Insights into L2 Connected Speech Segmentation: A Gating Experiment with Listeners of Different English Proficiency Levels

Naouel Zoghlami

Castledown

OPEN ACCESS

Australian Journal of Applied Linguistics
ISSN 2209-0959
https://www.castledown.com/journals/ajal/

Australian Journal of Applied Linguistics, 6(2), 94–113 (2023)
https://doi.org/10.29140/ajal.v6n2.1038

Insights into L2 Connected Speech Segmentation: A Gating Experiment with Listeners of Different English Proficiency Levels

NAOUEL ZOGHLAMI

Conservatoire National des Arts et Métiers, France
naouel.zoghlamiterrien@lecnam.net

Abstract

The relative contribution of bottom-up (i.e., acoustic-phonetic) and top-down (i.e., contextual) cues for successful L2 online segmentation is still a matter of debate. This study used the gating paradigm to investigate the segmentation processes of adult L2 English listeners with different proficiency levels, by looking at the type of cues they exploit and how they revise their hypotheses as connected speech is progressively revealed. Twenty-one French and Tunisian undergraduates were selected from a larger pool (n = 226) and identified as skilled (n = 11) and unskilled (n = 10) listeners based on their scores on standardized English listening and vocabulary tests. Descriptive statistics, analysis of variance and qualitative analysis were performed on the obtained data. Overall, this study provides supporting L2 evidence for the hierarchical nature of the multiple speech segmentation cues (Mattys et al., 2005). The results indicated an early effect of context on segmentation independent of L2 proficiency when the context is constraining. In non-constraining contexts, successful segmentation is delayed for both groups, with unskilled L2 listeners needing far more bottom-up information to process input and revise their segmentation hypotheses. We conclude that, in online L2 speech segmentation, what distinguishes proficient from non-proficient listeners is their efficient processing of bottom-up cues. Pedagogical implications are provided, hoping to help L2 English teachers (and materials developers) focus on bottom-up training to improve their learners' real-time comprehension competence.

Keywords: connected speech, segmentation, bottom-up and top-down cues, L2 English users, proficiency levels, gating

Introduction

As languages are generally spoken at about ten phonemes per second, and acoustic information is subject to fast decay (within one to two seconds, Baddeley et al., 2009), it is crucial for listeners to process oral input rapidly. Speech perception in the native language (L1) is immediate and effortless, given the lifelong exposure which begins before birth. However, it is far more challenging in a foreign language (L2). The ability to segment speech online (and hence recognize words) has been outlined as the major source of difficulty (Cutler, 2012; Snijders et al., 2007), being the strongest predictor of successful L2 listening comprehension (Adringa et al., 2012; Leonard, 2019; Matthews & Cheng, 2015; Matthews et al., 2023; Tsui & Fullilove, 1998), particularly determining the performance of low-ability listeners (Zoghlami, 2015). Adringa et al. (2012), for example, found that segmentation explains 40% of the variance for natives and 37% for non-natives. Segmenting speech is indeed an important skill that determines later language development as it correlates positively with vocabulary size at the age of two (Cutler, 2012, p. 279).

According to cognitive models of listening (e.g. Cutler & Clifton, 1999), segmentation is one of the main processes involved in speech perception. It is the process by which the brain breaks the continuous speech stream into potentially meaningful units (e.g., syllable or word) in order to understand what speakers say. Segmentation has been the object of extensive research in cognitive sciences including psycholinguistics – particularly in the L1 – and studies exploring cues to and probabilities of word boundaries is one way to approach lexical segmentation. There is indeed now a general consensus that this process involves the exploitation of bottom-up (the speech signal including phonological and prosodic patterns) and top-down (syntax, semantics, pragmatics and background knowledge) information in parallel and online as speech unfolds (Field, 2019; McClelland & Elman, 1986; Marslen-Wilson & Tyler, 1980). The findings so far also agree that language users with different experiences (native vs non-native) rely on different segmentation cues (Dobrego et al., 2022; Sanders et al., 2002). However, the relative contribution of bottom-up and top-down cues to solve the segmentation problem is still a matter of debate (Cutler, 2012; Eysenck & Keane, 2020). In particular, the exact role of contextual cues is yet to determine – especially with some researchers' strongly claiming that no contextual feedback is ever necessary (Cutler, 2012; Norris et al., 2000).

The segmentation process is problematic and extremely tentative in the L2, with listeners forming and possibly revising segmentation hypotheses as speech proceeds (Field, 2008a), given the nature of speech in terms of the high variability[1] of phonemes and spoken word forms and the absence of systematic cues to syllable/word boundaries. Co-articulatory effects (sounds being deleted, added, reduced, or converted into other sounds under the effect of adjacent phonetic context) are quite frequent in a language like English. Field (2019, p. 292) cites resyllabification, linking, elision and assimilation as the major articulatory phenomena present in all natural English speech. Take the example of regressive assimilation at word boundary in the frequent greeting expression *good morning* which is often pronounced /gʊb ˈmɔːnɪŋ/ – the alveolar /d/ occurring before the bilabial /m/ assimilates into another bilabial /b/. Despite its importance, L2 studies have paid little attention to the variability problem and the perception of casual speech. How L2 listeners use bottom-up and top-down information to segment connected speech and resolve frequent perception ambiguities (such as *night rate* vs *nitrate*) remains unclear. Accordingly, the goal of the present study is to address this issue by exploring the segmentation hypotheses that L2 users of different proficiency levels form while processing English-connected speech with certain phonetic features. The gating paradigm (Grosjean, 1980) as

[1] Of course, variability can be extra-linguistic and thus even more unpredictable, referring to specificities of speakers (accent, age, sex, shape of vocal tract, emotional and physical state) and speaking situations (noise, formality, multi-talker). In particular, noise and accent have been intensively studied, see Cutler (2012) for a comprehensive discussion of their role.

an online psycholinguistic measure was deemed appropriate to account for the complex interaction between signal and context cues and the time course of (successful) segmentation.

Literature Review

The Segmentation Problem: Bottom-Up or Top-Down?

Segmenting speech is a demanding cognitive task. Speech is continuous with no systematic (acoustic) spaces between words as in writing. In English, pauses only occur every six to eight words on average (Klatt, 1980). So how would a listener segment a phonemic sequence such as [ənaɪskri:m] which might correspond to the following spoken utterances *an ice cream*, *a nice cream*, *and I scream*? Listeners, in fact, exploit multiple sources of information, which are of different natures.

Every language has phonotactic (or possible-word) constraints[2] that govern the allowed sequence of sounds in a syllable. For example, /bm/ and /nf/ are illegal clusters in English syllables, hence necessarily signal syllable (*infant*) or word (*in front*) boundary. Similarly, the string /lp/ is not allowed in the onset position, but is legal in the coda position (*help*). Experimental research, particularly word-spotting studies, has shown that listeners use such phonotactic rules to segment speech and decide on word possibility in L1 as well as in L2 (Al-jasser, 2008; McQueen, 1998; Weber, 2000).

Prosody is another acoustic bottom-up cue which plays a crucial role in locating word boundaries (Cutler & Butterfield, 1992; Sanders et al., 2002). Recent evidence suggests that, even if languages have different rhythmic structures (for example, stress for English and Dutch, the syllable for French and Italian, and the mora for Japanese), the rhythmic segmentation strategy is less dependent on listeners' L1[3] than commonly assumed (Dobrego et al., 2022; Endress & Hauser, 2010; Katayma, 2015; Ip & Cutler, 2020). In English, the first syllable of most content words (like nouns and verbs) is usually stressed. Relying on a monitoring task, Katayama (2015) showed that Japanese speakers of distinct L2 English proficiency responded to target English words (e.g. *biscuit*, *picture*) faster when stress was on the first CVC syllable. They, in fact, relied on prosodic cues different from those of their L1, as they ruled out phonotactic knowledge and mora, and instead used stress to identify the target syllables. Using this *metrical strategy,* i.e. stressed syllables marking the beginning of content words, an L2 English listener would have a 90% chance of achieving correct segmentation (Cutler, 1990).

