



**HAL**  
open science

## The perception of word boundaries in a second language

Evelyn P. Altenberg

► **To cite this version:**

Evelyn P. Altenberg. The perception of word boundaries in a second language. *Second Language Research*, 2005, 21 (4), pp.325-358. 10.1191/0267658305sr250oa . hal-00572082

**HAL Id: hal-00572082**

**<https://hal.science/hal-00572082>**

Submitted on 1 Mar 2011

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# The perception of word boundaries in a second language

**Evelyn P. Altenberg** *Hofstra University*

Adult Spanish second language (L2) learners of English and native speakers of English participated in an English perception task designed to investigate their ability to use L2 acoustic-phonetic cues, e.g., aspiration, to segment the stream of speech into words. Subjects listened to a phrase and indicated whether they heard, e.g., *keep sparking* or *keeps parking*. The results indicate that learners are significantly worse than native speakers at using acoustic-phonetic cues, and that some types of stimuli are easier for learners to segment than others. The findings suggest that various factors, including transfer and markedness, may be relevant to success in L2 segmentation.

## I Introduction

In order to comprehend language, part of what the listener needs to do is to segment the continuous stream of incoming speech into meaningful units. A number of researchers (e.g., Quené, 1992; Gow and Gordon, 1995) suggest that speakers of a language need to use acoustic-phonetic information to segment speech and retrieve words, probably in conjunction with other information, for example, syntactic information. The majority of research in second language (L2) speech has explored speech production (Leather, 1999); those studies dealing with L2 speech perception have focused largely on the perceptual categorization of L2 speech sounds. (For reviews, see, for example, Leather, 1999; Major, 2001.) Thus, our knowledge of second language speech perception in general is limited, our knowledge of how well second language

learners are able to use acoustic-phonetic information in L2 speech perception even more so. The goal of the study reported here is to explore the abilities of adult learners of English to use acoustic-phonetic cues to segment the second language stream of speech into words. Learners listened to English stimuli such as *keep sparking* and indicated whether they heard, for example, *keep sparking* or *keeps parking*. Such an investigation adds to our knowledge of what learners do and do not know in terms of L2 speech perception and provides an avenue for the exploration of relevant issues in second language acquisition.

The study addresses the following two questions:

- How do second language learners compare to native speakers in their ability to use acoustic-phonetic cues to segment natural speech into words? In particular, is there evidence that learners are able to acquire and use second language acoustic-phonetic cues?
- Are some stimuli easier than others for second language learners to segment and, if so, what factors can account for these differences? In particular, is there evidence of transfer at the allophonic and/or syllabic levels, and/or of universal factors, such as markedness, in L2 speech segmentation?

Two issues that have been central to the study of L2 phonology, transfer and markedness, are relevant to the investigation here. Research has typically indicated that transfer from the first language plays a significant role in second language perception of segments. A number of theories of L2 speech perception – for example, Best's Perceptual Assimilation Model (1995), Flege's Speech Learning Model (1995) and Brown's (1998) Feature-Geometry-based model – while they differ in significant ways, rest on the claim that characteristics of the first language determine non-native speakers' perception of segments in the second language.

While substantial evidence exists that first language syllable structure influences second language production (e.g., Broselow and Finer, 1991; Eckman and Iverson, 1993), the evidence regarding the role of first language syllable structure in second language perception is mixed. Hallé *et al.* (1998), Dupoux *et al.* (1999) and Tench (2003) are examples of studies that provide evidence supporting the role of first language

(L1) syllable structure in L2 perception. However, Brown (1998: 171) found that 'L1 syllable structure does not have the same 'blocking' effect on acoustic perception that an L1 feature geometry does.' And Altenberg (2002) found evidence for transfer of L1 phonotactic constraints in an L2 production task but not in a task involving the perception of L2 word-initial consonant clusters. Therefore, the question of whether or not L1 syllable structure necessarily plays a role in L2 speech perception remains open. A speech segmentation task such as that used here allows one to further explore the potential role of L1 information, including L1 syllable structure, in L2 perception.

As Bohn (1995) says, with respect to second language speech perception, 'transfer doesn't tell it all.' Another potentially relevant factor is markedness, which relates to the notion of naturalness. Markedness is typically defined according to implicational hierarchies and/or statistical frequencies and can also be related to first language order of acquisition and to historical change (e.g., Major, 2001). A number of studies (e.g., Eckman and Iverson, 1993; Major, 1996) suggest that less marked sequences and segments may be easier for an L2 learner to produce than those that are more marked. However, in terms of perception, something that is more marked may in fact stand out because of its markedness, and hence be noticed more than an unmarked segment. In fact, Yavaş (1998) refers to the unmarked as the expected, and the marked as the unexpected. Thus, something that is more marked may be more perceptually salient. If that is the case, then one would expect, for example, that speakers will be more successful at segmenting a juncture with a more marked onset than one with a less marked onset, all other factors being equal. The ability of markedness to account for speech segmentation data is examined below.

Turning now to native language speech segmentation, various kinds of information have the potential to play a role in the segmentation process for speakers of a language, including lexico-semantic cues (e.g., Quené, 1992), syntactic cues (e.g., Sanders and Neville, 2000) and acoustic-phonetic cues (e.g., Church, 1987). Thus, speakers may use their lexical knowledge by segmenting the speech stream where they recognize a familiar word; syntactic and/or morphological knowledge may be used, for example, by segmenting the speech stream after a verb's *-ing* ending in English; speakers may use their acoustic-phonetic

knowledge by, for example, segmenting a syllable before an aspirated voiceless stop in English. It is unlikely that listeners rely on only one kind of cue when segmenting speech (e.g., Cutler *et al.*, 1986; Grosjean and Gee, 1987; Brent and Cartwright, 1996; Vroomen *et al.*, 1996; McQueen, 1998; Sanders and Neville, 2000).

With respect specifically to acoustic-phonetic cues, there are different kinds of acoustic-phonetic characteristics that can serve as cues to syllable and word boundaries in a language. These include phonotactic constraints (e.g., Brent, 1997; McQueen, 1998), allophonic variation (e.g., Nakatani and Dukes, 1977; Wells, 1990), durational cues (e.g., Quené, 1993) and prosodic cues such as rhythmic patterns (e.g., Grosjean and Gee, 1987; Cutler and Norris, 1988) and, possibly, word intonation (Ramana Rao and Srichand, 1996). Rhythmic cues in particular have recently received attention (e.g., Cutler *et al.*, 1986; 1989; 1992; Cutler and Butterfield, 1992; Sebastián-Gallés *et al.*, 1992; Goetry and Kolinsky, 2000).

Some of these phonetic cues appear to play a more significant role than others in speech segmentation. Nakatani and Dukes (1977: 719) found that the strong cues for juncture in English were 'glottal stops, laryngealization, aspiration on voiceless stops ... and distinct /l/ and /r/ allophones.' Others (e.g., Christie, 1974) have also shown aspiration to be a primary cue in English speech segmentation. Nakatani and Dukes also found, as Cohen (1987) largely found for Dutch, that quantitative cues such as duration, amplitude and rate of formant transitions did not provide strong juncture cues in English. A distinction has also been made (Suomi, 1985: 212) between positive word boundary signals, 'indicating the presence of a word boundary at the location they point to,' and negative signals, 'signaling the absence of a word boundary at some point(s) of an utterance.' For example, the presence of aspiration in *keeps talking* is a positive word boundary signal; its absence in *keep stalking* is a negative word boundary signal. It is reasonable to hypothesize that positive signals may be more perceptually salient for listeners than negative signals.

Second language learners often report that the second language is spoken too quickly. This observation suggests that they may be having difficulties with speech segmentation. The fact that even highly proficient non-native speakers can be shown to have a global 'perceptual foreign accent' (McAllister, 1996) supports this. Of relevance here is

the fact that word boundary cues can vary from language to language. Thus, in Finnish, vowel harmony can provide a segmentation cue for word onsets (Suomi *et al.*, 1997), and in Skolt-Sami, also a Finno-Ugric language, durational cues do not signal boundaries (McRobbie-Utasi, 1996). Cutler *et al.* (1986) suggest that speakers of different languages may have different segmentation strategies, depending on the language's rhythmic characteristics. Spanish, unlike English, has no aspirated voiceless stops (e.g., Macpherson, 1975; Goldstein, 2001), so that the presence or absence of aspiration in voiceless stops is not a potential word boundary cue in Spanish as it is in English.<sup>1</sup> This fact of the Spanish language is of particular interest in the study described here.

There has been limited exploration of the segmentation of speech by second language learners and proficient bilinguals. Lamminmäki (1979) examined the perception of English juncture by ninth grade Finnish learners of English. The percentage of correct discrimination was 61.5 and of correct identification was 54.2, leading the author to suggest that the task was difficult for the learners. However, there is no native English speaker control group in this study and no statistical analysis of the data, making it difficult to interpret the results.

The bulk of the relevant literature has focused in particular on the role of the syllable in L2 segmentation. Broselow (1988) found that English speakers learning Arabic make perception errors involving the segmentation of speech into words and that these can be accounted for by cross-linguistic differences in prosodic organization. That is, there appears to be prosody transfer from the first language. In Cutler *et al.* (1989: 229), bilinguals proficient in both French and English participated in a syllable monitoring task. Based on this study and on earlier research (e.g., Cutler *et al.*, 1986), the authors state that different languages have different segmentation routines, specifically, that 'speakers of French process spoken words syllable by syllable, but speakers of English do not'; English speakers rely on a stress-based strategy (Cutler *et al.*, 1992).<sup>2</sup> Cutler *et al.* suggest that syllabic segmentation is marked

<sup>1</sup>Note that the term aspiration is also used to refer to the realization, in certain environments, of /s/ as [h] in some dialects of Spanish. This usage is not relevant here.

