glottal stop (*Asp+/Glott-* stimuli). The acoustic analysis not only confirmed these acoustic characteristics but also suggested additional acoustic cues for word segmentation.

In *Asp-/Glott+* stimuli, positive VOTs were not seen in either of the stimuli, and all stimuli contained silence or random noise that indicated glottal stop or creaky voice before the word-initial vowel with one exception, *light old*. In this particular token, there was a 108 ms of silence before the vowel, but the silence originated as a stop occlusion so it was not possible to judge whether a glottal stop was included in the silence or not. In addition to these expected acoustic cues, in *Asp-/Glott+* stimuli with a bilabial or velar stop (i.e., *grape in*, *weep at*, *make art*, *like old*), 38–46 ms stop bursts were observed between the stop occlusion and the glottal stop/creaky voice, which can be a strong word juncture cue. In *Asp+/Glott-* stimuli, 105–122 ms VOTs, which were totally absent in *Asp-/Glott+* stimuli, were observed, and no signs of glottal stop/creaky voice were seen. Additional possible juncture cues included shorter stop silences in *Asp-/Glott+* stimuli (50–113 ms) than in *Asp-/Glott+* stimuli (83–129 ms), except for the *light old/lie told* pair. Moreover, the segments preceding the stop occlusion in *Asp-/Glott+* stimuli (117–339 ms) were always shorter than those in corresponding *Asp-/Glott+* stimuli (152–410 ms).

4. Summary

Table II summarizes the acoustic cues confirmed by the acoustic analysis. The acoustic analysis revealed that the difference in VOT values was present as the most distinctive acoustic cue in Aspiration stimuli and further indicated that the duration of the /s/ frication, and possibly that of the preceding phoneme, may also signal the word juncture. For

TABLE II. Summary of acoustic cues for word juncture.

Stimulus type	Main cues	Secondary cues	Other cues
Aspiration stimuli			
Positive stimuli (e.g., <i>loose tops</i>)	Longer VOT	Shorter /s/ friction	Shorter preceding vocalic segment
Negative stimuli (e.g., <i>Lou stops</i>)	Shorter VOT	Longer /s/ friction	Longer preceding vocalic segment
Glottal stop stimuli			
Positive stimuli (e.g., <i>an itch</i> , <i>wife ill</i> , <i>seal of</i>)	Presence of glottal stop	<ul style="list-style-type: none"> • Shorter nasal murmur or /f/, /v/ frication • Post-vocalic /l/ or /r/ 	
Negative stimuli (e.g., <i>a niche</i> , <i>why fill</i> , <i>sea love</i>)	Absence of glottal stop	<ul style="list-style-type: none"> • Longer nasal murmur or /f/, /v/ frication • Pre-vocalic /l/ or /r/ 	
Double cue stimuli			
Asp−/Glott+ (e.g., <i>grape in</i>)	<ul style="list-style-type: none"> • Absence of +VOT • Presence of glottal stop 	Stop burst of /p/ or /k/ (but not /t/)	<ul style="list-style-type: none"> • Shorter stop silence • Shorter preceding segment
Asp+/Glott− (e.g., <i>grey pin</i>)	<ul style="list-style-type: none"> • Presence of +VOT • Absence of glottal stop 		<ul style="list-style-type: none"> • Longer stop silence • Longer preceding segment

Glottal stop stimuli, the clear difference of presence or absence of the glottal stop/creaky voice was the most reliable acoustic cue. The duration of the nasal murmur, or the /f/ or /v/ frication, was an additional cue for the word juncture. In the case of the Liquid group, the allophonic change of /r/ or /l/ suggested a strong cue for word juncture. For Double cue stimuli, the combination of the presence or absence of long VOT and glottal stop/creaky voice indicated strong juncture cues. In addition, the acoustic analysis suggested that the stop burst of /p/ or /k/ (but not /t/), the duration of the stop silence, and that of the preceding segment could also be juncture cues. The longer duration of frication, nasal murmur, and stop silence at word-initial position observed across all stimulus types correspond with the findings of the lengthening of word-initial consonants in early studies (e.g., Oller, 1973; Furgeron and Keating, 1997).

D. Procedures

The participants were tested individually in a sound-attenuated room in the Speech-Language-Hearing Sciences Department at the CUNY Graduate Center, using computer software, PARADIGM ID (written by Tagliaferri) in which participants responded by clicking on one of two response alternatives printed on the screen.⁶ The experiment consisted of the main experiment and the familiarity questionnaire.