Top-down contextual information including information provided by the input (e.g. lexical, semantic, syntactic), pragmatic information and knowledge of the language (e.g. formulaic sequences, word frequency), also typically influence speech recognition. However, the exact role of context is still subject to considerable discussion, particularly when and how it operates during speech perception. According to *autonomous* theories (e.g. Shortlist, Norris 1994), only bottom-up acoustic information is used to activate lexical candidates, and context has no effect on the earliest stages of word recognition. However, the current evidence-based trend is in favour of *interactive* models (e.g. TRACE, McClelland & Elman 1986) where acoustic-phonetic input and contextual cues interact all through the different levels of speech processing. Previous and subsequent context can modulate the online analysis of the signal, making speech processing a feed-forward feed-back operation with constant adjustments being made (Grosjean, 1985, p. 309).

---

[2] According to Alexeeva et al. (2017), this does not seem to apply to Russian as it contains some single-consonant words without a vowel.

[3] See Cutler (2012, pp. 115–153) for a comprehensive review on the language-specificity of phonological cues to segmentation.

Context might influence segmentation – and thus spoken word recognition (SWR[4]) – even at early stages. Lucas (1999) reviewed 25 single-word priming studies investigating context effects where the presentation of the target word occurred before or at the end of the prime word. Effect sizes could be calculated in 17 studies with actual context effects being relatively delayed, i.e. some time after the offset of the prime word. However, as Brock and Nation (2014, p.114) argued, even if Lucas interpreted this result as evidence for an early contextual effect, it is quite difficult in such studies to determine the moment in processing at which the influence of context "kicks in". Brock and Nation (2014) directly addressed this issue of timing and used the visual paradigm to provide strong evidence of the immediate effect of context. Thirty-two native adult speakers of English listened to neutral and constraining sentences while viewing a computer display presenting four objects. The participants were asked to click on any object mentioned in the spoken sentences. Eye movements showed that the effect of the cohort displayed (e.g. *butter*) was significantly reduced when the preceding verb made the cohort implausible (e.g. *Sam fastened the but-ton*). The difference in eye movements (short fixations in constraining contexts vs long fixations in less-constraining contexts) was apparent early on in sentence processing. The authors' fine-grained analysis of the temporal evolution of cohort and context effects showed that their time-course was almost identical, as the effect was statistically significant a mere 30 milliseconds after the onset of the target word.

Findings about contextual constraints have been further corroborated by recent evidence coming from brain-imaging studies (e.g. Grisoni et al., 2017) showing that certain forms of top-down information such as word predictability – associated with immediate activation of language areas within the left frontal cortex – can indeed accelerate successful segmentation and leads to early recognition. Mattys et al. (2005) provided a hierarchical account of how the different cues operate simultaneously for (L1 English) spontaneous speech segmentation. Their model suggests that in normal hearing conditions, adult listeners firstly rely on sentential context (including pragmatic, semantic, and syntactic cues), followed by in descending weights lexical, segmental (phonotactics) and prosodic cues. Lower-level cues tend to be used when higher-level cues are unavailable or ineffective. However, this account has been recently challenged by Dobrego et al. (2022), who, in a chunking study comparing native and fluent L2 users of English (of different L1s[5]), showed that prosody outperforms syntax in both groups, and argued that this discrepancy with Mattys et al.'s evidence may lie in the nature of the speech material being spontaneous authentic (English) speech. This comparative study also showed that L1 users rely only slightly more on prosody than L2 users, and conclude that segmentation does not depend on language experiences at high levels of fluency.

Why Gating for L2 Continuous Speech Segmentation?

Though not as widely used as cross-modal priming and eye-tracking, the gating paradigm (Grosjean, 1980) is recognized as a powerful method that studies online SWR processes, accounting for the temporal evolution of the effects of the multiple segmentation cues. In gating, a spoken stimulus is presented in increments (called 'gates') of gradually increasing duration (in milliseconds) until complete revelation. After each gate, subjects are asked to report their guess of the stimulus and to give it a confidence rating. The dependent variables analysed in this paradigm are the isolation points (IPs, i.e. the moment at which the listener correctly identifies the stimulus), the confidence ratings and the different word candidates provided by the subjects after each gate. Apart from the time course of word

[4] Successful segmentation leads to SWR – which refers to the rapid encoding of acoustic signals into lexical representations. SWR involves the activation and competition of multiple lexical candidates. Segmenting isolated words is different from segmenting continuous speech as less competition is involved thanks to the lexical context. Davis et al. (1997, p.167) argue that "competition between lexical hypotheses that span potential word boundaries ensures that only words making up a consistent segmentation of the speech stream are activated" (as in the TRACE model).

[5] Such as Finnish, Russian, Turkish, Vietnamese, Italian, French, Greek, German and Korean.

identification, the qualitative data provided in the guesses is probably the main strength of gating as it reveals the information extracted from the acoustic-phonetic signal as well as the segmentation hypotheses formed by the subjects and how they revise them as the utterance proceeds (Field, 2008a).

The great majority of gating experiments (like the other experimental paradigms) have been carried out with native speakers (of English mainly) and with isolated words as stimuli. Grosjean (1980)'s seminal study was the first to show that contextual influence can occur even prior to contact with any bottom-up sensory input. He also calculated that the time needed for word activation is reduced by 50% when the target word is highly predictable in a specific context (1997). Two other classic studies (Grosjean, 1985; Bard *et al.*, 1988) provided evidence for late confident recognitions, those occurring sometime after the acoustic offset of targeted words. Grosjean (1985) gated infrequent monosyllabic and polysyllabic words with a preceding context that provided little semantic constraint, and found that listeners isolated with high confidence more than 50% of the short words only after their offsets. Bard et al. (1988) sought to offer a normative account of the delay in recognitions through the use of sentences taken randomly from conversational speech and presented in word increments. They concluded that late recognitions in spontaneous speech are fairly frequent, occurring roughly in every one in five recognized words. Though, this finding was found to be particularly true for function words, presenting speech in word increments could have distorted segmentation processes at word boundaries where coarticulation effects often occur.

To the best of my knowledge, only two gating studies (Field, 2008a; Shockey, 2003) investigated continuous speech segmentation processes of *non-native* speakers of English. Field (2008a) compared L1 and L2 English listeners' segmentation hypotheses of phonetically ambiguous short utterances (e.g. the stem [ˈdraɪvə]) and analysed when and how subjects revised their hypotheses when disambiguating input was revealed (*driver killed*, *drive a car*, *drive away*). Although Field's study did not directly tap into the type of cues used to segment speech, it showed that L2 listeners were not only slower than their L1 counterparts to adjust their interpretations on the basis of the incoming signal, but were also significantly reluctant to do so. Shockey (2003)'s study also showed how tentative L2 segmentation processes are when dealing with natural speech. She gated a whole sentence (*The screen play didn't resemble the book at all*) containing instances of coarticulation features. Her data provided further evidence for the retroactive nature of SWR as her L2 subjects – in comparison to L1 subjects – relied heavily on syntactico-semantic information as they waited for more spoken input to disambiguate reductions and interpret the sentence correctly. Interestingly, both Shockey (2003) and Field (2008a) concluded that relying on top-down information is a systematic strategic behaviour. This runs counter hypotheses of *bottom-up dependency* (Field, 2004) in L2 segmentation, i.e. directing too much attentional capacity towards the signal. However, as the data generated from these two studies come from subjects who are somewhat proficient in L2 English (advanced in Shockey 2003; intermediate to upper intermediate in Field 2008a), a question can be raised about the extent to which the use of contextual cues is direct in L2 connected speech, particularly for less proficient L2 listeners.