<sup>2</sup>This is due to the fact that 'French has relatively clear syllable boundaries and syllable-based timing patterns, whereas English has relatively unclear syllable boundaries and stress-based timing' (Cutler *et al.*, 1989: 229).

and that therefore French speakers can acquire unmarked, nonsyllabic routines in another language while English speakers cannot acquire the marked syllabic segmentation routines of French. That is, they suggest that the markedness of the segmentation routine of the bilingual's dominant language plays a role in determining the routines such a speaker has at his/her command in both of his/her languages. Cutler *et al.* (1992: 409) suggest that 'In some aspects of their processing, therefore, bilinguals as a consequence of this limitation may be functionally monolingual.' Golato (2002) also concludes that language dominance influences segmentation routines in both languages. Thus, these studies have led to the notion that individuals are constrained, in terms of segmentation routines, by transfer of the routines, and by the markedness of the routines of their dominant language and are therefore not necessarily able to acquire the segmentation routines of their other language.

In contrast, other studies suggest that non-native speakers are able to acquire the segmentation routines of their second language. Goetry and Kolinsky (2000: 143) conclude that 'the typical word stress pattern of the second language seems to be actually exploited in speech segmentation by bilingual listeners who have attained a high proficiency level in that language.' Sanders *et al.* (2002: 527) conclude that 'non-native speakers are able to learn new segmentation cues,' though not necessarily all segmentation cues. The evidence of Bradley *et al.* (1993: 198) suggests that segmentation strategies in the first language, at the syllabic level, change upon learning a second language. However, they caution that task-specific factors may be responsible for their results, and conclude that 'the case for language specificity in perceptual routines has yet to be made.' Thus, among the questions to be considered here is whether there is evidence that L2 segmentation routines are learnable or whether bilinguals remain, in some ways, functionally monolingual, as Cutler *et al.* suggest.

As indicated, while the use of rhythmic cues by second language learners has received some attention, there is minimal evidence regarding the abilities of non-native adult language learners to utilize cues such as allophonic distribution and/or phonotactic constraints in order to segment speech into words in their second language. The goal of this preliminary study is to discover just what the abilities of second

language learners are with regard to the use of acoustic-phonetic, but nonrhythmic, cues in speech segmentation and to explore the factors that may be playing a role in their speech segmentation strategies at the acoustic-phonetic level.

In order to explore this issue, a study was conducted with native English speakers and non-native speakers of English whose first language is Spanish, using an adaptation of the Nakatani and Dukes (1977) design. Participants hear a stimulus, for example, *keeps ticking*, and indicate whether they hear *keep sticking* or *keeps ticking*. No context is provided and word stress is the same for both choices, so that lexical, syntactic and rhythmic cues are unavailable. Speakers are forced to rely on other information, such as the presence or absence of aspiration, to segment such stimuli successfully.

There are a number of facts of Spanish and English that are relevant to this task. Borden *et al.* (2003: 124) point out, with particular reference to English, that ‘a great many durational, intonational, assimilative, and coarticulatory effects are used to establish the junctural differences between the members of such pairs as ‘nitrate’ versus ‘night rate,’ ‘it sprays’ versus ‘its praise,’ and ‘why choose’ versus ‘white shoes,’ in which the sequence of segmental phonemes is essentially identical.’ They further point out (2003: 174) that ‘internal juncture ... can be cued by a number of acoustic features, such as silence, vowel-lengthening, and the presence or absence of phonation or aspiration.’ Thus, in English, the phrase *keeps talking* can be distinguished from *keep stalking* by:

- the presence of aspiration in *talking* (e.g., Christie, 1974);
- a longer /s/ in *stalking* than in *keeps* (Klatt, 1974);
- a longer closure of the lips in *keep* than in *keeps* (Ladefoged, 1975); and
- higher amplitude of the /s/ in *stalking* than in *keeps* (Umeda and Coker, 1974).

Also relevant are the facts that glottal stop and/or creaky voice is often inserted before word-initial vowels in English (Borden, 2003), and English allows word-initial /s/ clusters.

A number of relevant facts of Spanish that could influence the ability of native speakers of Spanish to segment English input in this task are listed below. In terms of allophonic variation, Spanish has no

aspirated voiceless stops (Macpherson, 1975); Spanish can insert a glottal stop word-initially before a vowel and word-finally after a vowel in emphatic speech (Stockwell *et al.*, 1956); there are many variants of consonants word-finally in Spanish, including variants of /r/, /l/, /n/, and /s/ (Dalbor, 1969; Macpherson, 1975; Canfield, 1981; Harris, 1983; Estapa, 1989). For example, /n/ is velarized word-finally in many dialects (Cotton and Sharp, 1988); /s/ becomes a kind of /h/ word-finally in many dialects; /r/ is realized as a voiced alveolar trill word-initially and, word-finally, 'if the next word begins with a consonant but not if it begins with a vowel' (Harris, 1983: 66); and in the Cibaeño dialect of the Dominican Republic, optional liquid gliding occurs word-finally (Harris, 1983). In terms of phonotactic constraints, Spanish allows only the consonants /s, n, r, l, d/ word-finally (Goldstein, 2001); there are no word-final consonant clusters in Spanish except in a few loanwords (Dalbor, 1969); there are no word-initial /s-/ clusters (Stockwell and Bowen, 1965). Regarding syllabification, in casual speech, intervocalic consonants syllabify with the following vowel, within a word and across word boundaries (Harris, 1983; Hualde, 1991). And in terms of segments, Spanish /t/ and /d/ are dental (Harris, 1983); /f/ is realized as a bilabial fricative for most speakers of the Caribbean (Cotton and Sharp, 1988); in Colombia and Ecuador, /p, t, k/ are often voiced (Cotton and Sharp, 1988); in some Caribbean and Central American areas, [n] and [ŋ] contrast intervocalically, and [s], [h], and [x] contrast intervocalically (Cotton and Sharp, 1988).

## II Method

### 1 Participants

Background information about all the participants is provided in Table 1.

*a Native speakers of English:* Twenty undergraduate students, 13 women and 7 men, served as a native speaker control group. Participants ranged in age from 18 to 22 years, with an average age of 19.7. All considered themselves to be native speakers of English. Nineteen had studied a foreign language in school; three had also learned another language, not a Romance language, at home.

**Table 1** Biographical information, in years, native and non-native speakers

	Mean	Minimum	Maximum	SD
<i>Native speakers:</i>				
Age	19.7	18	22	1.1
<i>Non-native speakers:</i>				
Age	23.1	17	44	5.8
Age at beginning English study	17.2	7	42	7.3
Number of years of English study	5.1	1	20	3.8
Age at arrival in the USA	17.7	6	41	6.7
Length of time in the USA	5.8	0	20	3.8

*b Non-native speakers of English:* The non-native speakers were English as a second language students at community colleges or universities in the metropolitan New York area. There were 29 participants, all of whom spoke Spanish as their first language. Twenty participants were female; 9 were male. Fourteen participants were classified as intermediate learners and 15 as advanced, based on teachers' classifications of the level of class each student had been placed in. Students' placement was based on their performance on written English exams; no phonological criteria played a role. Participants' average backgrounds were as follows: their age was 23.1; their age at beginning English study was 17.2; they had studied English for 5.1 years; they had arrived in the USA at age 17.7; they had been in the USA for 5.8 years. Participants were from a number of Latin American countries, with the largest group, 17 participants, from the Dominican Republic; the next largest groups were from Mexico and Panama, with 2 participants each. (Four participants omitted at least one question on the questionnaire they were given.)

None of the participants indicated any history of hearing problems. All subjects were paid for their participation.

## 2 Materials

The stimuli are presented in Appendix 1. Of the 42 pairs of phrases, 18 consisted of phrases such as *chief's cool*, *chief school*, in which the presence or absence of aspiration provides a strong segmentation cue for native speakers of English (Nakatani and Dukes, 1977). As mentioned above, Spanish has no aspirated consonants. Eighteen pairs

consisted of phrases such as *a nice man*, *an ice man*, in which the presence or absence of glottal stop and/or creaky voice, often inserted before word-initial vowels in English (Borden *et al.*, 2003), provides a strong segmentation cue for native speakers of English (Nakatani and Dukes, 1977). Spanish can insert a glottal stop in emphatic speech. There were also six pairs for which both aspiration and glottal stop/creaky voice provide juncture cues, e.g., *like old*, *lie cold*. The three groups of stimuli will be referred to as the aspiration stimuli, the glottal stop stimuli, and the double cue stimuli, respectively. It is hypothesized, on the basis of L1 transfer, that the aspiration stimuli will be more difficult for L2 learners to segment than the glottal stop stimuli. On the assumption that two strong cues provide more relevant cues than one, it is also hypothesized that segmentation will be best for the double cue stimuli.