1. Main experiment

The stimuli were presented in random order, except that the members of each stimulus pair were separated by at least 15 other stimuli. The stimuli were presented binaurally through headphones (Telephonics TDH 39). The sound level was set in such a way that a 1000-Hz pure tone with the same amplitude as the median rms value of all stimuli was presented at 70 dB SPL. For each trial, the printed stimulus pair appeared on the computer screen approximately 1 s after

the presentation of the auditory stimulus. One member of each stimulus pair on the screen always matched the auditory stimulus. Each pair was placed side by side and the position of the correct answer was randomized. Participants were informed by written instructions on the screen that they would hear two to three English words embedded in the sentence “Say___ again,” and that they would be asked to click on one of the written alternatives that represented the speech stimulus. After choosing one member of the pair, they were asked to indicate how confident they were in their answer by clicking on one of the seven boxes corresponding to the seven levels of confidence rate (1=*not sure* to 7=*very confident*). They were specifically instructed that though some of the target phrases might sound awkward or unfamiliar, their choice should be based only on what they heard, not on the familiarity of the words. Participants responded to six sample stimulus pairs in the familiarization test before doing the main experiment.

2. Familiarity questionnaire

The main experiment was followed by the familiarity questionnaire that asked the participants whether one of each stimulus pair was more familiar than the other. The written stimulus pairs were presented on the screen without speech stimuli, and participants were asked to click on the member of the pair that was more familiar to them, or to click on *Same* when the two alternatives were equally familiar to them.

III. RESULTS

This section reports the results in the following order: (1) Japanese listeners’ performance vs AE listeners as a group, (2) item analysis, and (3) correlation between Japanese listeners’ LOR and performance. In the first part, the overall performance by the Japanese group, including their

TABLE III. Mean percent correct for each stimulus type (positive vs negative): japanese and AE subjects.

Stimulus type	American subjects (n=10)			Japanese subjects (n=30)		
	Mean	SD ^a	SE ^b	Mean	SD	SE
Aspiration (<i>n</i> =36) (e.g., Lou stops/loose tops)	95.8	4.4	1.46	73.1	9.4	1.74
Positive aspiration (<i>n</i> =18; e.g., loose tops)	98.3	3.7	1.25	73.5	15.5	2.88
Negative aspiration (<i>n</i> =18; e.g., Lou stops)	93.3	9.0	3.00	72.8	13.0	2.41
Glottal stop (<i>n</i> =36) (e.g., a niche/an itch)	96.9	1.6	0.53	91.3	6.2	1.15
Positive glottal stop (<i>n</i> =18; e.g., an itch)	98.9	2.3	0.78	95.9	7.0	1.30
Negative glottal stop (<i>n</i> =18; e.g., a niche)	95.0	3.2	1.05	86.7	11.5	2.13
Double cue (<i>n</i> =12) (e.g., grey pin/grape in)	99.2	2.6	0.88	93.6	7.8	1.45
Double cue (Asp-/Glott+) (<i>n</i> =6; e.g., grape in)	100.0	0.0	0.00	99.4	3.0	0.57
Double cue (Asp+/Glott-) (<i>n</i> =6; e.g., grey pin)	98.3	5.3	1.76	87.8	15.1	2.81
All stimuli (<i>n</i> =84)	96.8	2.1	0.70	83.8	5.7	1.06

^aSD=standard deviation.^bSE=standard error.

performance on each stimulus type, is reported in relation to that of the AE group first, then the influence of the presence or absence of the acoustic cues (denoted as positive and negative) on the listeners' performance is reported. Furthermore, Japanese listeners' performance on the subcategories of each stimulus type with positive and negative distinctions is compared. The second part compares the most difficult items for Japanese listeners with those for AE listeners and also examines the performance on the items with irregular acoustic properties (e.g., the positive glottal stop items that do not contain glottal stop) by each language group. Furthermore, whether or not the confidence ratings showed close relations with the error rates of test items was investigated. In the third part, the correlation of the Japanese subjects' performance with their LOR in English-speaking countries is reported.

A. Comparisons of Japanese vs AE groups

1. Overall results and results on each stimulus type

The mean percentage of correct responses by Japanese and AE subjects for each stimulus type is presented in Table III. Nonparametric Median tests evaluated by chi-square were chosen for statistical analysis of the comparisons between the AE and Japanese groups because (a) the numbers of subjects in the two groups were different, and (b) the American group's performance showed ceiling effects. The AE group's performance was significantly higher overall than that of the Japanese group (χ^2 [1df]=12.06, $p < 0.001$). Three separate Median tests were conducted to compare the performance on aspiration stimuli (χ^2 [1df]=12.06, $p < 0.001$), on glottal stop stimuli (χ^2 [1df]=8.888, $p < 0.01$), and on double cue stimuli (χ^2 [1df]=5.004, $p < 0.05$). The

AE group's performance was significantly better than that of the Japanese group for all three types of contrasts.

The data also show that the Japanese group was much less accurate on Aspiration stimuli (73.1%) than on Glottal stop stimuli (91.3%) or on Double cue stimuli (93.6%) while the AE group performed above 95% on all stimulus types. In addition, one can assume, from the Japanese group's rather high standard deviations (5.7–9.4) compared to those of the AE group (1.6–4.4), that Japanese subjects' performance was more variable than that of the AE group.