The Current Research

If we are to study online cognitive processing of speech, its connected nature needs to be taken into account. This is the first research gap the present study seeks to address by using stimuli that are more valid, reflecting continuous natural speech with common coarticulation features at word boundaries. Research has also mainly focused on L1 speech perception, and our knowledge of L2 speech processing remains limited. In particular, the relative contribution of bottom-up and top-down cues for L2 online segmentation is still unclear. The role of contextual top-down cues is a source of debate, with some researchers accrediting no effect to context (Cutler, 2012; Norris et al., 2000). Evidence exists on the use of different segmentation cues by native and non-native speakers (e.g. Dobrego et al., 2022;

Sanders et al., 2002). However, *when* and *what cues* L2 listeners of different abilities use to segment speech and recognize words have seldom been addressed. Thus, this research seeks to investigate potential differences between proficient and less-proficient L2 English listeners when trying to segment target spoken words presented in different types of discourse contexts: constraining and non-constraining syntactico-semantic contexts.

In addition, research has indicated that L1 influences L2 speech segmentation. Particularly, the role of bottom-up cues (e.g. rhythm and phonetic sequencing) seem to depend on listeners' L1 (Cutler, 2012). This study explored potential segmentation differences existing between listeners of distinct first languages (French vs Tunisian Arabic). As I also wanted to obtain fine-grained data about segmentation hypotheses, how certain properties of the target words including frequency and phonetic realization (which depends on coarticulatory features) would influence segmentation was taken into consideration. A variant of the gating paradigm was used to address the issue of timing (Brock & Nation, 2014), to explore which cues are being exploited, and to qualitatively tap into the segmentation guesses formulated by L2 English listeners as speech unfolds.

Thus, the present study provides an avenue for better understanding these issues in L2 speech segmentation by addressing the following research questions:

- In constraining contexts, which cues are used to segment connected speech into words? Is there a difference in listeners' responses across L1s and levels of L2 proficiency?
- In non-constraining contexts, which cues are used to segment connected speech into words? Is there a difference in listeners' responses across L1s and levels of L2 proficiency?
- Do some features of the stimuli (frequency and articulatory phenomena) account for segmentation differences?

Method

Participants

Twenty-one first-year English majors with an average age of 18.7 participated in the gating experiment. The subjects were native speakers of French and Tunisian Arabic who totalled about 750 hours of L2 English learning by the end of their secondary schooling. They were selected from a larger sample (n = 226)[6] according to their listening performance based on their scores in an FCE Listening test. FCE is a standardized certification assessing L2 English competence at an upper-intermediate level, and it is aligned with level B2 of the CEFR[7]. The FCE listening paper comprised four sections (multiple-choice, sentence completion, multiple matching, and sentence evaluation) with 30 items in total. Each correct answer was worth one point for a maximum score of 30. According to their obtained scores, eleven participants were identified as *Skilled* L2 listeners (top scores; MS = 22.09), the other ten as *Unskilled* L2 listeners (bottom scores; MU = 5). Mann Whitney test confirmed that the difference between these two groups was statistically significant (U = 0; p < .001).

Given the role of L2 lexical knowledge in listening comprehension, it was deemed necessary to control the subjects' size of English vocabulary, and the Nation and Beglar (2007)'s Vocabulary Size Test was used for this purpose. The skilled listeners (MS = 6609.09) had a much higher vocabulary volume

[6] The gating experiment is actually part of a larger study the author has been undertaking on the contributions of bottom-up and top-down processes in L2 listening comprehension.
[7] Common European Framework of Reference.

than their unskilled counterparts (*MU* = 4790). This means that, on average, a proficient participant in the gating experiment knew about 1800 more words than a non-proficient one. This difference in vocabulary knowledge is significant ($U = 5.500$; $p < .001$). No participants of the study had a known hearing impairment.

Stimuli

Four short affirmative sentences representing examples of naturally connected English speech were constructed for the experiment:

1) He lives on the ***first floor***.
2) It really was a very ***good concert***.
3) What nice ***blue earrings***.
4) Yesterday I found ***ten pence***.

The highlighted target items were noun structures appearing in the final position in the sentences. The items contained different coarticulation features at word boundaries, including assimilation, elision and linking. They were controlled for word length (one- and two-syllable words) and frequency using BNC K20 vocabulary levels in the "VocabProfile" program provided on the *Compleat Lexical Tutor* website. The gated items represent familiar vocabulary within the proficiency level of the L2 subjects, except for one word (*earrings*), which could be challenging as it is a low-frequency word (K7). Item predictability was also controlled as sentence 1 provides a more constraining context compared to the others. Table 1 describes the target items in terms of frequency level, spoken form (i.e. actual pronunciation in the recordings[8]), and the coarticulatory phenomena occurring at word boundaries. For item 2 (*good concert*), the retained spoken form did not contain the regressive assimilation feature we initially predicted /gʊgkənˈsɜːt/, as it is often the case in RP-connected English. The item was kept to measure the effect of word frequency and context on word recognition.

Table 1 *Description of the gated items*

Target Item	Word 1 Frequency	Word 2 Frequency	Citation Form	Spoken Form	Phonetic Features
first floor	K1	K1	/ˈfɜːrst flɔːr/	/ˈfɜːrsflɔː/	Elision of the alveolar plosive /t/ (within a consonant cluster) and final consonant deletion of /r/.
good concert	K1	K3	/gʊd kənˈsɜːrt/	/gʊdkənˈsɜːt/	Absence of assimilation: The /d/ in good is clearly pronounced. Included for evidence of the effect of other features (e.g. context) on segmentation.
blue earrings	K1	K7	/'bluː ˈɪə. rɪŋz/	/'bluːwɪərɪŋz/	Linking /w/: occurs when the first word ends in a rounded vowel (or a diphthong) and the next word starts with a vowel.
ten pence	K1	K1	/'ten 'pens/	/'tem'pens/	Regressive assimilation (of place): the alveolar nasal /n/ becomes a bilabial /m/ because it is followed by a bilabial sound /p/

[8] The sentences were pronounced by a native speaker of American English and recorded in a soundproofed room.

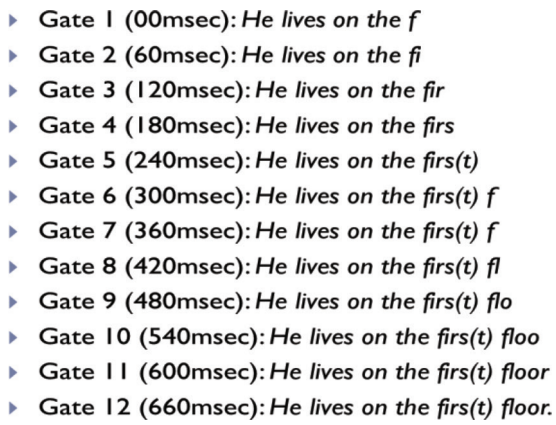

Figure 1 *Example of sequential gated presentation.*[9]

Procedure and Analysis

In this gating experiment and following Shockey's (2003) recommendations, the items were gated in time increments of 60ms which is slightly longer than time gates in previous research. Figure 1 illustrates an example of a gated presentation. The first gate includes the pre-item context along with the onset of the item. The following gates present sequentially longer gates (+60ms for each gate) until the whole sentence is revealed.

The gating experiment was run in E-prime 1.2 (Schneider, Eschman & Zuccolotto, 2002). The gates were presented to subjects individually and following the same order: utterance 1 (12 gates), utterance 2 (14 gates), utterance 3 (13 gates), and utterance 4 (12 gates). Participants were instructed to listen carefully to each segment, type out their guesses of the word being presented after each gate and indicate their confidence level using a 4-point scale (1 = very sure; 2 = fairly sure; 3 = fairly unsure; 4 = very unsure). To provide the L2 listeners with an opportunity of comprehending what was requested, a practice sentence – *She's a freelance translator* – was presented prior to the experiment with *translator* (/trænsˈleɪtə/) being the gated item.