There were different environments within the aspiration and within the glottal stop stimuli, so that difficulty with either of these types of stimuli could not be attributed to a difficulty with any one specific environment. In addition, the distinctions between some of the environments allow for interesting comparisons in terms of markedness and of transfer of phonotactic constraints. The aspiration stimuli are broken down into three groups; these vary in terms of whether they have a vowel or a consonant before the /s/ at the juncture, and whether the /s/ is followed by one or two consonants. Note that of these groups, only the first has one choice which meets the phonotactic constraints of Spanish; in the other two groups, both choices violate Spanish phonotactic constraints. Thus, if learners rely on L1 phonotactic constraints rather than L2 acoustic cues wherever possible, segmentation on the first group should be worse than for the other two groups due to those cases in the first group where learners are misled by L1 phonotactic constraints. According to Yavaş (1998), CV is the most unmarked syllable type, and all others are marked. According to Carlisle (1998), clusters of two consonants are less marked than clusters of three consonants. If clusters are more perceptually salient as they increase in markedness, segmentation should be better on the second group than on the first, and better on the third group than the second.

When the aspiration and double cue stimuli are combined, 16 contain a potential aspiration cue with /p/, 16 with /t/ and 16 with /k/

(e.g., *keepspinning*, *keepsticking*, *keepscaring*). Of potential relevance here are the facts that /p/ and /t/ may be less marked than /k/ (Yavaş, 1998) and that /t/ has higher intensity spectral energy than /p/ and /k/ (Ferrand, 2001), which may make it more perceptually salient.

The glottal stop stimuli differ in terms of whether a nasal (/m/ or /n/), an obstruent (/f/ or /v/) or a liquid (/l/ or /r/) is the pivotal phoneme; e.g., *seeneither*, *gra(y)v(e)at*, *see(sea)lo(f)(ve)*. According to the findings of Nakatani and Dukes (1977), the allophones of /l/ and /r/ are particularly strong juncture cues for native speakers of English, while they found no such pattern for nasals or obstruents. The same subgroups are included here in order, again, to provide varied environments and for comparison with the Nakatani and Dukes results.

A different breakdown of the glottal stop stimuli is also relevant to the question of phonotactic constraints. Glottal stop stimuli with /f/ at the juncture as well as the double cue stimuli each contain one choice that meets the phonotactic constraints of Spanish (e.g., *why fill*) and one which does not meet the phonotactic constraints of Spanish (e.g., *wife ill*). (Recall that Spanish does not allow word-final /f/.) On the other hand, all the stimuli with a liquid at the juncture meet Spanish phonotactic constraints, since Spanish allows word-final as well as word-initial liquids. Again, non-native speakers should do worse on stimuli in which only one choice violates Spanish phonotactic constraints if learners rely primarily on L1 phonotactic constraints wherever possible. In addition, a comparison of glottal stop stimuli with /m/ vs. /n/ at the juncture allows one to compare a consonant that occurs word-finally in Spanish (/n/) with one that occurs in Spanish, but not in final position (/m/).

The terms positive and negative, aspiration and glottal stop will be used to designate subsets of the stimuli with regard to the strong cues of aspiration and glottal stop that researchers have found for English. Thus, stimuli in which the presence of aspiration serves as a cue for native English speakers (e.g., *keeps ticking*) will be referred to as the positive aspiration stimuli; stimuli such as *keep sticking*, in which its absence is a cue, will be referred to as the negative aspiration stimuli. It is important to remember, however, as discussed above, that other potential cues, in addition to the strong cues, are present in all of the stimuli. For example, the duration of /s/ is different in *stalking* than in *keeps* (Klatt, 1974); word-final vowels and diphthongs are longer than

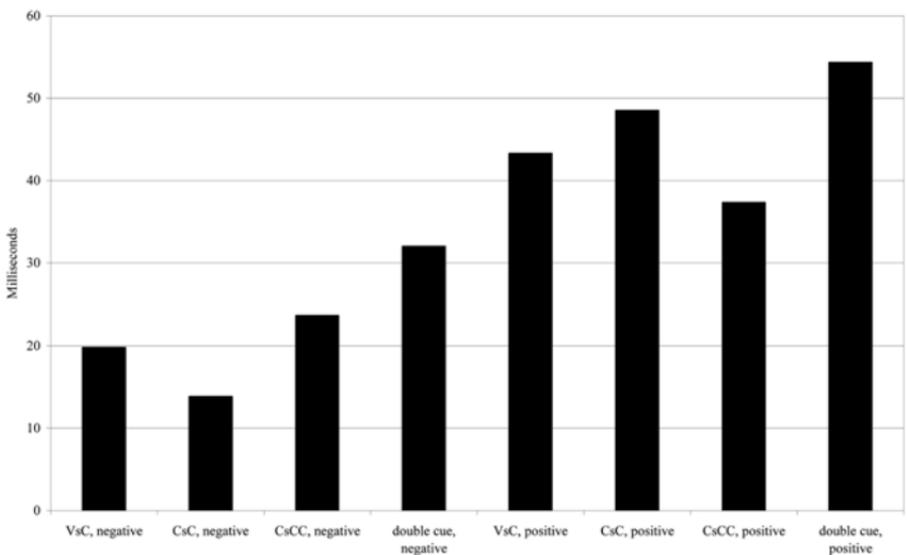
those in other positions (Ladefoged, 1975); word-final [n] is longer than word-initial [n] (Borden *et al.*, 2003). Thus, there are always, given natural stimuli, a variety of cues available.

In order to control for the naturalness of each phrase used, the phrases in the Appendix were selected from a larger set of 168 phrases (84 sets of two phrases) which had been constructed. Seven native speakers of English, who were not participants in the main experiment, participated in a pretest. They were presented with the 168 phrases and were given the following instructions: 'For each phrase, indicate its degree of naturalness by circling a number from 1 to 7, where 1 = completely natural and 7 = completely unnatural.' Of the 84 sets tested, 42 (84 phrases) were selected as stimuli, 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*). Note that the goal of the pretest was not only to maximize the naturalness of the final 84 stimuli, but also and primarily to balance the naturalness within each set. The point was to control naturalness as much as possible in order to maximize the likelihood of subjects selecting a response in the main experiment on the basis of their perception rather than on the basis of the naturalness of the phrase. The instructions of the main experiment (below) were also designed to encourage attention to the form rather than the content of the utterance.

The stimuli were recorded by an adult, male, university-educated native speaker of New York English with college radio broadcasting experience. The recording was made in a quiet room using a professional quality analogue tape recorder. While the speaker was not informed as to the specific nature of the stimuli, he was informed that non-native speakers as well as native speakers would be listening to the recording. He was instructed to speak naturally but at a slightly slower rate than usual. This instruction was given in order to increase the likelihood of observing second language knowledge of juncture cues, i.e., to minimize the likelihood of performance errors. Natural speech, rather than synthetic speech, was used since natural speech provides second language learners with all available acoustic-phonetic cues. That is, the goal was to determine how well second language speakers do with the odds stacked in their favour.

The stimuli were input into the Computerized Speech Lab and digitized at 10 000 Hz, and the spectrographic analysis program of Kay Elemetrics Multi-Voice Program was used to analyse them. The duration of the release burst was measured for the positive aspiration stimuli (e.g., *chief's cool*) along with the relevant double cue stimuli (e.g., *grey pin*) and compared to the duration of the release burst of the negative aspiration stimuli (e.g., *chief school*) along with the relevant double cue stimuli (e.g., *grape in*). The mean duration of the release burst for the stimuli expected to have aspiration was 46 msec (SD = 16.0) for the negative stimuli, it was 22 msec (SD = 10.77). This difference was significant at  $p < .001$  ( $t(23)=5.879$ ). Figure 1 indicates the mean duration of the release burst for each of the subgroups of the aspiration and relevant double cue stimuli.

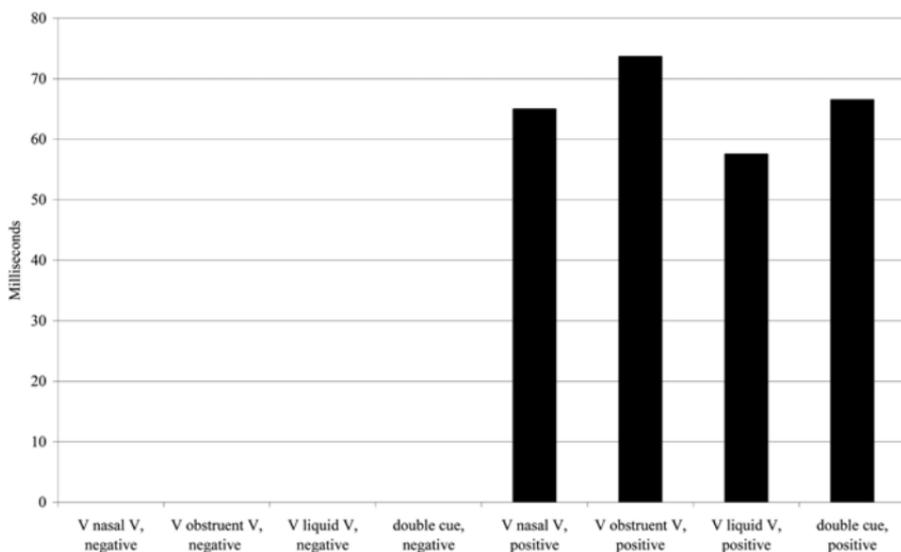
The acoustic analysis of the glottal stop stimuli indicates that the positive glottal stop stimuli (e.g., *seen either*), including the relevant double cue stimuli (e.g., *grape in*), are characterized by a silent gap or period of creaky voice before the vowel at the juncture. (The majority of the stimuli were characterized by the silent gap.) Creaky voice was defined as an audible creaky voice along with, acoustically, random noise spread throughout the frequency spectrum but more intense



**Figure 1** Release burst, aspiration and double cue stimuli

between 500 and 2000 Hz. The average duration of the silent gap/creaky voice was 77 msec (SD = 34.27). The negative glottal stop stimuli (e.g., *see neither*) are all characterized by the absence of a silent gap or creaky voice at the juncture, with one surprising exception: the three stimuli beginning with /v/ (*grey vat*, *we vend* and *say vamps*) were each characterized by silence before the /v/, of an average duration of 65 msec. This will be addressed further below. Figure 2 indicates the mean duration of the period of silence or creaky voice for each of the subgroups of the glottal stop and relevant double cue stimuli, excluding the stimuli with /v/.

