Published in Journal of Phonetics 49, 67-76, 2015,
which should be used for any reference to this work

The role of vowel phonotactics in native speech segmentation

Katrin Skoruppa [a,*], Andrew Nevins [b], Adam Gillard [b], Stuart Rosen [c]

[a] *German Seminar, University of Basel, Switzerland*
[b] *Department of Linguistics, UCL, UK*
[c] *Department of Speech, Hearing and Phonetic Sciences, UCL, UK*

A B S T R A C T

Numerous studies have shown that listeners can use phonological cues such as word stress and consonant clusters to find word boundaries in fluent speech. This paper investigates whether they can also use language-specific restrictions on vowel positioning for native speech segmentation. We show that English adults can exploit the fact that typical English words do not end in a lax vowel (e.g. [*diːtʊ]) in order to segment unknown words in a nonsense phrase-picture matching task, in contrast to the null results in prior studies using lexical tasks. However, they only used this cue in quiet listening conditions, and not in the presence of background noise. Thus, like consonant clusters, the lax vowel constraint is vulnerable in adverse listening conditions.

Keywords: Speech segmentation ; Phonological cues ; Phonotactics ; Lax vowel constraint

1. Introduction

Breaking down continuous speech into word units is challenging because there are no clear acoustic correlates for word boundaries (Klatt, 1980). Recent work has shown that listeners are sensitive to a variety of cues signalling word boundaries, including lexical viability (e.g. Norris, McQueen, Cutler, Butterfield, & Kearns, 2001), transitional probabilities (e.g. Saffran, Newport, & Aslin, 1996) and phonological cues, including stress (e.g. Cutler & Norris, 1988) and phonotactic (i.e. sound positioning) regularities (e.g. McQueen, 1998; Mattys, White, & Melhorn, 2005). With regard to phonological cues, English listeners will generally assume that words begin with stressed syllables (Cutler & Norris, 1988), and therefore detect embedded words (like *mint*) more easily before weak syllables (as in the string ˈ*mintesh*) than before stressed syllables (as in the string *minˈtayve*). Mattys et al. (2005) show that in a cross-modal fragment priming task, English listeners also take consonant clusters into account for the purposes of word segmentation, and weigh them more heavily than stress cues. In their study, fragments (e.g. [kʌstə]) primed corresponding words (e.g. *customer*) more effectively when they were embedded in consonant clusters such as [mk] that rarely appear within words (here,[gɑste**mk**ʌstə]) than in clusters such as [ŋk] in [gɑsteŋ**k**ʌstə].

Most of this work on phonotactic constraints in word segmentation has focused on consonants, whereas vowels have received less attention. On the one hand, consonants and vowels may indeed play different roles in speech segmentation. Nespor, Pena, and Mehler (2003) claim that consonants are more important for lexical access, whereas vowels are crucial for phrasal intonation and its relation with syntactic structure. Indeed, English adults seem to use consonants, but not vowels, during lexical access in tasks involving word reconstruction (Van Ooijen, 1996; Sharp, Scott, Cutler & Wise, 2005), word learning (Creel, Aslin, & Tanenhaus, 2006) and auditory priming (Delle Luche et al., 2014). Thus, one may indeed hypothesize that native language speech segmentation can be influenced by consonant phonotactics, but not by vowel phonotactics. However, given that some researchers have found vowel effects under certain conditions (e.g. for vowel-initial bisyllabic words, Delle Luche et al., 2014), the distinction between vowels and consonants may be less categorical, and there may be some influence of vowels in tasks involving lexical processing.