Since the AE group's performance was at or near ceiling on all stimulus types, within-group comparisons were performed only for the Japanese data. A nonparametric Friedman's test and Wilcoxon signed-ranks tests were chosen for examining differences in performance across stimulus types because of the unequal variances, skewed distributions, and ceiling effects seen in performance on Glottal stop and on Double cue stimuli. Friedman's test showed that Japanese listeners' performance differed significantly by stimulus types (χ^2 [2df]=46.41, $p < 0.001$). Wilcoxon tests further revealed that their performance on Aspiration stimuli was significantly poorer than that on Glottal stop stimuli ($z=4.62$, $p < 0.001$) and on Double cue stimuli ($z=4.78$, $p < 0.01$), while the difference in performance on Glottal stop and Double cue stimuli did not reach significance ($z=1.71$, $p > 0.05$).

2. Results on positive and negative stimuli

Table III also shows the mean percent correct responses by Japanese and AE subjects for Positive and Negative stimuli of each stimulus type. Since the AE group's performance was near ceiling on both Positive and Negative stimuli (though it appears to show slightly better perfor-

FIG. 2. Japanese listeners' mean percent correct: Aspiration stimuli (left panel) and Glottal stop stimuli (right panel). (Error bar = standard error.)

mance on Positive stimuli than Negative stimuli), statistical comparison between the performance on Positive and Negative stimuli was performed only for the Japanese data.

Wilcoxon signed-ranks tests revealed that Japanese listeners' performance on Positive and Negative stimuli did not differ significantly for Aspiration stimuli ($z=0.55$, $p>0.05$) whereas their performance did differ significantly for Glottal stop stimuli ($z=3.07$, $p<0.01$) and for Double cue stimuli ($z=3.30$, $p<0.01$). Regarding the perception of Glottal stop stimuli, Japanese listeners performed significantly better when the glottal stop cue was present than when it was absent. Their performance on Double cue stimuli also indicated that they performed better on the stimuli with a positive glottal stop cue without aspiration cue than on the stimuli with an aspiration cue but no glottal stop.

Figure 2 shows Japanese listeners' mean percentage of correct responses on subcategorized groups of Aspiration stimuli and Glottal stop stimuli. Here, again, the data for the subcategories of Glottal stop stimuli indicate that the negative stimuli of each subcategory, indicated by “-” in the graph, were always lower than the corresponding positive stimuli, indicated by “+” while those of Aspiration stimuli do not show obvious patterns of such tendencies.⁷ Thus, Japanese listeners apparently did not utilize the presence/absence of aspiration cues effectively for word segmentation whereas they counted heavily on the presence of glottal stop cues as word juncture cues.

B. Item analysis

This section examines whether or not Japanese listeners and AE listeners made frequent errors on the same items and whether or not their confidence ratings reflected their difficulty in correctly perceiving the items. Figure 3 illustrates the relations between the error rate of each item and corresponding mean confidence rating by Japanese listeners (left panel) and by AE listeners (right panel) in scatter plots.

The items that were mistaken most by Japanese listeners were *keeps canning* (73.3% errors), *tops crawled* (60.0% errors), *keep stalking* (56.7% errors), and *Lou skis* (56.7% errors), and the items that were most error-inducing for AE listeners were *tea mat* (40% errors), *cook screams* (30% errors), *say vamps* (30% errors), and *cook sprints* (30% errors), showing no overlap of the most difficult items for the two

groups. Furthermore, the error-prone items for Japanese listeners were all Aspiration stimuli whereas the error-prone items for AE listeners included both Glottal stop stimuli (*tea mat* and *say vamps*) and Aspiration stimuli (*cook screams* and *cook sprints*).

Although the error rates were not as high as for the items mentioned above, the Japanese group showed higher error rates on some of the items with irregular acoustic properties than on the other items in the same stimulus type. The AE group did not show such a tendency. The error rates of *an ice man* and *save amps*, which were categorized as Positive Glottal stop stimuli but did not have glottal stop/creaky voice, were 13.3% and 26.7%, respectively, for the Japanese group while the error rate for the rest of Positive Glottal stop stimuli was 1.7% on average. On the other hand, the AE group's error rate for *an ice man* was 0% and that for *save amps* was 10% (i.e., only one person answered incorrectly), showing no distinct difference from the mean error rate of the rest of Positive Glottal stop stimuli (0.6%). These results indicate the possibility that both Japanese and AE listeners were using acoustic cues other than glottal stop/creaky voice, although Japanese listeners were more disadvantaged by the lack of the glottal stop/creaky voice cue. For another irregular item *light old*, which was categorized into Double cue (Asp-/Glott+) stimuli but the glottal stop was not distinguishable from the stop silence, the error rate was 3.3% for the Japanese group and 0% for the AE group, showing no greater difficulty than the rest of Asp-/Glott+ stimuli by either the Japanese (6.1%) or by the AE (0%) group.