The 84 stimuli (42 pairs of phrases) were recorded in a random order. Each phrase was placed within the carrier phrase, 'Say \_\_\_\_\_ again,' which has frequently been used in controlled speech studies in order to maximize natural speech (e.g., Klatt, 1974; Stockman and Pluut, 1999). This particular frame is useful because it is one in which any class of word can be inserted. Thus, participants heard, for example, 'Say chief's cool again.' Each stimulus was separated from the other member of its pair by at least 15 other stimuli, with an average of 41 items separating the members of each pair. Two stimuli from the same



**Figure 2** Period of silence or creaky voice, glottal and double cue stimuli

subgroup in the Appendix never followed one another. The correct response was the first choice on the answer sheet half of the time.

### 3 Procedure

Participants were given the following instructions, both orally and in writing: 'You will hear two or three words. The words will always be in the sentence, "Say \_\_\_\_\_ again." Decide which words you have heard, and, on your answer sheet, circle either choice A or choice B.' Both members of each pair were written down on an answer sheet. Subjects did two practice examples, and the instructions continued, indicating that participants might sometimes hear words they did not know but that they were not to worry about this; the important thing was for them to listen carefully. The goal of these instructions was to encourage subjects to attend to form rather than content. Subjects heard each phrase once and had 6 seconds between stimuli, determined by pre-testing as a comfortable amount of time in which to respond. They were tested in small groups in a quiet room. The stimuli were played on a professional quality analogue tape recorder and the volume was adjusted until each participant indicated that he/she could hear comfortably. After participating in the experiment, subjects completed a language background questionnaire.

## III Results

The means and standard deviations for each stimulus are indicated in Appendix 1. As mentioned above, stimuli with word-initial /v/ were, unexpectedly, characterized by a period of silence at the juncture. On the assumption that there may have been something unusual about the speaker's pronunciation of words with /v/ at the juncture, and given that they did not share the acoustic characteristics of the other stimuli in this group, the six stimuli with /v/ at the juncture were removed from all subsequent analyses, leaving a total of 78 stimuli.

A comparison of second language learners placed in intermediate vs. advanced classes indicates no significant difference between the groups in terms of overall percentage of correct responses:  $t(27) = 1.14$ ,  $p > 1.0$ . The mean was 74.1 for the intermediate group and 76.3 for the

advanced group. Consequently, scores for the intermediate and advanced learners are combined into one group of non-native speakers in the analyses which follow.

The mean percentages correct for native and non-native speakers on the word boundary perception task are indicated in Table 2.<sup>3</sup> The results here indicate that there was a significant effect of participant group:  $t(39.83) = 20.051, p < .001$ . That is, native speakers had significantly more correct responses than non-native speakers on the task. An analysis of variance was also conducted. There was a significant main effect of stimulus type:  $F(1.989, 93.506) = 138.808, p < .001$  and a significant interaction of stimulus type and group:  $F(1.989, 93.506) = 110.135, p < .001$ . Post-hoc analysis indicates that, for the native speakers, there was no significant difference among the three stimulus groups. For the non-native speakers, the mean percentage correct for the glottal stop and double cue stimuli were significantly better than for the aspiration stimuli; there was no significant difference between the glottal stop and double cue stimuli.

Although the non-native speakers' mean percentage correct for the aspiration stimuli is 58.5%, close to the 50% one would expect by

**Table 2** Native and non-native speakers, mean percentage of correct responses for each stimulus type

Stimulus type	Native speakers ( $n = 20$ )		Non-native speakers ( $n = 29$ )	
	Mean	SD	Mean	SD
Aspiration ( $n = 36$ ) (e.g., <i>Lou stops/loose tops</i> )	96.7	3.4	58.5	8.3
Glottal stop ( $n = 30$ ) (e.g., <i>a niche/an itch</i> )	97.0	3.2	88.4	7.0
Double cue ( $n = 12$ ) (e.g., <i>grey pin/grape in</i> )	99.6	1.9	92.5	7.5
All stimuli ( $n = 78$ )	97.3	2.0	76.3	4.8

<sup>3</sup>Levene's Test was employed in order to examine the differences in the standard deviations of the native speakers and non-native speakers. Its results indicate that there was significantly more variation in the scores of the non-native speakers than the native speakers ( $F = 87.292, p = .006$ ). In order to take this difference in variation into account, all between-group comparisons with significant Levene's Test results have been conducted using  $t$ -tests with the Welch correction. Further, given the unequal numbers in some of the stimulus groups and the effect that this can have on variability, the Box Test of Equality of Covariance Matrices was employed wherever applicable to determine the equality of the observed covariance matrices of the dependent variables. Where this is significant, analysis of variance results are reported using the Huynh-Feldt correction. Where Box's Test is inapplicable,  $t$ -tests are used.

chance (since participants had only two choices), the 58.5% is significantly different from 50%, suggesting that non-native speakers were still doing more than just guessing:  $t(28) = 5.510, p < .001$ .

Analyses were also conducted comparing native and non-native speakers' responses to positive stimuli, i.e., stimuli in which the strong cue of aspiration or glottal stop/creaky voice was present (e.g., *keeps parking, seen either*), with their responses to negative stimuli, i.e., stimuli in which the strong cue was absent (e.g., *keep sparking, see neither*). These results are presented in Table 3. The difference between the two subject groups was again significant:  $t(39.783) = 20.32, p < .001$ . All subjects were more accurate in identifying stimuli with the cue present than stimuli with the cue absent:  $t(48) = 6.079, p < .001$ . For native speakers, there was no significant difference, overall, between positive and negative stimuli:  $t(39) = 1.144, p > .1$ . A breakdown of stimulus types indicates that, for native speakers, negative aspiration stimuli were identified more accurately than positive aspiration stimuli:  $t(19) = 3.040, p = .007$ , while positive glottal stop stimuli were identified more accurately than negative glottal stop stimuli:  $t(19) = 4.158, p = .001$ . For non-native speakers, positive aspiration stimuli were identified more accurately than negative aspiration stimuli:  $t(28) = 5.746, p < .001$ , and positive glottal stop stimuli were identified more accurately than negative glottal stop stimuli:  $t(28) = 8.751, p < .001$ .

Further analyses of the various subgroups of each type of stimulus reveal the following results. For each of these analyses, the difference between the native and non-native speakers, overall, was always significant. An analysis of variance comparing all subjects' responses to the

**Table 3** Native and non-native speakers, mean percentage of correct responses, positive vs. negative stimuli

Stimulus type	Native speakers ( $n = 20$ )		Non-native speakers ( $n = 29$ )	
	Mean	SD	Mean	SD
Positive aspiration (e.g., <i>chief's cool</i> ) ( $n = 18$ )	95.0	5.5	73.0	15.8
Positive glottal stop (e.g., <i>an itch</i> ) ( $n = 15$ )	100.0	0.0	99.1	2.3
Total positive ( $n = 33$ )	97.3	2.8	84.8	8.5
Negative aspiration (e.g., <i>chief school</i> ) ( $n = 18$ )	98.3	3.2	44.1	16.1
Negative glottal stop (e.g., <i>a niche</i> ) ( $n = 15$ )	94.0	6.5	77.0	13.8
Total negative ( $n = 33$ )	96.4	3.6	59.0	11.5

**Table 4** Native and non-native speakers, mean percentage of correct responses, aspiration stimuli, /p/ vs. /t/ vs. /k/ at juncture

Phoneme	Native speakers ( <i>n</i> = 20)		Non-native speakers ( <i>n</i> = 29)	
	Mean	SD	Mean	SD
/p/ ( <i>n</i> = 12) (e.g., <i>Lou spills/loose pills</i> )	95.6	4.6	64.0	10.3
/t/ ( <i>n</i> = 12) (e.g., <i>Lou stops/loose tops</i> )	99.1	2.3	72.4	10.2
/k/ ( <i>n</i> = 12) (e.g., <i>Lou skis/loose keys</i> )	97.5	4.7	65.1	10.1

**Table 5** Non-native and native speakers, mean percentage of correct responses, aspiration stimuli formats

Format	Native speakers ( <i>n</i> = 20)		Non-native speakers ( <i>n</i> = 29)	
	Mean	SD	Mean	SD
VsC ( <i>n</i> = 12) (e.g., <i>lay stable, lace table</i> )	97.1	4.9	54.0	12.7
CsC ( <i>n</i> = 12) (e.g., <i>keep sparking, keeps parking</i> )	97.5	4.8	57.8	13.2
CsCC ( <i>n</i> = 12) (e.g., <i>top scrawled, tops crawled</i> )	95.4	5.7	63.8	13.0

aspiration stimuli with juncture at the /p/ vs. /t/ vs. /k/ revealed a significant effect of stop phoneme:  $F(2,94) = 7.556$ ,  $p = .001$  and no interaction of subject group by stop phoneme:  $F(2,94) = 1.940$ ,  $p = .149$  (see Table 4). Post-hoc analysis reveals a significant difference between stimuli with /t/ at the juncture and stimuli with /p/ and /k/; there was no significant difference between /p/ and /k/ stimuli. That is, juncture with /t/ was easier to perceive correctly than juncture with /p/ or /k/. While the scores were in this direction for all participants, the difference was significant only for the non-native speakers.