A total of 1071 gate/answer entries were registered in E-prime and transferred in an Excel sheet. The data were analysed quantitatively and qualitatively. We identified the IPs for both words composing each item across subjects and for each sub-group. Descriptive statistics of the IPs were performed including modes, means and standard errors (*SE*[10]). For each IP, we also calculated the mean of the confidence ratings (*M*C). An analysis of variance was then run with L2 proficiency and L1 as independent variables and duration (i.e. the amount of acoustic-phonetic input measured in ms) as the dependent variable. A qualitative analysis of the obtained word guesses was performed to account for the segmentation hypotheses and differences between skilled and unskilled L2 listeners, knowing that erratic spellings were excluded from the counting of correct entries. To account for all the issues in L2 segmentation addressed in the research questions, a thorough analysis is provided for each of the target items.

[9] For illustration, an audio file of the gates presented successively is accessible here https://www.cjoint.com/c/KJlkVfsFlPw

[10] Here, the *SE*s correspond to the percentage of subjects' recognition errors; i.e. when items were not recognized by presentation of the last gate.

Results

L1 Effect

The L2 participants of this study have very distinct L1s: French and Tunisian Arabic. However, no L1 effect was found. In fact, ANOVA tests showed no statistically significant segmentation differences between the two groups ($p > .05$), regardless of the syntactico-semantic context and the features of the spoken items (word frequency and articulatory phenomena).

Isolation Points

An examination of the IPs allows us to determine the time it took for the target word/s to be correctly recognized. Table 2 presents the descriptive statistics of all target word IPs for the whole participant group ($n = 21$). Means are provided in gate number (G), while modes are given in both gate number (G) and its respective duration in milliseconds (ms).

The *Mean* values of IPs shows that successful word recognition can occur at early gates depending on the context. The low *SE* values of items 1 and 2 indicate that they were recognized accurately by most of the participants. The identification of both words in item 1 *first floor* occurred before their offset was completely presented. Compared with the other three sentential contexts, the utterance "*He lives on the"* provides a strong discursive constraint: few noun phrases are possible at the final gate. For item 2, *good* was equally highly predictable and was recognized by the majority of the L2 subjects from the very first acoustic evidence with a *M*IP of 1.81, i.e. occurring between gate 1 (00ms) and gate 2 (60ms). The preceding discursive context and the high frequency of *good* (K1) probably played a key role in its early recognition. Identification of *concert* was, in contrast, delayed (*M*IP = 10.67) as subjects had to wait until the input of its offset (second syllable) was revealed to rule out possible candidates. This can be explained by the fact that the pre-item context is semantically open, and the word *concert* is less frequent (K3).

Table 2 *Descriptive statistics for IPs (n = 21)*

		Mode (G/ms)	Mean (G)	SD (G)
Item 1	first (K1)	4 (180ms)	4.14	2.372
***n*G = 12**				
***SE* = 9.5%**	floor (K1)	8 (420ms)	7.24	3.113
***MC* = 2.33**				
Item 2	good (K1)	1 (00ms)	1.81	0.873
***n*G = 14**				
***SE* = 14.28%**	concert (K3)	11 (600ms)	10.67	2.938
***MC* = 1.76**				
Item 3	blue (K1)	14 (720ms)	10.14	4.976
***n*G = 13**				
***SE* = 57.14%**	earrings (K7)	14 (720ms)	12.62	2.014
***MC* = 1.95**				
Item 4	ten (K1)	1 (00ms)	6.10	4.253
***n*G = 12**				
***SE* = 85.7%**	pence (K1)	13 (720ms)	12.43	1.805
***MC* = 2.1**				

nG = total number of item gates; a gate = 60ms.

The standard errors of items 3 (*blue earrings*; *SE* = 57.14%) and 4 (*ten pence*; *SE* = 85.7%) show that their online segmentation was more problematic. It seems that, although the word *blue* is monosyllabic and highly frequent, its recognition was delayed occurring for the majority between gates 9 and 10 (480–540ms), with acoustic evidence for *earrings* starting at gate 4. In the absence of contextual constraint, the delay is probably due to the phonetic feature at the word boundary (/'bluːwɪərɪŋz/), combined with the low frequency of the word *earrings* (K7). As for item 4, only three subjects segmented it correctly with a mean IP of the word *ten* between gates 6 and 7 (300–360ms) – i.e. far after the word was revealed completely. Segmentation of *pence* was problematic though its phonological evidence started at gate 4 (180ms).

The analysis of the confidence ratings at the different IPs showed that the respondents generally became more confident about their answers as the size of the acoustic input increased – though this observed pattern was not statistically significant. Interestingly, the confidence ratings of all items are rather at odds with the accuracy of their recognition. The participants' *MC* index for item 1 was the lowest, despite its correct and rapid recognition. One could wonder if this is actually an experimental effect, as participants might still be adjusting to the gating paradigm. Similarly, for items 3 and 4 – although most of the subjects failed to recognize them, they seemed rather confident about their segmentations (*MC*3 = 1.95; *MC*4 = 2.1).

Online Segmentation Differences Across L2 Levels

An examination of Figure 2, which shows the recognition time course and the frequencies of correctly identified words across L2 listening levels, indicates that unskilled listeners generally need more time – and thus more bottom-up acoustic evidence – to achieve an accurate segmentation of the items. The analysis of variance revealed that this difference was significant across items, except for item 1. The qualitative analysis showed a considerable variation in the way skilled and unskilled L2 listeners form and revise their online segmentation hypotheses, in particular, at word boundaries. An account of this variation for each item is provided in the following sections.

First Floor - /ˈfɜːrsflɔː/

As shown in Figure 2, both skilled and unskilled listeners recognized item 1 early in the time course of the gated sentence. The majority required only 180ms – actually corresponding to the revelation of /ˈfɜːrs/ as acoustic evidence – to recognize the word *first*. Knowledge of phonotactic constraints – the cluster /rst/ is only allowed at syllable final position in English – might have helped segmenting the word. However, the most striking result is that most of the proficient subjects identified *floor* prior to or by hearing its very first acoustic evidence starting at 300ms. It is very likely that these listeners were able to predict the target item thanks to the constraining lexico-semantic context (top-down information) though it illustrated typical connected speech reductions at word boundaries. Unskilled listeners generally took more time to recognize the whole item. However, recognition difference between the proficiency sub-groups for this item was statistically non-significant ($p > .07$). One unskilled subject could even isolate *floor* as early as gate 4, i.e. at 180ms.

Table 3 presents the gate responses and confidence ratings provided by a representative listener in each subgroup. Accordingly, the unskilled L2 listener does not seem to use the context, and rather relies heavily on the acoustic signal to the point of inventing words as seen in the tentative segmentations at gates 420 and 480ms. The entry *firstful* shows that the subject did not segment the word boundary correctly; the cluster /rstf/ is phonotactically illegal word-internally in English. She revises her segmentation hypothesis in the following gate, but keeps on using a bottom-up strategy, *lexical substitution* (Field, 2004, p. 373), as the revealed signal seems to be used to approximately match the input