Spearman rank-order correlations were calculated to examine the relations between the confidence ratings and error rates for both Japanese and AE listeners.⁸ The ρ values were -0.778 for Japanese listeners and -0.377 for AE listeners.

FIG. 3. Mean confidence rating and error rate by Japanese listeners (left panel) and by AE listeners (right panel).

FIG. 4. Japanese listeners' LOR and performance: overall performance (top left panel), performance on Glottal stop stimuli (top right panel), on Double cue stimuli (bottom left panel), and on Aspiration stimuli (bottom right panel).

ers after the correction for tied ranks (Siegel and Castellan, 1988, p. 239). For the AE group, the correlation was limited by the large number of errorless scores (ceiling effects). For the Japanese group, the correlation was quite high, but the range in confidence ratings was limited, with relatively large variations across subjects in ratings for both easy and difficult items. The mean confidence ratings remained quite high even when more than half of the listeners answered incorrectly (see left panel of Fig. 3).

C. Correlation between Japanese listeners' LOR and performance

In order to investigate whether greater L2 experience was associated with listeners' better use of L2 acoustic-phonetic cues for word segmentation, Spearman rank-order correlations between Japanese listeners' LOR in English-speaking countries and their accuracy in the experiment were computed. Figure 4 illustrates the relation between their LOR and overall performance and those between LOR and performance on each of the three different stimulus types. All correlations were positive, although some were rather low, due to ceiling effects that restricted variability. Overall accuracy rates (top left panel) ($\rho = +0.633$) increased rapidly over the first couple of years but flattened after that without reaching ceiling. However, performance seems to have increased again for participants with LOR of 9 years or greater. Performance on Glottal stop stimuli (top right panel) shows a similar tendency to that for overall performance ($\rho = +0.573$), demonstrating a rapid increase in response accuracy

in the first couple of years of LOR with rates near the ceiling after that. The correlations of LOR with performance on Double cue stimuli (bottom left panel) ($\rho = +0.385$) and Aspiration stimuli (bottom right panel) ($\rho = +0.469$) were lower than for overall performance and Glottal stop stimuli. Performance on Double cue stimuli showed ceiling effects because many subjects obtained perfect scores regardless of their LORs, although subjects with lower LOR varied more among themselves. For Aspiration stimuli, accuracy rates were widely distributed across LOR, although the subjects with more than 9 years of LOR tended to have higher scores than the rest of the subjects. Comparing these results with overall performance, it appears that the rapid overall improvement in the first few years of LOR is attributable to L2 listeners picking up on the glottal stop cue while the late improvement (i.e., the higher performance after 9 years of LOR) is due to their acquiring the ability to tap into the aspiration cue.

D. Effects of familiarity with phrases on Japanese listeners' performance

The results for the familiarity test for Aspiration stimulus pairs and Glottal stop stimulus pairs by Japanese listeners are summarized in Table IV. As can be seen, the numbers of more familiar items for Positive stimuli and for Negative stimuli in both Aspiration and Glottal stop stimuli were almost evenly distributed, showing no strong bias toward either Positive or Negative stimuli.

For Aspiration stimulus pairs, a tendency that may indicate response bias favoring more familiar items in perfor-

TABLE IV. Results of Japanese listeners' familiarity questionnaire.

Stimulus pair type	Familiarity answer	No. of pairs	No. of more familiar item better identified	Example (% correct in parentheses)
Aspiration stimulus pairs (18 pairs)	Positive item as more familiar	8	7/8 (87.5%)	<i>loose keys</i> vs <i>Lou skis</i> (90.0%) vs (43.3%)
	Negative item as more familiar	10	7/10 (70%)	<i>keep scanning</i> vs <i>keeps canning</i> (76.7%) vs (26.7%)
Glottal stop stimulus pairs (18 pairs)	Positive item as more familiar	10	9/10 (90%)	<i>grave at</i> vs <i>grey vat</i> (100%) vs (66.7%)
	Negative item as more familiar	8	0/8 (0%)	<i>we vend</i> vs <i>weave end</i> (80.0%) vs (93.3%)
Double cue stimulus pairs (6 pairs)	Asp-/Glott+ item as more familiar	3	2/3 (66.7%) (1 pair was both 100% correct)	<i>make art</i> vs <i>may cart</i> (100%) vs (73.3%)
	Asp+/Glott- item as more familiar	3	0/3 (0%)	<i>grey pin</i> vs <i>grape in</i> (73.3%) vs (100%)