Another analysis of variance was conducted to examine the three formats of the aspiration stimuli, as they are grouped in the Appendix (see Table 5). The overall difference among the three types of aspiration formats was not significant:  $F(2,94) = 2.00$ ,  $p = .141$ , although the interaction between subject group and stimulus type was significant:  $F(2,94) = 4.213$ ,  $p = .018$ . Post-hoc analysis indicates that, for non-native participants, the mean difference between the VsC subgroup and the CsCC subgroup was significant, with the latter perceived correctly more often; all other comparisons were not significant.

**Table 6** Native and non-native speakers, mean percentage of correct responses, glottal stop stimuli formats

Format	Native speakers ( $n = 20$ )		Non-native speakers ( $n = 29$ )	
	Mean	SD	Mean	SD
nasal ( $n = 12$ ) (e.g., <i>see neither, seen either</i> )	93.8	7.6	79.6	13.5
obstruent ( $n = 6$ ) (e.g., <i>grey vat, grave at</i> )	95.0	5.7	84.8	7.1
liquid ( $n = 12$ ) (e.g., <i>I learn, I'll earn</i> )	99.6	1.9	92.8	7.9

A repeated measures analysis of variance indicates a significant effect of type of glottal stop stimulus (nasal, obstruent, or liquid at the juncture):  $F(1.65, 77.37) = 17.102$ ,  $p < .001$  (see Table 6). The interaction of stimulus type and subject group was also significant:  $F(1.65, 77.37) = 4.038$ ,  $p = .028$ . *Post-hoc* analysis indicates, for non-native speakers only, a significant difference between those stimuli with a nasal at the juncture and the other two stimulus groups. That is, learners' responses were worst with a nasal at the juncture.

A comparison of the glottal stop stimuli with /f/ combined with the double cue stimuli to the glottal stop stimuli with a liquid at the juncture was conducted. Recall that the former have only one choice that meets the phonotactic constraints of Spanish, while for the latter, both choices meet Spanish phonotactic constraints. The results indicate no significant effect of stimulus type:  $F(1, 47) = .001$ ,  $p > .1$ , and no interaction of stimulus type and subject group:  $F(1, 47) = .184$ ,  $p > .1$ . That is, the status of Spanish phonotactics did not appear to play a role in subjects' responses to these stimuli. Finally, stimuli with word-final /n/ were compared to stimuli with word-final /m/ for non-native speakers; only /n/ is allowed word-finally in Spanish, although it may be realized phonetically as [ŋ]. Subjects had the same mean score on positive /m/ and positive /n/ stimuli (99% correct on each); they scored better on negative /m/ (63.3 %) than negative /n/ (57.3%) stimuli, but this difference was not significant:  $t(28) = .740$ ,  $p > .1$ . In these comparisons, Spanish phonotactics again did not appear to be playing a role.

A few stimuli contained a function word in one choice only. A comparison of glottal stop and double cue stimuli with vs. without only one member of each pair containing a function word was conducted.

(No aspiration stimuli contained function words.) There was no significant difference in the percentage correct for each type of stimulus:  $t(48) = 1.560, p > .1$ , over all subjects.

There were no significant correlations, for non-native speakers, between their overall percentage of correct responses and:

- their age at beginning English study ( $r = -.113, p = .59$ );
- the number of years of English study ( $r = .036, p = .852$ );
- their age of arrival in the USA ( $r = -.015, p = .940$ ); and
- the number of years they had been in the USA ( $r = .084, p = .669$ ).

#### **IV Discussion**

A central finding of the study is that second language learners, even those placed in intermediate or advanced classes, have difficulty segmenting speech in their second language on the basis of acoustic-phonetic cues, segmenting correctly only 76% of the time, compared to 96% for native speakers. This is the case even though the participants' task was simplified by the availability of both choices on the answer sheet as they listened to the stimulus. Thus, second language learners are unable to fully use the kinds of acoustic-phonetic cues available to native English speakers. If speakers of a language need to use acoustic-phonetic information to segment speech and retrieve words (probably in conjunction with higher level information), as a number of researchers suggest (e.g., Quené, 1992; Gow and Gordon, 1995), then the second language learner is missing information crucial to the process of lexical retrieval.

Another central finding of this study is that some stimuli are easier than others for participants to segment. The reasons for this greater ease or difficulty are not always clear, although the data suggest that various factors may be involved. For the non-native speakers, stimuli in which the presence or absence of glottal stop/creaky voice is a strong juncture cue for native English speakers were easier to segment than stimuli for which the presence or absence of aspiration is a strong cue. Transfer may be responsible for the non-native speakers' low scores on the aspiration stimuli, since a glottal stop occurs in emphatic speech in Spanish while aspiration does not occur in Spanish. Another possibility is that glottal stop insertion is less marked and hence easier to acquire than

aspiration. According to Borden *et al.* (2003: 75), 'Some speakers initiate phonation with what is called a glottal attack.' Summarizing the responses to a query on the *Linguist List* (1996) regarding obligatory onsets, Ratcliffe notes that utterance initial vowels in English must be preceded by a glottal stop, which may be due in some cases to 'a universal phonetic, i.e., physiological effect.' Further, the glottal stop 'seems to be the universal phonetic default.' The reference here must be only to utterance initial vowels, since at internal word boundaries, e.g., *seem able*, it is common, but not required, to insert a glottal stop before the vowel in English. Nevertheless, these factors suggest that inserting a glottal stop may well be a relatively unmarked phenomenon. In terms of the results of this experiment, then, transfer, markedness, or a combination of both may be contributing to the learners' greater success rate with the glottal stop stimuli than the aspiration stimuli.

For both groups, there was no advantage of the double cue stimuli over the glottal stop stimuli. With regard to the native speakers, it is possible that a more challenging task (e.g., one with noise) might show an advantage of the double cue stimuli. With regard to the non-native speakers, the fact that there was no advantage of the double cue stimuli over the glottal stop stimuli can support the finding that non-native speakers were limited in their ability to use aspiration as a cue: if they were not able to effectively use aspiration, there is no reason for the double cue stimuli (with both aspiration and glottal stop as potential cues) to be better than the glottal stop stimuli.

Non-native participants were also better at segmenting stimuli with a positive strong cue than a negative one, while for native speakers the results were mixed. Various researchers have found (Nakatani and Dukes, 1977; Gow and Gordon, 1995) that more salient cues are word-initial rather than word-final; the positive cues here were all word-initial. Thus, learners may be better at segmenting the positive stimuli for at least two reasons: cues at the beginning of a word are more salient than cues at the end; the presence of a cue is more perceptually salient than its absence. In either case, the evidence suggests that perceptual salience is a relevant factor in this task. Perceptual salience may also account for the finding that aspiration stimuli with /t/ were easiest for non-native participants to segment, since /t/ and /d/ have higher intensity spectral energy than the other stop consonants.

There are also differences among the subgroups of the glottal stop. Those stimuli in which a liquid (/l/ or /r/) or obstruent (/f/) can potentially be on either side of the word boundary were easier for second language learners to segment than those with a nasal. Recall that, according to Nakatani and Dukes, the allophones of /l/ and /r/ are particularly strong juncture cues for native speakers of English. Related to this is Brown's (1998) finding that Japanese L2 learners of English were able to discriminate /l/ and /r/ only in coda position. She attributes this to the fact that liquid codas have a strong effect on the preceding vowel. The findings here indicate that liquids provide strong cues for non-native speakers in this task as well, although no stronger than /f/. The fact that nasals are weak intensity sounds (Ferrand, 2001) may also be a factor in the lower success rate for those stimuli with a nasal, as opposed to /f/ or a liquid, at the juncture.

The question of transfer from Spanish can be further explored by examining subgroups of the data that do or do not conform to Spanish phonotactic constraints. The results here are mixed. The two relevant comparisons of glottal stop stimuli (the glottal stop and double cue stimuli that violate Spanish phonotactics on one choice vs. glottal stop stimuli that do not violate Spanish phonotactics, and /m/ vs. /n/ at the juncture) reveal no difference between the stimuli groups and, hence, no apparent effect of Spanish phonotactics on non-native speakers' responses on this task.

In contrast, non-native speakers performed more accurately on the CsCC aspiration stimuli, in which both choices violated Spanish phonotactics, than the VsC stimuli, in which one choice violated Spanish phonotactics. This finding suggests that the non-native speakers may have been incorrectly relying on first language phonotactics to segment where that was an option. However, those stimuli that violate Spanish phonotactics contain a consonant cluster, and are hence also more marked than those that do not violate Spanish phonotactics, making them potentially more salient. Thus, two conclusions are possible here. One is that non-native speakers relied on L1 phonotactics only with those stimuli, the aspiration stimuli, about which they were most uncertain and not with the glottal stop stimuli. The other is that the difference found is due, not to transfer, but to the presence of a more marked, and hence perhaps perceptually more salient, consonant

cluster. Either interpretation, however, leaves unexplained the behaviour of the CsC group of stimuli. These should pattern with the CsCC group, but not the VsC group, in the case of transfer; they should be significantly different from both other groups in the case of markedness. However, the CsC group was not significantly different from either of the other two, although the data is in the direction hypothesized if increasing markedness results in increasing perceptual salience. Thus, the one comparison out of three that suggests that first language phonotactics may be playing a role can be accounted for as well (or as poorly) by markedness, so that the data supporting the role of L1 syllable structure in this task is not compelling.