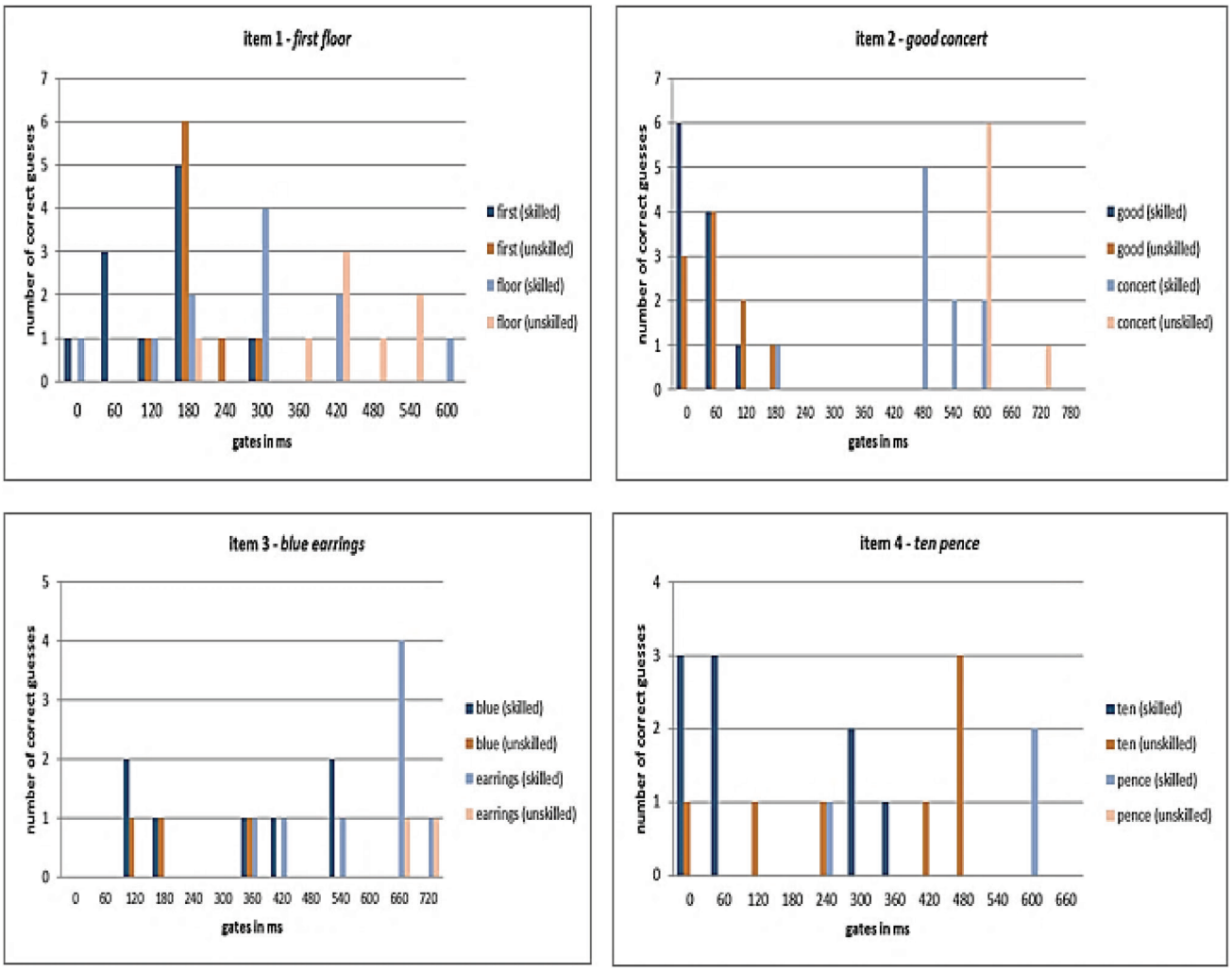

Figure 2 *Time course of item recognition for skilled and unskilled L2 listeners.*

Table 3 *Example of item 1 segmentation hypotheses (first floor)*

Gates in ms	Skilled L2 Listener		Unskilled L2 Listener	
	Response	**Confidence**	**Response**	**Confidence**
00	he lives on the	4	-----	2
60	on the f	2	-----	1
120	on the fa	2	-----	4
180	first	2	first	1
240	first	1	first	1
300	first floor	2	first	1
360	first floor	1	first	2
420	first floor	1	firstful	2
480	first floor	1	first fluence	2
540	first floor	1	first	2
600	first floor	1	first flore	1
660	first floor	1	first flore	1

to a familiar lexical candidate (*fluence* is most likely matched to *influence*). The IPs – signaled by the grey lines – show that it takes twice as long – and the acoustic input almost completely revealed – for the weak listener to activate a lexical representation quite confidently (1 = very sure). Yet, the final transcription (*flore*) gives the impression that the word *floor* might not be part of the subject's vocabulary – and I am not sure if an accurate meaning activation actually occurred.

***Good Concert* - /gʊdkənˈsɜːt/**

It is apparent from the data displayed in Table 4 on listeners' segmentation hypotheses of item 2 and the second histogram of Figure 2 that – in an utterance where the target nominal group is less constrained by the context (*It really was a very good concert*) – all listeners need more bottom-up linguistic information to activate the correct candidates confidently. The adjective *good* was recognized quite early in the time course. Six of the proficient subjects (*nS* = 11) even guessed it correctly at gate 00ms, probably by exploiting either coarticulatory, lexical or collocational information. Differences between the skilled and the unskilled listeners are better observed in the isolation frequencies and segmentation hypotheses of the word *concert* whose acoustic information started to be revealed at gate 240ms. Skilled listeners were consistently faster than their less-skilled counterparts in identifying *concert*, with most of the skilled subjects segmenting it correctly and confidently at gate 480ms, i.e. immediately at hearing the right first acoustic evidence of the second syllable [ˈsɜːrt]. One skilled listener even had an anticipated correct activation as early as gate 180ms.

As can be seen in Table 4, some of the subjects' responses (*n*=21) included blanks and simple segment transcriptions of what they decoded from the acoustic increments (examples: *c*, *ca*, *can*, *cans*, *conse*). The subjects were apparently hesitant to segment the word in the absence of sufficient bottom-up information. Among the activated candidates to *concert*, answers included *company*, *conversation*, *contact*, *cop*, *car*, *coffee*, *cause*, *country*, *camping*, *competition* and *contest*, which are in the majority K1 words. The data suggest that when the context is not constraining, L2 listeners rely on the signal and the words available in their L2 mental lexicon (frequency effect). Though provided with low confidence levels,

Table 4 *Example of item 2 segmentation hypotheses (good concert)*

Gates in ms	Skilled L2 Listener		Unskilled L2 Listener	
	Response	Confidence	Response	Confidence
00	It really was a very good	1	----	4
60	good	1	good	1
120	good place	2	good	1
180	good place	1	good	1
240	compa	2	good c	4
300	company	2	good campagny	4
360	comp	3	good campagny	3
420	cont	3	good can	4
480	concert	1	good cans	4
540	concert	2	good consonant	4
600	concert	1	good concern	3
660	concert	1	good concern	2
720	concert	1	good concert	1
780	concert	1	good concert	1

some segmentation hypotheses of non-proficient listeners were inaccurate given the lexico-syntactic context of the item. Examples included *consonant* and *concern*, but also *concentration*, *cancer* and *compassion*. It seems then that, for unskilled L2 subjects, the bottom-up signal prevails over top-down information, which can be at the expense of semantic coherence. One-way ANOVA showed that the difference in recognizing the word *concert* between skilled and unskilled listeners was quite significant ($F(1,10) = 9.451$, $p = .006$).

***Blue Earrings* - /'blu:wɪə.rɪŋz/**

As clearly seen in Table 2 and Figure 2, less than half the participants achieved a correct segmentation of item 3. In fact, only seven skilled and two unskilled subjects could identify the item as whole, and all did with some delay (*M*S = 11.63 vs *M*U = 13.7). Given word frequency and the coarticulatory features at word boundaries (Table 1), this result was partially expected for *earrings* (K7) but not for *blue* (K1). One proficient listener even recognized *earrings* but failed to identify *blue*. Confident recognition of *blue* occurred at the minimum 180ms after the revelation of its offset. Nevertheless, segmentation differences across L2 proficiency for this item were significant for *earrings* ($F(1,19) = 6.989$; $p = .01$) but not for *blue* ($p > .3$). Examples of the segmentation problems encountered by listeners across different levels are provided in Table 5.