mance was observed. The more familiar items tended to be better identified than their counterparts. Seven out of eight positive items and seven out of ten negative items that were chosen as more familiar had higher accuracy rates than the corresponding members of the pairs did. This means that Japanese listeners chose the more familiar items as the answers in the experiment significantly more often than their less familiar counterparts (14/18, significantly different from chance [$p < 0.02$] by the binomial expansion). This tendency was not seen in Glottal stop stimulus pairs. In Glottal stop pairs, none of the more familiar negative items was identified more accurately than their positive counterparts while nine out of ten more familiar positive items were better identified than their negative counterparts. The better performance on Positive Glottal stop stimuli than on their negative counterparts, regardless of familiarity, indicates a lack of response bias for familiar items in identification of word onsets cued by glottal stops (9/18, not significantly different from chance). A similar pattern is seen in Double cue stimulus pairs. None of Asp+/Glott- items were better identified than their Asp-/Glott+ counterparts irrespective of their relative familiarity with their counterparts (2/6, not significantly different from chance). These results suggest that familiarity with words or phrases does not seriously affect how listeners segment speech when they can effectively use allophonic segmentation cues, but may affect segmentation when they cannot take advantage of the allophonic cues.

IV. DISCUSSION

This section summarizes the results of the study, offering answers to the three research questions presented in the Introduction, and then considers the limitations of the study and some directions for future studies.

A. Perceptual differences between Japanese and AE listeners

The results clearly indicated that Japanese L2 listeners, as a group, perceived English word boundaries more poorly

than native AE listeners in many ways. First, Japanese listeners' overall performance was significantly worse than that of AE listeners, suggesting that Japanese listeners could not use the available cues as effectively as native listeners did. Second, Japanese listeners showed a significant difference in their performance on different stimulus types while AE listeners' performance was near ceiling on all stimulus types. Japanese performance on Aspiration stimuli was significantly worse than that on Glottal stop stimuli and on Double cue stimuli while the AE group's correct rates were consistently above 95%. Third, the items that were mistaken most by Japanese listeners and by AE listeners did not overlap. The four most error-prone items for Japanese listeners were all Aspiration stimuli, whereas those of AE listeners included both Glottal stop stimuli and Aspiration stimuli, implying different perceptual patterns by Japanese and AE listeners. Fourth, Japanese listeners seemed to have a poorer sense than AE listeners of how well they were able to perceive the word boundaries. The scatter plots of confidence ratings and correct performance showed that the Japanese listeners' confidence ratings did not decline much even when more than half of the listeners answered incorrectly, while the AE listeners' confidence ratings better reflected their error rates on the few items that were difficult. Finally, Japanese listeners seemed to be more disadvantaged than native AE listeners by the lack of the most distinctive acoustic cue for segmentation. Japanese listeners misperceived the word boundaries of the irregular Positive Glottal stop stimuli that did not have glottal stop/creaky voice more often than the stimuli with glottal stop/creaky voice present, while AE listeners did not show such a difference. These results suggest that when a reliable acoustic cue is absent, Japanese listeners are not capable of taking advantage of the remaining acoustic cues to the same extent as AE listeners. These results generally replicate those reported by [Altenberg \(2005\)](#) for Spanish late L2 learners of English on the same materials spoken by the different speaker.

B. Relative perceptual problems with different types of segmentation cues

Japanese listeners made significantly more errors on pairs differentiated by aspiration cues than they did on pairs differentiated by glottal stop cues. Furthermore, the data comparing Japanese listeners' performance on Positive and Negative stimuli within each stimulus type show that they performed significantly better when the glottal stop cues were present than when they were absent. However, the presence or absence of the aspiration cues did not significantly affect their performance. This pattern of results replicates those reported by [Altenberg \(2005\)](#) for Spanish L2 learners with the exception that her subjects showed better relative performance on both Positive Aspiration stimuli and Positive Glottal stop stimuli.

We can tentatively conclude that, independent of differences in the use of aspiration and glottal stops in the L1, aspiration cues (VOT and correlated duration parameters) are harder to learn by L2 learners than are glottal stop cues (and related parameters) for segmentation. In the course of discussing the cause of relative difficulty for Spanish listeners, [Altenberg \(2005\)](#) explored the concept of markedness as one of the possible factors. That is, "glottal stop insertion is less marked and hence easier to acquire than aspiration" (pp. 344–345). It has been observed by speech scientists that initiating phonation with a glottal stop ([Borden et al., 2003](#)) and inserting a glottal stop at the boundary of abutting words ending and beginning with vowels ([Kent and Read, 2002](#)) are quite common by some speakers. Furthermore, glottal stop insertion at a word juncture preceding a word-initial vowel is commonly observed in many languages other than English, Japanese, and Spanish (e.g., [Kohler, 1994](#); [Rubach, 2000](#)). Thus, it might have been the case that the less marked glottal insertion was learned more easily as a segmentation cue by Japanese listeners than the aspiration cue.