This tentative conclusion is strengthened when the stimuli are examined from the perspective of Spanish syllabification. In Spanish, in casual speech, intervocalic consonants syllabify with the following vowel, within a word and across word boundaries. If Spanish syllabification is transferred to English in this task, and assuming that the stimuli are perceived as casual speech, then one would expect the non-native speakers to prefer the segmentation of, for example, *a niche* over *an itch*. They should therefore have higher scores on the negative glottal stop stimuli, in which the pivotal consonant is part of the following word, than on the positive glottal stop stimuli; however, the opposite is, in fact, the case. Spanish syllabification would also predict, for the double cue stimuli, that stimuli such as *grey pin* would be favoured as a choice over those such as *grape in*, leading to a greater error rate for the latter. In fact, while there was no significant difference between the two types of double cue stimuli for the non-native speakers, the difference was in the opposite direction: non-native speakers had a lower percentage correct on the former than on the latter ( $t(28) = 1.314$ ,  $p > .1$ ). These data thus suggest that Spanish syllable structure was not playing a role in subjects' responses (and further support the claim that the non-native speakers have difficulty utilizing aspiration as a cue).

The finding appears to conflict with that of Broselow (1988: 301), who attributes the misperceptions of Arabic word boundary made by native speakers of English to the differences between English and Arabic in terms of 'the inventory of syllable types and the principles for organizing syllables into words.' However, her analysis rests on the assumption that 'the phonetic cues indicating a segment's position in

a syllable are roughly equivalent in English and Arabic'; this assumption may not be true. Further, the misparsings Broselow reports are those commonly made by English speakers learning Arabic, presumably when they are listening to Arabic. The task is different from that in this study, in which individuals are given a choice of two possible L2 strings, each of which they know to be grammatical in the second language. Given that participants know that both parsings are grammatical and that therefore phonotactics (neither L1 nor L2) cannot provide directly useful information, the task may encourage participants to listen for acoustic-phonetic cues. Thus, the nature of the task here may account for the sparse evidence supporting transfer from L1 to L2 at the syllabic level.

The issue of word frequency needs to be considered in examining these results: when in doubt, were L2 learners more likely to select high frequency words than low frequency words? While the naturalness of each phrase was controlled, the possibility remains that the frequency of the individual words comprising each phrase played a role in second language learners' responses, despite the instructions encouraging them not to pay attention to the words' meaning. The frequency of the individual content words was therefore tabulated. The 16 instances of function words, out of a total of 158 different stimulus words, were omitted because their often extremely high frequency would present a distorted picture of the average word frequencies. (For example, the relatively high frequency of *see* is 772; that of *in* is 21,341.) Proper noun stimuli were omitted from the tabulation and any other word not appearing in the database was assigned a frequency of zero. The frequencies of 87.3% of the stimuli were thus calculated, using the Kučera and Francis (1967) frequency count at the MRC Psycholinguistic Database ([http://www.psy.uwa.edu.au/mrcdatabase/uwa\\_mrc.htm](http://www.psy.uwa.edu.au/mrcdatabase/uwa_mrc.htm)). The frequency of the base form was always used, for example, the frequency of *talk* rather than of *talking*.<sup>4</sup>

The overall average frequencies for the aspiration, glottal stop and double cue stimuli were 97.1, 436.6 and 267.7, respectively. If frequency were playing a primary role in participants' responses, then one

<sup>4</sup>It should be noted that the relation of these frequencies to the experiences of the non-native speakers may not be as close as for the native speakers.

would expect the highest percentage correct for second language learners on the glottal stop stimuli. While the difference in percentage correct between the glottal stop and double cue stimuli for the second language learners was not significant, it was in the opposite direction from that predicted by word frequency, with responses to the double cue stimuli higher. Furthermore, the average frequency for the positive glottal stop stimuli was 145, that of the negative glottal stop stimuli was 728.3. The average correct responses to these stimuli were 99.1% and 77%, in the opposite direction of what one would expect if word frequency were playing a role in non-native participants' responses. As for the aspiration stimuli, those about which participants were apparently most uncertain, the average frequency of the positive aspiration stimuli was 97.2, that of the negative aspiration stimuli was 96.8, while their correct responses were 73% and 44.1% respectively, so that again the possibility of word frequency playing a role in second language participants' responses to these stimuli is unlikely. Thus, individual word frequency does not appear to play a significant role in determining which group of stimuli was easiest for learners to segment.

The roles of other measures of frequency, in addition to word frequency, need to be considered as well. This includes the relative frequency of sounds in different positions, e.g., the frequency of /f/ word-initially vs. word-finally. Using data provided by Shriberg and Kent (2003) (adapted from Mines *et al.*, 1978), a number of facts are relevant here. The phoneme /s/, the pivotal phoneme in the aspiration stimuli, occurs 37.1% of the time word-initially and 30.8% of the time word-finally. However, the aspiration stimuli in which /s/ occurs word-finally (i.e., the positive aspiration stimuli) were identified more accurately than those in which /s/ occurs word-initially (i.e., the negative aspiration stimuli), in the opposite direction of what one would expect if relative frequency were playing a role here. Similarly, according to Shriberg and Kent, /l/ is as frequent word-initially (23.72% of the time) as word-finally (24.31%), yet stimuli with word-final /l/ were identified accurately by the non-native speakers significantly more often than stimuli with word-initial /l/ (99% and 85%, respectively). And /f/ is far more frequent in word-initial than final position (63.15% vs. 12.87%, respectively), yet was identified correctly by the non-native participants significantly more often in final position than in initial position

(100% and 94.33%, respectively). Thus, it appears that the relative frequency of the pivotal phoneme is also not likely to be playing a role in subjects' decisions.

The situation is complicated by the fact that two words are involved in every choice. Thus, for the pair *see neither/seen either*, the relative frequency of word-final /i/, as in *see*, is greater (60.01%) than its frequency word-initially, as in *either* (2.56%); subjects should prefer *see neither* on this basis. However, the relative frequency of word-final /n/, as in *seen*, is greater (46.48%) than its frequency word-initially, as in *neither* (12.42%); subjects should prefer *seen either* on this basis. Under these circumstances, it is difficult to evaluate the relative weight of segment frequency in each position for each phrase.

Frequency can be measured from other perspectives as well, including the relative frequency of sound classes in different positions (e.g., the relative frequency of vowels word-initially vs. word-finally) and probabilistic frequency (e.g., Kessler and Treiman, 1997). While the data provided suggest that word frequency and the frequency of the pivotal phoneme in different positions is not likely to be playing a role in subjects' responses, the role of other measures of frequency cannot be ruled out. Further, while the instructions encouraged participants to pay attention to acoustic-phonetic cues, it is possible that participants may have not known some of the words used, and that this influenced their choices, at least in some cases. Future research can examine the role of lexical knowledge in this task by, for example, comparing participants' responses to stimuli where both choices contain words to stimuli where one choice contains an appropriately constructed nonword.

An issue raised earlier is the question of whether or not second language segmentation routines are fully attainable. Recall that while some researchers, working with rhythmic cues, have suggested that individuals use the same segmentation strategies in each of their languages, others state that individuals are able to acquire new rules, appropriate for their second language. The fact that learners scored above chance on the aspiration stimuli, apparently utilizing a rule that does not exist in Spanish and that relies on perceiving phones that do not exist in Spanish, suggests that some relevant learning has taken place, although without a monolingual Spanish subject group, one cannot be certain

that individuals with no knowledge of English would not perform similarly. Note that the above chance difference is due to learners' perception of the positive aspiration stimuli (73% correct), since responses by non-native speakers to the negative aspiration stimuli (44.1% correct) were below chance. Nonetheless, while some learning may have taken place, it is clear that in this perception task non-native speakers are far from achieving native proficiency. Other researchers (e.g., Brown, 1998) have also found that the attainment of native-like proficiency in perception is limited. Future research with more advanced non-native speakers will be relevant to the issue of whether native-like proficiency in this domain is attainable.

Hohne and Jusczyk (1994: 614), in discussing young infants' abilities to detect word boundaries, note that a number of factors are needed for infants to use allophonic variation as a word boundary cue. These include the fact that (1) the relevant differences need to be perceived, (2) their distribution needs to be known and (3) there must be 'some capacity to use this distributional information during fluent speech perception.' The study here has begun the investigation of second language learners' abilities to use nonrhythmic acoustic-phonetic cues with results that suggest that adult learners do not use these cues in their second language as effectively as native speakers. It remains to be determined which of the factors, or which combination of the factors outlined by Hohne and Jusczyk, is specifically responsible for their difficulties. However, the fact that learners scored above chance on the positive aspiration stimuli suggests that there may be a knowledge of the relevant distributional information and that the difficulty lies with the capacity to use this information.

There was no difference between learners in intermediate vs. advanced classes and no correlation found between variables such as age, number of years of English study, etc., and scores on the segmentation task. However, only 4 of the 29 non-native speakers began learning English before the age of 13; only 4 had lived in the USA for more than 9 years. Thus, research with wider ranges of the relevant variables is needed before the role of language experience in phonetic speech segmentation can be fully evaluated. Nonetheless, it is interesting to note that Brown (1998), in her investigation of the perception of L2 segment structure, similarly found no significant correlation between

subjects' performance and variables such as number of years of English study and number of years in North America.