The reported guesses demonstrate how coarticulation complicates the process of L2 speech segmentation, even for proficient L2 learners. As expected, the phonetic feature (linking /w/) at the word boundary (/'blu:wɪə.rɪŋz/) did blur segmentation. This is clearly shown at gate 480ms in the first unskilled subject's response '*blue wearing*', knowing that acoustic evidence for *earrings* started at gate 360ms. More interestingly, segmentation was even more problematic because of another unpredictable (or at least not initially controlled in this gating experiment) phonological feature: the voiceless final /s/ sound in the word *nice* influenced the pronunciation of the following word-initial /b/ sound in *blue*. /b/ was partially devoiced and perceived as /p/ during several gates by one-third of the total sample ($n = 21$). In addition to the ones reported in column three, wrong segmentation hypotheses due to this devoicing effect included *splourings*, *pluriarings*, *plurial wings*. Non-proficient listeners did not

Table 5 *Example of item 3 segmentation hypotheses (blue earrings)*

Gates in ms	Skilled L2 Listener		Unskilled L2 Listener (1)		Unskilled L2 Listener (2)	
	Response	Confidence	Response	Confidence	Response	Confidence
00	what a nice p	2	place	1	pol	4
60	what nice pl	2	place	1	plu	4
120	what nice blu	3	blue	1	plu	4
180	what nice blue	2	blue	1	plui	4
240	what nice	1	blue	2	pluit	4
300	what nice	1	blue	2	pluriel	4
360	blue earing	2	blue	4	plurielly	4
420	blue earring	2	blue	4	plurielly	3
480	blue earring	1	blue wearing	3	plurieling	4
540	blue earring	1	blue earing	1	pluring	3
600	blue earring	1	blue earring	1	pluring	2
660	blue earrings	1	blue earrings	1	plurings	2
720	blue earrings	1	blue earrings	1	plurings	2

abandon these incorrect segmentations even in light of disambiguating acoustic evidence at later gates, which points to their confidence in the signal as underscored by the final good confidence levels indicated by the second unskilled listener. Another possible interpretation would be that the infrequent English word *earrings* (K7) may be simply unknown, at least in its spoken form, to the unskilled subjects, suggesting that segmentation differences across L2 linguistic levels also lie in variation in L2 lexical and phonological knowledge.

***Ten Pence* - /'tem'pens/**

Analysis of the fourth histogram in Figure 2 shows that the word boundary feature (regressive assimilation of /n/ into /m/) posed a major problem for L2 learners when attempting to segment item 4. Compared to proficient listeners who were faster (gates 00–60ms) and more confident ($M = 2.4$) when isolating the word *ten*, the majority of the weak listeners needed more acoustic input, and only recognized it after its offset (minimum of 300ms). The difference between the groups was significant ($F(1,19) = 5.114$; $p = .03$). As for *pence*, only three skilled listeners recognized it, and none of the unskilled ones did. Two of the successful subjects recognized *pence* at gate 11 (600ms), while the third isolated it 6 gates earlier (240ms). It is to be noted that robust tests of equality of means could not be performed for pence. Some segmentation attempts of the L2 listeners are provided in Table 6 and further reveal the differences between skilled and unskilled listeners.

The guesses provided show how the revision of segmentation hypotheses differed depending on L2 proficiency level and use of available cues. The transcription of the last gate in example 4 (*tempends*) indicates that the unskilled listener did not use the context to recognize (and understand) the item. Again here, although the proposed non-word is semantically incoherent, the subject did not revise her hypothesis relying exclusively on the signal in a linear fashion. Interestingly, a similar behavior was observed in the data as far as the recognition of *pence* is concerned. As we can see in examples 2 and 3, the listeners confidently transcribed their guess of the last gate as *ten pens* – which is actually the case for over half of the total sample ($n = 21$), skilled and unskilled listeners alike. It seems that listeners used bottom-up information coming from the gated input, disregarding phonological knowledge about the pronunciation of the plural form of the word *pen*, being /'penz/ and not /'pens/. Early transcriptions in Example 2 show that the skilled subject had a confident guess on the gated

Table 6 *Example of item 4 segmentation hypotheses (ten pence)*

Gates in ms	Skilled Listener (1)		Skilled Listener (2)		Unskilled Listener (3)		Unskilled Listener (4)	
	Response	Confidence	Response	Confidence	Response	Confidence	Response	Confidence
00	ta	3	ten dolars	2	I found	1	test	4
60	te	3	ten dollars	2	te	1	test	4
120	tem	3	ten bax	1	ten	3	ten	4
180	temp	3	ten pounds	1	temp	3	tem	4
240	tempe	2	ten pounds	1	temperature	2	tempeture	4
300	tempe	2	ten pounds	1	temperature	2	tempeture	4
360	ten pen	3	ten pounds	1	temp	2	tempen	4
420	ten pen	2	ten pounds	1	ten pen	4	tempen	4
480	ten pen	2	ten pen	1	ten pen	2	tempen	4
540	ten pens	2	ten pens	2	ten pens	2	tempend	4
600	ten pence	2	ten pens	1	ten pens	2	tempends	4
660	ten pence	2	ten pens	1	ten pens	2	tempends	3

word (*dollars*, *bax*[11], *pounds*), very likely through a syntactic analysis of the item context – even if she finally failed to recognize accurately the target item. This was the case for other skilled listeners who used the syntactico-semantic context to formulate inaccurate but plausible hypotheses.

Discussion: Which Cues? When? Why?

The present gating experiment provided further evidence on which cues L2 English listeners of different proficiency levels attend to and how they exploit them to segment connected speech. The items selected represented authentic input with coarticulation features at word boundaries presented in constraining (item 1) and less-constraining contexts (items 2, 3 and 4). The overall data showed that the skilled L2 listeners were consistently faster – needing less bottom-up information to revise their segmentation hypotheses – and generally more confident than their unskilled counterparts in identifying words (IPs).

The gating data confirmed one of the robust findings in speech recognition research: (early) word recognition depends largely on the degree of contextual constraint regardless of L2 listening proficiency. The answers of the majority of the subjects (IPs and segmentation guesses) to the first gated stimulus (*he lives on the first floor*) provided evidence for the role of top-down contextual information. The preceding syntactico-semantic context was constraining: the verb collocation *live on* allowed subjects – including the unskilled listeners – to recognize the item quickly despite speech reductions. This finding contradicts Dobrego et al. (2022)'s results suggesting the importance of bottom-up cues over top-down ones and further supports the hierarchical nature of the cues relied on when segmenting continuous speech (Mattys *et al.*, 2005): contextual top-down cues, when available, override acoustic-phonetic bottom-up cues. The collocational effect is significant as some of the L2 listeners could predict the item accurately even before the revelation of any of its acoustic information. These results are also in line with L1 studies showing that activation time is reduced when the word is predictable (Brock & Nation, 2014; Grosjean, 1997), and that word identification can occur before hearing target onsets (Grosjean, 1980). Overall, the findings indicate that listening to L2 speech entails processing input at a supra-lexical level – something that research focused on recognition of isolated words might lead us to forget, which is further evidence for *interactive* accounts of SWR.

Another important finding is that listeners need more time to recognize L2 words in phonologically-reduced input, independently of word frequency and L2 proficiency. In fact, the recognition of some of the highly frequent (monosyllabic) words was problematic because of coarticulation features at word boundaries even for skilled listeners – as particularly seen in the segmentation results of the K1 words *blue* and *ten*. The recognition of these words was delayed occurring far after their complete revelation, which corroborates earlier findings in the literature of L1 spoken word recognition showing that belated identification of 50% of frequent short words is quite common in speech processing (Grosjean, 1985; Bard *et al.*, 1988).

The processing lag observed in the present gating experiment with skilled and unskilled L2 English listeners is also consistent with previous studies comparing native and non-native English speakers on word identification in complete sentences (Shockey, 2003) and on segmentation of ambiguous stems (Field, 2008a). The recognition time course results clearly show that the less proficient L2 listeners need more time and, therefore more acoustic input to recognize lexical items. According to the different guesses provided after each gate as the stimuli were gradually revealed, the segmentation hypotheses of the unskilled listeners seem to rely heavily on bottom-up information from the signal, particularly in non-constraining contexts. The unskilled subjects' hypotheses included a higher num-

[11] Subject's incorrect spelling of *bucks*.

ber of blanks, mere transcriptions of recognized sounds (*fa*, *compa*, *cans*, *plu*, *pol*, *blu*, *temp*, *te*), and non-existent words (*firstful*, *fluence*, *splourings*, *pluring*, *tempends*). Beyond the neologisms and the syllables as clear segmentation attempts, we agree with Field (2008a) that listeners' blank reactions do not necessarily reflect an absence of segmentation but a hesitant or incorrect one. As he put it, this type of segmentation decision shows "*an adherence, in the absence of contrary evidence, to an inappropriate segmentation*" (p. 48). Other lexical candidates proposed by the unskilled listeners showed a reliance on the acoustic input at the expense of semantic coherence, as observed in some incorrect/unconfident segmentation hypotheses: *It really was a very good (cancer)* or *Yesterday, I found (temperature)*. It seems that – even after the complete revelation of the stimulus, and therefore revelation of the sentential context – the non-proficient listeners continued to trust the signal, failing to integrate top-down information from the context that could limit the potential candidates.