Another possible explanation may be related to the influence of the L1, or L1 transfer, as also suggested by [Altenberg \(2005\)](#). In comparing the pattern of relative difficulty across Japanese and Spanish listeners the two groups showed comparable performance on Glottal stop and Double cue stimuli, whereas Japanese listeners performed relatively better than Spanish listeners did on Aspiration stimuli (although worse than AE controls). This pattern of relative difficulty in utilizing aspiration cues corresponds to the extent of word-initial aspiration of voiceless stops in the three languages (AE > Japanese > Spanish), as discussed earlier. It has to be noted, however, that the data reported here and those of [Altenberg \(2005\)](#) are not completely suitable for comparison because of differences with regard to subject inclusion criteria. [Altenberg \(2005\)](#) tested college-level English as a second language students, whereas the present study included participants who ranged from those with almost no immersion experience to those with more than 10 years of immersion experience.

It is also useful to consider if constraints of L1 syllable structure affect Japanese listeners' L2 perception of word

boundaries. Japanese syllable structure does not allow a consonant cluster within a syllable because the majority of Japanese syllables are CV syllables, with some exceptions of CVC syllables involving syllable-final moraic consonants (moraic final nasal and geminate consonants). If L1 syllable constraints influence the perception of L2 syllable patterns, therefore, one may predict that Japanese listeners would have more difficulty in segmenting words correctly as the consonant cluster at the word juncture in Aspiration stimuli became more complex. The results, however, did not show any obvious tendency. In the Aspiration stimuli, the consonant cluster at word juncture becomes more complex in the following order: VsC group, CsC group, and CsCC group. The mean accuracies on the above groups by Japanese listeners were 73.9%, 74.4%, and 71.1%, respectively, and did not differ from each other.⁹

Thus, the results regarding L1 transfer seem to be mixed. The relation between the degree of aspiration in L1 and that in L2 seems to suggest L1 transfer in perceiving L2 word boundaries whereas L1 syllable patterns do not seem to affect the perception of L2 syllable at word boundaries.

C. Relationship between immersion experience and use of allophonic cues

The positive relationships between Japanese listeners' performance and their LOR suggest that increased English experience does improve Japanese listeners' use of English allophonic cues. Their overall response accuracy increased as a function of their LOR, although it was not linearly related. A rapid difference in response accuracy across the first 1–3 years of LOR followed by a plateau was observed. Performance appeared to improve again after 9–10 years, although the sampling of participants with long LOR was limited. While Japanese listeners seem to learn to exploit glottal stop cues for English word segmentation within the first few years of residence, their mastery of aspiration cues may come much later. The pattern seen in the overall performance seems to reflect this "time lag" of learning different allophonic cues. These tendencies need to be followed up in future studies with larger numbers of participants with extensive immersion experience.

D. Other issues: Limitations and future studies

In addition to the findings related to the research questions, the present study suggested that familiarity with words or phrases may have affected Japanese listeners' responses in the experiment when they were not using the allophonic cues effectively. The results of the familiarity questionnaire revealed that for Aspiration stimuli, on which Japanese listeners performed significantly worse than the other types of stimuli, items that were chosen by Japanese listeners as more familiar had higher accuracy rates than the corresponding members of the pairs, indicating a response bias toward the familiar items. Since this tendency was not seen in Glottal stop stimuli, the results imply that familiarity with words or phrases affected performance only when Japanese listeners

could not fully use the acoustic-phonetic segmentation cues. This agrees with the finding reported by [Mattys and Melhorn \(2007\)](#) that lexical cues influence segmentation decisions when allophonic cues are ambiguous. [Altenberg \(2005\)](#) attempted to minimize lexical effects when constructing the stimuli; however, the present results suggest that it may not be possible to eliminate the effects of lexical bias completely from word segmentation studies of this kind. Future studies must take these effects into consideration when interpreting results.

As mentioned in the Introduction, cross-language research on L2 word segmentation is not extensive. There are fewer studies dealing with individual acoustic-phonetic cues than those examining rhythmic cues. The present study examined the use of only two of the many allophonic cues used in English word segmentation. Furthermore, the issue of coarticulatory effects across word boundaries, such as assimilation and deletion, which may cause significant changes in acoustic/phonetic properties of phonetic segments, has not been addressed sufficiently in L2 word segmentation studies, in spite of the fact that such coarticulatory variations commonly occur in connected speech. Future studies should examine these issues. In the process of L2 learning, non-native listeners must learn to perceive the distinctive segments (phoneme contrasts), the phonotactic patterns, and the allophonic variations of phonetic segments that serve as segmentation cues. Research on all aspects of L2 phonological learning will provide us with a better view of the problems associated with late language learning and guide our efforts to improve L2 language processing through intensive L2 training methods.

ACKNOWLEDGMENTS

We thank Evelyn Altenberg for her helpful comments and suggestions in preparing this manuscript. This research was supported in part by a grant to W.S. (NIH Grant No. DC00323).