Natural stimuli were used in this preliminary exploration of learners' ability to use nonrhythmic cues to segment L2 speech because natural stimuli provide information about what learners can and cannot do with all available cues at their disposal. However, one of the consequences of using natural stimuli is that it is not possible to isolate with certainty all the factors that may be playing a role in learners' perception of the stimuli. Future research with synthetic stimuli can be directed towards examining those factors, transfer and/or markedness, positive and/or word-initial cues, which could not be teased apart here. Future research should ideally include a monolingual Spanish subject group, in order to identify any cues which are relevant to native Spanish speakers with no knowledge of English. Any deviation by English learners from the response pattern of those with no knowledge of English is likely to be due to the process of acquiring English. The results here also suggest that further exploration of the role of L1 syllable structure constraints in L2 speech perception is warranted.

Word segmentation is a potentially fruitful area in which to explore issues, such as transfer and markedness, which have received far more attention in second language production and second language phoneme perception. The experimental paradigm used here holds promise for facilitating this exploration.

### *Acknowledgements*

Special thanks to Mark Osipow for his assistance in preparing the tape-recording, to Lisa Ross for her wonderful job as research assistant, to Paul Camhi and Cindy Greenberg for their assistance in obtaining participants, to Michael Barnes for help with the statistics and to Carole T. Ferrand for her invaluable help with the acoustic analysis of the stimuli. I also greatly appreciate the insightful comments and suggestions of Ronald L. Bloom, Helen Smith Cairns, James J. Drieves, Carolyn Sobel and Robert M. Vago, as well as those of three anonymous *Second Language Research* reviewers. I also appreciate the support provided to this study by a Hofstra University Faculty Development Grant.

## V References

- Altenberg, E.P.** 2002: The acquisition of syllable onsets in a second language: judgments, perception, and production. Paper presented at the Annual International Linguistics Association Conference, Toronto.
- Best, C.T.** 1995: A direct realist view of cross-language speech perception. In Strange, W., editor, *Speech perception and linguistic experience*. Timonium, MD: York Press, 171–204.
- Bohn, O-S.** 1995: Cross-language speech perception in adults: first language transfer doesn't tell it all. In Strange, W., editor, *Speech perception and linguistic experience*. Timonium, MD: York Press, 273–304.
- Borden, G.J., Harris, K.S. and Raphael, L.J.** 2003: *Speech science primer*. 4th edition. Philadelphia, PA: Lippincott, Williams and Wilkins.
- Bradley, D., Sánchez-Casas, R. and García-Albea, J.** 1993: The status of the syllable in the perception of Spanish and English. *Language and Cognitive Processes* 8, 197–233.
- Brent, M.R.** 1997: Toward a unified model of lexical acquisition and lexical access. *Journal of Psycholinguistic Research* 26, 363–75.
- Brent, M.R. and Cartwright, T.A.** 1996: Distributional regularity and phonotactic constraints are useful for segmentation. *Cognition* 61, 93–125.
- Broselow, E.** 1988: Prosodic phonology and the acquisition of a second language. In Flynn, S. and O'Neill, W., editor, *Linguistic theory in second language acquisition*. Dordrecht: Kluwer, 295–308.
- Broselow, E. and Finer, D.** 1991: Parameter setting in second language phonology and syntax. *Second Language Research* 7, 35–59.
- Brown, C.A.** 1998: The role of the L1 grammar in the L2 acquisition of segmental structure. *Second Language Research* 14, 136–93.
- Canfield, D.L.** 1981: *Spanish pronunciation in the Americas*. Chicago, IL: University of Chicago Press.
- Carlisle, R.S.** 1998: The acquisition of onsets in a markedness relationship. *Studies in Second Language Acquisition* 20, 245–60.
- Christie, W.M.** 1974: Some cues for syllable juncture perception in English. *Journal of the Acoustical Society of America* 55, 819–21.
- Church, K.** 1987: Phonological parsing and lexical retrieval. *Cognition* 25, 53–69.
- Cohen, A.** 1987: Juncture revisited. In Chanon, R. and Shockey, L., editors, *In honor of Ilse Lehiste*. Dordrecht: Foris Publications, 7–18.
- Cotton, E. and Sharp, J.** 1988: *Spanish in the Americas*. Washington, DC: Georgetown University Press.
- Cutler, A. and Butterfield, S.** 1992: Rhythmic cues to speech segmentation: evidence from juncture misperception. *Journal of Memory and Language* 31, 218–36.
- Cutler, A., Mehler, J., Norris, D. and Segui, J.** 1986: The syllables' differing role in the segmentation of French and English. *Journal of Memory and Language* 25, 385–400.

- 1989: Limits on bilingualism. *Nature* 340, 229–30.
- 1992: The monolingual nature of speech segmentation by bilinguals. *Cognitive Psychology* 24, 381–410.
- Cutler, A. and Norris, D.** 1988: The role of strong syllables in segmentation for lexical access. *Journal of Experimental Psychology: Human Perception and Performance* 14, 113–21.
- Dalbor, J.B.** 1969: *Spanish pronunciation: theory and practice*. New York: Holt, Rinehart and Winston.
- Dupoux, E., Kakehi, K., Hirose, Y., Pallier, C. and Mehler, J.** 1999: Epenthetic vowels in Japanese: a perceptual illusion? *Journal of Experimental Psychology: Human Perception and Performance* 25, 1568–78.
- Eckman, F.R. and Iverson, G.K.** 1993: Sonority and markedness among onset clusters in the interlanguage of ESL learners. *Second Language Research* 9, 234–52.
- Estapa, R.** 1989: Internal open juncture in Spanish [La juntura interna abierta en español]. *Verba* 16, 97–116.
- Ferrand, C.T.** 2001: *Speech science: an integrated approach to theory and clinical practice*. Boston: Allyn and Bacon.
- Flege, J.E.** 1995: Second language speech learning: theory, findings, and problems. In Strange, W., editor, *Speech perception and linguistic experience*. Timonium, MD: York Press, 233–77.
- Goetry, V. and Kolinsky, R.** 2000: The role of rhythmic cues for speech segmentation in monolingual and bilingual listeners. *Psychologica Belgica* 40, 115–52.
- Golato, P.** 2002: Word parsing by late-learning French-English bilinguals. *Applied Psycholinguistics* 23, 417–46.
- Goldstein, B.** 2001: Transcription of Spanish and Spanish-influenced English. *Communication Disorder Quarterly* 23, 54–60.
- Gow, D.W. and Gordon, P.C.** 1995: Lexical and prelexical influences on word segmentation: evidence from priming. *Journal of Experimental Psychology: Human Perception and Performance* 21, 344–59.
- Grosjean, F. and Gee, J.P.** 1987: Prosodic structure and spoken word recognition. *Cognition* 25, 135–55.
- Hallé, P., Segui, J., Frauenfelder, U. and Meunier, C.** 1998: Processing of illegal consonant clusters: a case of perceptual assimilation? *Journal of Experimental Psychology: Human Perception and Performance* 24, 592–608.
- Harris, J.W.** 1983: *Syllable structure and stress in Spanish*. Cambridge, MA: The MIT Press.
- Hohne, E.A. and Jusczyk, P.W.** 1994: Two-month-old infants' sensitivity to allophonic differences. *Perception and Psychophysics* 56, 613–23.
- Hualde, J.I.** 1991: On Spanish syllabification. In Campos, H. and Martinez-Gill, F., editors, *Current studies in Spanish linguistics*. Washington, DC: Georgetown University Press, 475–93.

- Kessler, B. and Treiman, R.** 1997: Syllable structure and the distribution of phonemes in English syllables. *Journal of Memory and Language* 37, 295–311.
- Klatt, D.H.** 1974: The duration of [s] in English words. *Journal of Speech and Hearing Research* 17, 51–63.
- Kučera, H. and Francis, W.N.** 1967: *Computational analysis of present-day American English*. Providence, RI: Brown University Press.
- Ladefoged, P.** 1975: *A course in phonetics*. New York: Harcourt Brace Jovanovich.
- Lamminmäki, R.** 1979: The discrimination and identification of English words, consonants, junctures and sentence stress by Finnish comprehensive school pupils. In Lehtonen, J. and Sajavaara, K., editors, *Papers in contrastive phonetics*. Jyväskylä Cross-language Studies Series 7. Jyväskylä University of Jyväskylä, 165–229.
- Leather, J.** 1999: Second-language speech research: an introduction. *Language Learning* 49, 1–56.
- Macpherson, I.R.** 1975: *Spanish phonology: descriptive and historical*. Manchester: Manchester University Press.
- Major, R.** 1996: Markedness universals. *Studies in Second Language Acquisition* 18, 69–90.
- 2001: *Foreign accent*. Mahwah, NJ: Lawrence Erlbaum.
- McAllister, R.** 1996: Perceptual foreign accent: L2 users' comprehension ability. In James, A.R. and Leather, J.H., editors, *Second-language speech*. Berlin: Mouton de Gruyter, 119–32.
- McQueen, J.M.** 1998: Segmentation of continuous speech using phonotactics. *Journal of Memory and Language* 39, 21–46.
- McRobbie-Utasi, Z.** 1996: The implications of temporal pattern for the prosody of boundary signaling in connected speech. Paper presented at the Fourth International Conference on Spoken Language Processing, Philadelphia, PA. Available at: [www.asel.udel.edu/icslp/cdrom/session.htm](http://www.asel.udel.edu/icslp/cdrom/session.htm) (October 2004).
- Mines, M., Hanson, B. and Shoup, J.** 1978: Frequency of occurrence of phonemes in conversational English. *Language and Speech* 21, 221–41.
- MRC Psycholinguistic Database:** Available at: [http://www.psy.uwa.edu.au/mrcdatabase/uwa\\_mrc.htm](http://www.psy.uwa.edu.au/mrcdatabase/uwa_mrc.htm) (October 2004).
- Nakatani, L.H. and Dukes, K.D.** 1977: Locus of segmental cues for word juncture. *Journal of the Acoustical Society of America* 62, 714–19.
- Quené, H.** 1992: Durational cues for word segmentation in Dutch. *Journal of Phonetics* 20, 331–50.
- 1993: Segment durations and accent as cues to word segmentation in Dutch. *Journal of the Acoustical Society of America* 94, 2027–35.
- Ramana Rao, G.V. and Srichand, J.** 1996: Word boundary detection using pitch variations. Paper presented at the Fourth International Conference

on Spoken Language Processing, Philadelphia, PA. Available at: <http://www.asel.udel.edu/icslp/cdrom/session.htm> (October 2004).