The L2 skilled participants also occasionally applied a *linear* bottom-up processing. However, they were able to adjust their segmentation hypotheses more rapidly when disambiguating acoustic evidence was revealed. A clear example comes from the results of the problematic item *ten pence* presented in a non-constraining lexico-semantic context. The findings showed that even proficient subjects processed the signal linearly and failed to use top-down phonological knowledge (spoken form of *pence* vs *pens*) that could have helped them recognize the target candidate. As only a minority of proficient listeners identified *pence*, and knowing that it is not their top-down processing which is at fault as was observed in the early guesses of the skilled listener 2, one can wonder whether linear bottom-up processing is related to the absence of the target candidate in the mental lexicon rather than an inability to segment speech and recognize words. The skilled listeners would, in such cases, resort to lexical substitution (*pens*) in the same way as unskilled listeners. The gating experiment involves frequent English vocabulary that is likely to be known by the non-native participants. However, the phonetic realizations of this vocabulary in connected speech might be unfamiliar, in particular for non-proficient learners. This further highlights the importance of a large L2 phonological knowledge for successful segmentation.

The existing attempts to explain the processing lag are not fully satisfactory. Shockey (2003), for example, interestingly assigns recognition delay and a seemingly dependence on the signal to an over-reliance on syntactic-semantic information as listeners wait for context clues to be gathered from the incoming perceptual evidence. Field (2008a) speaks of a *perseveration effect*, where weak listeners would be reluctant to abandon an initial segmentation hypothesis due to a lack of confidence. Field even argues that if L2 listeners react this way under the easily-paced conditions of gating, they are very likely to behave similarly under the time constraints of authentic listening.

I alternatively believe that a more plausible explanation is related to the cognitive load and the role of working memory (Baddeley, 2000). Processing speech in the L1 is efficient and effortless because it is highly automatic, which refers to immediate, unconscious, and attention-free activation of word candidates. For the L2 listener, due to deficits in the target linguistic knowledge, formal processing consumes most of the attentional resources. For the unskilled participants in this study, focusing on the upcoming input and trying to segment it probably saturated their attentional capacity, leaving few resources for top-down processing. This better explains why these listeners were unable to revise and integrate their segmentation hypotheses with what they recognized/understood from the utterance to the point of inventing English words or providing lexical candidates that were completely incoherent in the sentential context. Caution is still needed when interpreting the apparent dependency of unskilled listeners on the acoustic signal. There is, in fact, no clear and direct evidence that listeners did not attempt to exploit top-down clues while getting bottom-up information in each gate – hence probably the necessity of having more brain imaging studies that would investigate the neural processing of L2 speech segmentation modulated by the degree of L2 proficiency.

Another interesting finding suggests that L2 English segmentation is not dependent on listeners' L1. In fact, the study did not reveal any significant differences cross-linguistically. French and Tunisian participants seemed to rely on the same cues in the same way, although French is a syllable-timed language and Tunisian Arabic is a stress language. The present data could be interpreted as further support for the universality of the rhythmic strategy (e.g. Katayama, 2015; Ip & Cutler, 2020) with caution as we did not explicitly tap into the role of stress. More experimental investigations on this topic are needed, especially involving Arabic languages as they are scarcely studied.

Conclusion and Implications for the L2 English Classroom

This paper is probably the first gating study that sought to compare speech segmentation processes of L2 users of English with different proficiency levels. It also addressed another gap in research by using experimental stimuli which better reflect the interconnected nature of speech. Nevertheless, the stimuli remained artificial as they were recorded in ideal circumstances (speaker reading out loud in a soundproofed room; subjects were equipped with headsets). Listening/communication conditions are completely different in real life involving environmental noises, multi-talker contexts and unpredictable speaker-variation. Speech perception studies relying on authentic speech are rare, and this needs to be addressed in future research.

This study provided further insights into the type of cues used by proficient and non-proficient L2 English listeners when segmenting continuous speech. It showed that predictability plays an important role. When the context is restraining, top-down information is prominent regardless of L2 proficiency. This means that skilled and unskilled listeners are able to use the context in a similar way. On the whole, it is an efficient use of bottom-up cues that distinguishes between L2 listeners with different abilities. This calls into question the privileged place that has been attributed to top-down processing in language classrooms and which presents the strategic use of context as a cure-all for any listening problem including perceptual ones. Undoubtedly, this is due to the dominance of the Communicative Approach over the past thirty years. Without questioning the very teachability of top-down strategies (Dörnyei, 1995), the empirical evidence provided by this study suggests that it is time to reverse this prevailing trend and call for a focus on bottom-up training to improve (particularly less-skilled) L2 English learners' real-time perceptual processing – and hence their listening comprehension competence.

Some of the segmentation hypotheses revealed listeners' inability to segment speech even when it is composed of frequent words. Frequent vocabulary may not be recognised because of coarticulation features in connected authentic speech which blurs word boundaries. These cues are, unfortunately, still widely neglected in language classrooms. Bottom-up perceptual training can simply start with raising learners' awareness about the variability of speech as a major cause of their segmentation delay and failure. Teachers can, for example, pick and teach the most challenging coarticulatory phenomena (assimilation, elision, linking, etc.) to their students. Field's (2008b, p. 140–162) list might be useful for this purpose as it provides a rich account of the unreliable nature of the English speech signal. Pedagogically, the findings also illustrate the importance of developing L2 learners' phonological knowledge. English learners need to be taught the different phonological realisations of (at least frequent) words (e.g. *pence* vs *pens*). Decoding and phoneme discrimination work also need to be considered in language classrooms – as seen with /p/ and /b/ being easily confused in some phonological (word boundary) contexts, like in the item *blue earrings*. To develop learners' speech segmentation skills and hence automatic SWR and comprehension, this bottom-up training needs to be explicit and systematic – probably encouraging learners to carry it out individually outside the classroom through CALL-based activities.

Disclosure Statement

There are no competing interests to declare.

Acknowledgments

I warmly thank the editors for their support and the reviewers for all the valuable comments which contributed to improving the quality of the manuscript. Of course, I am also grateful to all the French and Tunisian learners of English who participated in this project.

References

Adringa, S., Olsthoorn, N., van Beuningen, C., Schoonen, R., & Hultijn, J. (2012). Determinants of success in native and non-native listening comprehension: An individual differences approach. *Language Learning, 62*(Issue Supplement 2), 49–78. https://doi.org/10.1111/j.1467-9922.2012.00706.x

Alexeeva, S., Frolova, A. & Slioussar, N. (2017). Data from Russian help to determine in which languages the possible word constraint applies. *Journal of Psycholinguistic Research, 46*, 629–640. https://doi.org/10.1007/s10936-016-9458-7

Al-jasser, F. (2008). The effect of teaching English phonotactics on the lexical segmentation of English as a foreign language. *System*, 36(Issue 1), 94–106. https://doi.org/10.1016/j.system.2007.12.002

Baddeley, A., Eysenk, M., & Anderson, M. (2009). *Memory.* Psychological Press

Baddeley, A. (2000). The episodic buffer: a new component of working memory? *Trends in Cognitive Sciences, 4*, 417–423. https://doi.org/10.1016/S1364-6613(00)01538-2

Bard, E., Shillcock, R., & Altmann, G. (1988). The recognition of words after their acoustic offsets in spontaneous speech: effects of subsequent context. *Perception and Psychophysics, 44*, 395–408. https://doi.org/10.3758/BF03210424

Brock, J., & Nation, K. (2014). The hardest butter to button: Immediate context effects in spoken word identification. *The Quarterly Journal of Experimental Psychology, 67*(1), 114–123. https://doi.org/10.1080/17470218.2013.791331

Cutler, A. (1990). Exploiting prosodic possibilities in speech segmentation. In G. Altmann (Ed.), *Cognitive Models of Speech Processing* (pp. 105–121). MIT Press.