APPENDIX: STIMULUS PAIRS USED IN THE EXPERIMENT [TAKEN FROM [ALTENBERG \(2005\)](#)]

ASPIRATION STIMULUS PAIRS

Positive stimuli	Negative stimuli
<i>VsC group</i>	
loose pills	Lou spills
loose tops	Lou tops
loose keys	Lou skis
lace peach	lay speech
lace table	lay stable
lace car	lay scar
<i>CsC group</i>	
keeps parking	keep sparking
keeps talking	keep stalking
keeps canning	keep scanning
chief's port	chief sport
chief's tar	chief star

chief's cool	chief school
<i>CsCC group</i>	
cook's prints	cook sprints
cook's truck	cook struck
cook's creams	cook screams
tops pry	top spry
tops trains	top strains
tops crawled	top scrawled

GLOTTAL STOP STIMULUS PAIRS

Positive stimuli Negative stimuli

<i>Nasal group</i>	
seen either	see neither
an itch	a niche
an ice man	a nice man
seem able	see Mabel
team at	tea mat
claim annual	clay manual

<i>Obstruent group</i>	
wife ill	why fill
loaf ate	low fate
loaf ailing	low failing
grave at	grey vat
weave end	we vend
save amps	say vamps

<i>Liquid group</i>	
I'll earn	I learn
sail east	say least
seal of	see love
beer old	be rolled
cure ache	cue rake
beer ice	be rice

DOUBLE CUE STIMULUS PAIRS

Asp–/Glott+ Stimuli Asp+/Glott– Stimuli

grape in	grey pin
weep at	we pat
light old	lie told
might owe	my toe
make art	may cart
like old	lie cold

Stimulus pairs used in the familiarization task
[taken from [Nakatani and Dukes \(1977\)](#) except for *ice cream/I scream*]

ice cream/I scream	lawn chair/launch air
we'll own/we loan	stayed ill/stay dill
plate ought/play taught	no notion/known ocean

¹Throughout this study, we will refer to these cues as aspiration cues and glottal stop cues, although an acoustic analysis of the stimuli revealed that there were other acoustic parameters that differentiated word-initial consonants and vowels.

²Although [Altenberg \(2005\)](#) selected the stimuli to minimize lexical influences on the experiment by using native speakers' naturalness rating, it cannot be ascertained whether L2 listeners' lexical bias contributed to the results. [Altenberg \(2005\)](#) reported that word frequency was not likely to affect the results by comparing the average frequency counts of words in

each stimulus type and the corresponding Spanish listeners' performance (pp. 348–349). In the present study, however, the familiarity test was designed for a more direct investigation of the possible correlation between participants' familiarity with words or phrases and their performance on the segmentation task.

³All Japanese participants were asked to report experience in staying in English-speaking countries only if it lasted for 1 month or longer, except for one participant who had come to the United States only 2 weeks before the test.

⁴Although it can also be pronounced as /niʃ/, the word *niche* was pronounced as /nɪʃ/ in this experiment in order to make the phrase *a niche* nearly homophonous to the phrase *an itch*.

⁵The data for the Liquid group were not included in the statistic analysis because of the non-segmentability of the liquid portion in Positive stimuli.

⁶The original *Altenberg experiment* (2005) tested the participants in small groups in a quiet room with a fixed interval in which to respond on answer sheets.

⁷Wilcoxon signed-ranks tests showed that Japanese listeners' performance difference between Positive and Negative stimuli did not reach significance in any of subcategorized groups of Aspiration stimuli ($z=1.76$ for VsC group; $z=0.94$ for CsC group; $z=1.43$ for CsCC group).

⁸Because of the nonlinearity of the relationships, the non-parametric Spearman test that does not assume normal distribution or linearity of the relationship was employed (Siegel and Castellan, 1988).

⁹Wilcoxon signed-ranks tests revealed that none of the Japanese listeners' performance differences between subcategorized groups of aspiration stimuli reached significance ($z=0.08$ for VsC vs CsC group; $z=1.09$ for VsC vs CsCC group; $z=1.13$ for CsC vs CsCC group).

Altenberg, E. P. (2005). "The perception of word boundaries in a second language," *Second Lang. Res.* **21**, 325–358.

Bloch, B. (1950). "Studies in colloquial Japanese IV: Phonemics," *Language* **26**, 86–125.

Borden, G. J., Harris, K. S., and Raphael, L. J. (2003). *Speech Science Primer*, 4th ed. (Lippincott Williams and Wilkins, Philadelphia, PA).

Bradley, D. C., Sanchez-Casas, R. M., and Garcia-Albea, J. E. (1993). "The status of the syllable in the perception of Spanish and English," *Lang. Cognit. Processes* **8**, 197–233.

Christie, W. M. (1974). "Some cues for syllable juncture perception in English," *J. Acoust. Soc. Am.* **55**, 819–821.

Church, K. (1987). "Phonological parsing and lexical retrieval," *Cognition* **25**, 53–69.

Cutler, A., and Butterfield, S. (1992). "Rhythmic cues to speech segmentation: Evidence from juncture misperception," *J. Mem. Lang.* **31**, 218–236.