- Ratcliffe, R.** 1996: Languages with obligatory onsets and without distinctive length. Message posted to LINGUIST List electronic mailing list, archived at: <http://linguistlist.org/issues/7/7-1698.html#1> (October 2004).
- Sanders, L.D. and Neville, H.J.** 2000: Lexical, syntactic, and stress-pattern cues for speech segmentation. *Journal of Speech, Language, and Hearing Research* 43, 1301–21.
- 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 cues. *Journal of Speech, Language and Hearing Research* 45, 519–29.
- Sebastián-Gallés, N., Dupoux, E., Segui, J. and Mehler, J.** 1992: Contrasting syllabic effects in Catalan and Spanish: the role of stress. *Journal of Memory and Language* 31, 18–32.
- Shriberg, L.D. and Kent, R.D.** 2003: *Clinical phonetics*. Boston, MA: Allyn and Bacon.
- Stockman, I. and Pluut, E.** 1999: Segment composition as a factor in the syllabification errors of second-language speakers. *Language Learning* 49, 185–209.
- Stockwell, R.P. and Bowen, J.D.** 1965: *The sounds of English and Spanish*. Chicago, IL: The University of Chicago Press.
- Stockwell, R.P., Bowen, J.D. and Silva-Fuenzalida, I.** 1956: Spanish juncture and intonation. *Language* 32, 641–65.
- Suomi, K.** 1985: On detecting words and word boundaries in Finnish: a survey of potential word boundary signals. *Nordic Journal of Linguistics* 8, 211–31.
- Suomi, K., McQueen, J.M. and Cutler, A.** 1997: Vowel harmony and speech segmentation in Finnish. *Journal of Memory and Language* 36, 422–44.
- Tench, P.** 2003: Non-native speakers' misperceptions of English vowels and consonants: evidence from Korean adults in UK. *IRAL (International Review of Applied Linguistics in Language Teaching)* 41, 145–73.
- Umeda, N. and Coker, C.H.** 1974: Allophonic variation in American English. *Journal of Phonetics* 2, 1–5.
- Vroomen, J., van Zon, M. and de Gelder, B.** 1996: Cues to speech segmentation: evidence from juncture misperceptions and word spotting. *Memory and Cognition* 24, 744–55.
- Wells, J.C.** 1990: Syllabification and allophony. In Ramsaran, S.N., editor, *Studies in the pronunciation of English*. London: Routledge, 76–86.
- Yavaş, M.** 1998: *Phonology development and disorders*. San Diego, CA: Singular Publishing.

**Appendix 1** Stimuli, with mean percentage correct (and standard deviation in parentheses) for each item

	Native speakers	Non-native speakers		Native speakers	Non-native speakers
<i>Aspiration stimuli: vowel-/s/-consonant (VsC):</i>					
Lou spills	95 (22.4)	27.6 (45.5)	loose pills	100 (0.0)	72.4 (45.5)
Lou stops	95 (22.4)	17.2 (38.4)	loose tops	95 (22.4)	86.2 (35.1)
Lou skis	100 (0.0)	13.8 (35.1)	loose keys	100 (0.0)	69.0 (47.1)
lay speech	100 (0.0)	55.2 (50.6)	lace peach	85 (36.6)	65.5 (48.4)
lay stable	100 (0.0)	65.5 (48.4)	lace table	100 (0.0)	72.4 (45.5)
lay scar	100 (0.0)	34.5 (48.4)	lace car	95 (22.4)	69.0 (47.1)
<i>Aspiration stimuli: consonant-/s/-consonant (CsC):</i>					
keep sparking	95 (22.4)	10.3 (31.0)	keeps parking	90.0 (30.8)	75.9 (43.6)
keep stalking	100 (0.0)	17.2 (38.4)	keeps talking	100.0 (0.0)	96.6 (18.6)
keep scanning	100 (0.0)	44.8 (50.6)	keeps canning	95.0 (22.4)	55.2 (50.6)
chief sport	100 (0.0)	75.9 (43.6)	chief's port	100.0 (0.0)	51.7 (50.9)
chief star	100 (0.0)	86.2 (35.1)	chief's tar	100.0 (0.0)	65.5 (48.4)
chief school	95 (22.4)	51.7 (50.9)	chief's cool	95.0 (22.4)	62.1 (49.4)
<i>Aspiration stimuli: consonant-/s/-consonant-consonant (CsCC):</i>					
cook sprints	100.0 (0.0)	27.6 (45.5)	cook's prints	80.0 (41.0)	69.0 (47.1)
cook struck	100.0 (0.0)	41.4 (50.1)	cook's truck	100.0 (0.0)	79.3 (41.2)
cook screams	90.0 (0.0)	31.0 (47.1)	cook's creams	95.0 (22.4)	72.4 (45.5)
top spy	100.0 (0.0)	51.7 (50.9)	tops pry	90.0 (30.8)	79.3 (41.2)
top strains	100.0 (0.0)	62.1 (49.4)	tops trains	95.0 (22.4)	79.3 (41.2)
top scrawled	100.0 (0.0)	82.8 (38.4)	tops crawled	95.0 (22.4)	89.7 (31.0)
<i>Glottal stop / creaky voice stimuli: vowel-nasal-vowel:</i>					
see neither	95.0 (22.4)	62.1 (49.4)	seen either	100.0 (0.0)	100.0 (0.0)
a niche	100.0 (0.0)	55.2 (50.6)	an itch	100.0 (0.0)	96.6 (18.6)
a nice man	100.0 (0.0)	72.4 (45.5)	an ice man	100.0 (0.0)	100.0 (0.0)
see Mabel	90.0 (0.0)	51.7 (50.9)	seem able	100.0 (0.0)	100.0 (0.0)
tea mat	75.0 (44.4)	69.0 (47.1)	team at	100.0 (0.0)	100.0 (0.0)
clay manual	65.0 (48.9)	55.2 (50.6)	claim annual	100.0 (0.0)	96.6 (18.6)
<i>Glottal stop / creaky voice stimuli: vowel-obstruent-vowel:</i>					
why fill	100 (0.0)	96.6 (18.6)	wife ill	100 (0.0)	100.0 (0.0)
low fate	90 (0.0)	89.7 (31.0)	loaf ate	100 (0.0)	100.0 (0.0)
low failing	100 (0.0)	96.6 (18.6)	loaf ailing	100 (0.0)	100.0 (0.0)
grey vat	95 (22.4)	65.5 (48.4)	grave at	100 (0.0)	89.7 (31.0)
we vend	100 (0.0)	96.6 (18.6)	weave end	100 (0.0)	86.2 (35.1)
say vamps	95 (22.4)	37.9 (49.4)	save amps	60 (50.3)	55.2 (50.6)
<i>Glottal stop / creaky voice stimuli: vowel-liquid-vowel:</i>					
l learn	100 (0.0)	82.8 (38.4)	l'll earn	100 (0.0)	96.6 (18.6)
say least	100 (0.0)	86.2 (35.1)	sail east	100 (0.0)	96.6 (18.6)
see love	95 (22.4)	86.2 (35.1)	seal of	100 (0.0)	100.0 (0.0)
be rolled	100 (0.0)	89.7 (31.0)	beer old	100 (0.0)	100.0 (0.0)
cue rake	100 (0.0)	86.2 (35.1)	cure ache	100 (0.0)	96.6 (18.6)
be rice	100 (0.0)	93.1 (25.8)	beer ice	100 (0.0)	100.0 (0.0)

(continued)

**Appendix 1** continued

	Native speakers		Non-native speakers			Native speakers		Non-native speakers	
<i>Double cue stimuli: aspiration-glottal stop:</i>									
grape in	100	(0.0)	86.2	(35.1)	grey pin	100	(0.0)	96.6	(18.6)
weep at	95	(0.0)	82.8	(38.4)	we pat	100	(0.0)	96.6	(18.6)
light old	100	(0.0)	96.6	(18.6)	lie told	100	(0.0)	96.6	(18.6)
might owe	100	(0.0)	96.6	(18.6)	my toe	100	(0.0)	100.0	(0.0)
make art	100	(0.0)	100.0	(0.0)	may cart	100	(0.0)	79.3	(41.2)
like old	100	(0.0)	100.0	(0.0)	lie cold	100	(0.0)	79.3	(41.2)

*Note:* Each subject was assigned a score of 100% if an item was answered correctly and 0% if the item was answered incorrectly.