Cutler, A. (2012). *Native Listening: Language Experience and the Recognition of Spoken Words.* MIT Press.

Cutler, A., & Butterfield, S. (1992). Rhythmic cues to speech segmentation: Evidence from juncture misperception. *Journal of Memory and Language, 31*, 218–236. https://doi.org/10.1016/0749-596X(92)90012-M

Cutler, A., & Clifton, C. (1999). Comprehending spoken language: a blueprint of the listener. In C. Brown, & P. Hagoort (Eds.), *The Neurocognition of language* (pp. 123–166). Oxford: Oxford University Press.

Davis, M., Marslen-Wilson, W., & Gaskell, M. (1997). Ambiguity and competition in lexical segmentation. In *Proceedings of the Nineteenth Annual Conference of the Cognitive Science Society* (pp. 167–172). Mahwah: Lawrence Erlbaum ASSOC PUBL.

Dobrego, A., Konina, A. & Mauranen, A. (2022). Continuous speech segmentation by L1 and L2 speakers of English: the role of syntactic and prosodic cues. *Language Awareness*, https://doi.org/10.1080/09658416.2022.2131801

Dörnyei, Z. (1995). On the teachability of communication strategies. *TESOL Quarterley, 29*(1), 55–85. https://doi.org/10.2307/3587805

Endress, A. D. & Hauser, M. D. (2010). Word segmentation with universal prosodic cues. *Cognitive Psychology*, *61*(2), 177–199. https://doi.org/10.1016/j.cogpsych.2010.05.001

Eysenck, M.W., & Keane, M.T. (2020). *Cognitive Psychology: A Student's Handbook* (8th ed.). Psychology Press.

Field, J. (2004). An insight into listeners' problems: too much bottom-up or too much top-down? *System, 32*, 363–377. https://doi.org/10.1016/j.system.2004.05.002

Field, J. (2008a). Revising segmentation hypotheses in first and second language listening. *System, 36*, 35–51. https://doi.org/10.1016/j.system.2007.10.003

Field, J. (2008b). *Listening in the Language Classroom.* Cambridge: CUP.

Field, J. (2019). Second Language Listening: Current Ideas, Current Issues. In J. Schwieter & A. Benati (Eds.), The Cambridge Handbook of Language Learning (Cambridge Handbooks in Language and Linguistics, pp. 283–319). Cambridge University Press. https://doi.org/10.1017/9781108333603.013

Grisoni, L., Miller, T. M., & Pulvermüller, F. (2017). Neural correlates of semantic prediction and resolution in sentence processing. *The Journal of Neuroscience, 37*(18), 4848–4858. https://doi.org/10.1523/JNEUROSCI.2800-16.2017

Grosjean, F. (1980). Spoken word recognition processes and the gating paradigm. *Perception and Psychophysics, 28*(4), 267–283. https://doi.org/10.3758/BF03204386

Grosjean, F. (1985). The recognition of words after their acoustic offsets: evidence and implications. *Perception and Psychophysics, 38*, 299–310. https://doi.org/10.3758/BF03207159

Grosjean, F. (1997). Processing mixed language: Issues, findings and models. In A. M. B. de Groot & J. F. Kroll (Eds.), *Tutorials in bilingualism: Psycholinguistic perspectives* (pp. 225–254). Lawrence Erlbaum Associates Publishers.

Ip, M. H. K., & Cutler, A. (2020). Universals of listening: Equivalent prosodic entrainment in tone and non-tone languages. *Cognition*, *202*: 104311. https://doi.org/10.1016/j.cognition.2020.104311

Katayama, T. (2015). Effect of Phonotactic Constraints on Second Language Speech Processing. *I-Perception*, *6*(6). https://doi.org/10.1177/2041669515615714

Klatt, D. (1980). Speech perception: A model of acoustic-phonetic analysis and lexical access. In R. Cole (Ed.), *Perception and Production of Fluent Speech* (pp. 243–288). Hillsdale, NJ: Erlbaum.

Leonard, K.R. (2019). Examining the relationship between decoding and comprehension in L2 listening. *System*, 87. https://doi.org/10.1016/j.system.2019.102150

Lucas, M. (1999). Context effects in lexical access: A meta-analysis. *Memory & Cognition*, *27*, 385–398. https://doi.org/10.3758/BF03211535

Marslen-Wilson, W. & Tyler, L. (1980). The Temporal structure of spoken language understanding: The perception of words in sentences. *Cognition, 8*, 1–71. https://doi.org/10.1016/0010-0277(80)90015-3

Matthews, J., & Cheng, J. (2015). Recognition of high frequency words from speech as a predictor of L2 listening comprehension. *System*, *52*, 1–13. https://doi.org/10.1016/j.system.2015.04.015

Matthews, J., Masrai, A. Lange, K., McLean, S., Alghamdi, E. A., Kim, Y. A., Shinhara, Y., Tada, S. (2023). Exploring links between aural lexical knowledge and L2 listening in Arabic and Japanese speakers: A close replication of Cheng, Matthews, Lange and McLean (2022). TESOL Quarterly, https://doi.org/10.1002/tesq.3212

Mattys, S., White, L., & Melhorn, J. (2005). Integration of multiple speech segmentation cues: a hierarchical framework. *Journal of Experimental Psychology, 4*, 477–500. https://psycnet.apa.org/doi/10.1037/0096-3445.134.4.477

McClelland, J., & Elman, J. (1986). The TRACE model of speech perception. *Cognitive Psychology, 18*, 1–86. https://doi.org/10.1016/0010-0285(86)90015-0

McQueen, J. (1998). Segmentation of continuous speech using phonotactics. *Journal of Memory and Language, 39*, 21–46. https://doi.org/10.1006/jmla.1998.2568

Nation, I., & Beglar, D. (2007). A vocabulary size test. *The Language Teacher, 31*(7), 9–13.

Norris, D. (1994). Shortlist: a connectionist model of continuous speech recognition. *Cognition, 52*, 189–234. https://doi.org/10.1016/0010-0277(94)90043-4

Norris, D., McQueen, J., & Cutler, A. (2000). Merging information in speech recognition: Feedback is never necessary. *Behavioral and Brain Sciences, 23*, 299–370. https://doi.org/10.1017/S0140525X00003241

Sanders, L.D., Neville, H.J., & 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*(3), 519–30. https://doi.org/10.1044/1092-4388(2002/041)

Schneider, W., Eschman, A., & Zuccolotto, A. (2002). *E-Prime Users' Guide.* Pittsburgh: Psychology Software Tools Inc.

Shockey, L. (2003). *Sound Patterns of Spoken English.* Blackwell Publishing.

Snijders, T., Kooijman, V., Cutler, A. & Hagoort, P. (2007). Neurophysiological evidence of delayed segmentation in a foreign language. *Brain Research*, *1178*, 106–113. https://doi.org/10.1016/j.brainres.2007.07.080

Tsui, A., & Fullilove, J. (1998). Bottom-up or top-down processing as a discriminator of L2 listening performance. *Applied Linguistics, 19*(4), 432–451. https://doi.org/10.1093/applin/19.4.432

Weber, A. (2000). The role of phonotactics in the segmentation of native and non-native continuous speech. In A. Cutler, J. M. McQueen, & R. Zondervan (Eds.), *Proceedings of SWAP, Workshop on Spoken Word Access Processes*. MPI for Psycholinguistics. https://hdl.handle.net/11858/00-001M-0000-0013-2EE6-6

Zoghlami, N. (2015). *Foreign language listening: bottom-up and top-down processes—issues for EFL teaching and research*. Doctoral dissertation. Université Paris 8, Paris, France.