Cutler, A., and Norris, D. (1988). "The role of strong syllables in segmentation for lexical access," *J. Exp. Psychol. Hum. Percept. Perform.* **14**, 113–121.

Cutler, A., and Otake, T. (1994). "Mora or phoneme? Further evidence for language-specific listening," *J. Mem. Lang.* **33**, 824–844.

Cutler, A., Mehler, J., Norris, D., and Segui, J. (1986). "The syllables' differing role in the segmentation of French and English," *J. Mem. Lang.* **25**, 385–400.

Cutler, A., Mehler, J., Norris, D., and Segui, J. (1992). "Monolingual nature of speech segmentation by bilinguals," *Cogn. Psychol.* **24**, 381–410.

Fougeron, C., and Keating, P. A. (1997). "Articulatory strengthening at edges of prosodic domains," *J. Acoust. Soc. Am.* **101**, 3728–3740.

Goettry, V., and Kolinsky, R. (2000). "The role of rhythmic cues for speech

segmentation in monolingual and bilingual listeners," *Psychol. Belg.* **40**, 115–152.

Gow, D. W., and Gordon, P. C. (1995). "Lexical and prelexical influences on word segmentation: Evidence from priming," *J. Exp. Psychol. Hum. Percept. Perform.* **21**, 344–359.

Hattori, S. (1951). *Onseigaku (Phonetics)* (Iwanami, Tokyo).

Homma, Y. (1980). "Voice onset time in Japanese stops," *Onsei Gakkai Kaiho* **163**, 7–9.

Kawakami, S. (1977). *Nihongo Onsei Gaisetsu (An Outline of Japanese Phonetics)* (Ofusha, Tokyo).

Kent, R., and Read, C. (2002). *The Acoustic Analysis of Speech*, 2nd ed. (Singular, San Diego, CA).

Kohler, K. J. (1994). "Glottal stops and glottalization in German," *Phonetica* **51**, 38–51.

Mattys, S., and Melhorn, J. F. (2007). "Sentential, lexical, and acoustic effects on the perception of word boundaries," *J. Acoust. Soc. Am.* **122**, 554–567.

Mattys, S., White, L., and Melhorn, J. F. (2005). "Integration of multiple speech segmentation cues: A hierarchical framework," *J. Exp. Psychol. Gen.* **134**, 477–500.

McQueen, J. M. (1998). "Segmentation of continuous speech using phonotactics," *J. Mem. Lang.* **39**, 21–46.

Mehler, J., Dommergues, J. Y., Frauenfelder, U., and Segui, J. (1981). "The syllable's role in speech segmentation," *J. Verbal Learn. Verbal Behav.* **20**, 298–305.

Nakatani, L. H., and Dukes, K. D. (1977). "Locus of segmental cues for word juncture," *J. Acoust. Soc. Am.* **62**, 714–719.

Oller, D. K. (1973). "The effect of position in utterance on speech segment duration in English," *J. Acoust. Soc. Am.* **54**, 1235–1247.

Otake, T., Hatano, G., Cutler, A., and Mehler, J. (1993). "Mora or syllable? Speech segmentation in Japanese," *J. Mem. Lang.* **32**, 258–278.

Riney, T. J., Takagi, N., Ota, K., and Uchida, Y. (2007). "The intermediate degree of VOT in Japanese initial voiceless stops," *J. Phonetics* **35**, 439–443.

Rubach, J. (2000). "Glide and glottal stop insertion in Slavic languages: A DOT analysis," *Linguistic Inquiry* **31**, 271–317.

Sanders, L. D., and Neville, H. J. (2000). "Lexical, syntactic, and stress-pattern cues for speech segmentation," *J. Speech Lang. Hear. Res.* **43**, 1302–1321.

Sanders, L. D., and Neville, H. J. (2003a). "An ERP study of continuous speech processing II. Segmentation, semantics, and syntax in non-native speakers," *Brain Res. Cognit. Brain Res.* **15**, 214–227.

Sanders, L. D., and Neville, H. J. (2003b). "An ERP study of continuous speech processing I. Segmentation, semantics, and syntax in native speakers," *Brain Res. Cognit. Brain Res.* **15**, 228–240.

Sanders, L. D., Neville, H. J., and Woldorff, M. G. (2002). "Speech segmentation by native and non-native speakers: The use of lexical, syntactic, and stress-pattern," *J. Speech Lang. Hear. Res.* **45**, 519–530.

Siegel, S., and Castellan, N. J., Jr. (1988). *Nonparametric Statistics for the Behavioral Sciences*, 2nd ed. (McGraw-Hill, Boston, MA).

Stockwell, R. P., Bowen, J. D., and Silva-Fuenzalida, I. (1956). "Spanish juncture and intonation," *Language* **32**, 641–665.

Vance, T. J. (1987). *An introduction to Japanese Phonology* (State University of New York Press, New York).

Williams, L. (1977). "The voicing contrast in Spanish," *J. Phonetics* **5**, 169–184.