The present study focuses on positional restrictions on English vowels that listeners can potentially exploit for the purposes of word segmentation. Specifically, they could make use of the fact that English words can end in tense vowels like [u] in *shoe* or [iː] in *tea*,

* Correspondence to: German Seminar, University of Basel, Nadelberg 4, CH-4051 Basel.
E-mail address: kskoruppa@gmail.com (K. Skoruppa).

but typically not in lax vowels like [ʊ] or [ɪ].[1] Two previous studies (Newman, Sawusch, & Wunnenberg, 2011; Norris et al., 2001) have investigated the use of this constraint, amongst other phonological cues, in a word spotting task. In both studies, there was no effect of vowel quality; and listeners were as fast and accurate in identifying vowel-initial words like *apple* in strings with preceding lax vowels like *vuhfapple* [vʊfæpl] as in sequences with tense vowels like *veefapple* [viːfæpl], although the lax vowel constraint should facilitate segmentation in the former (i.e., [vʊf æpl] is a phonotactically legal possibility, but [vʊ fæpl] is not). However, as null results are hard to interpret, it is not clear whether this failure can be attributed to the lexical nature of the task, or to differences in the importance of consonants and vowels for speech segmentation. Specifically Mattys et al. (2005) have shown that lexical cues, if present, outweigh phonotactic cues in a cross-modal fragment priming task. Therefore, the fact that *apple* was the only possible lexical item in the strings described above may have cancelled out any phonotactic effect. Thus, despite the two null results in lexical tasks of Norris et al. (2001) and Newman et al. (2011), the lax vowel constraint could still be active during the segmentation of unknown words. Indeed, Newport and Aslin (2004) show that listeners can compute transitional probabilities both over consonants and over vowels when exposed to new artificial languages, suggesting that vowels are not completely ignored in segmentation tasks. Thus, Experiment 1 of the present study investigates the role of the lax vowel constraint in the segmentation of native speech in a non-lexical phrase-picture matching task.

Most of the studies on word segmentation cited above only investigate performance under optimal, quiet listening conditions. However, most natural speech processing is done in less ideal situations, and listeners' cue weighting in many speech and language processing tasks changes drastically with background noise (for a review see Mattys, Davis, Bradlow, & Scott, 2012). Specifically, Mattys et al. (2005) found that although listeners ranked consonant clusters higher than stress cues in quiet, this hierarchy was reversed when strong background noise was added. Thus, the second experiment investigates the lax vowel constraint under challenging listening conditions, that is, in stimuli presented in background noise, in order to find out whether it is vulnerable to background noise as well.

2. Experiment 1

2.1. Rationale

We designed our task such that listeners would have to segment unknown nonsense syllable sequences in a naturalistic environment. Listeners are told that they would learn words for new alien creatures and their colours in a language game, and that they would hear novel adjective-noun phrases describing the creatures, analogous to real English phrases like *yellow car* or *red balloon*.

Each trial begins with a *presentation phase*, a picture of a multicoloured alien creature, accompanied by a three-syllable nonsense sequence in a carrier phrase (e.g. *This is a* [naɪvʊʃaʊ]). An example is shown in Fig. 1 (left). We then test (in the *choice phase*) whether the quality of the middle vowel influenced how listeners break these sequences up into words. Therefore, we present a different sequence, and ask listeners to pick the corresponding creature (e.g. *Where is the* [naɪzʌteɪ]). Listeners can choose among three alternatives: (a) a different alien in the same color, (b) the same alien in a different color and (c) a different alien in a different color (see example in Fig. 1 (right)).

In the example above, the new sequence differs in the last two syllables from the first sequence, which leads to two different segmentation possibilities and answer patterns. First, listeners could segment the sequences into *adjective* [naɪ] and *noun* [vʊʃaʊ] during presentation, and *adjective* [naɪ]+*noun* [zʌteɪ] during the choice phase. A real word analogue for this possibility would be *blue* +*giraffe* and *blue*+*raccoon.* In this case, they should choose a different creature with the same color (henceforth, “early boundary response”). Second, they could segment the sequences into *adjective* [naɪvʊ]+*noun* [ʃaʊ] (presentation) and *adjective* [naɪzʌ]+*noun* [teɪ] (choice). A real word analogue for this second possibility would be *orange*+*cat* and *olive*+*dog*. In this case, they should choose a different creature with a different color (henceforth, “late boundary response”). Crucially, however, we expect listeners to avoid this latter segmentation since it contravenes the positional restrictions of English phonotactics by yielding a lax vowel at the end of a word ([naɪzʌ]). For comparison purposes, we also test listeners' response to (otherwise identical) items with tense middle vowels (e.g. [naɪvuːʃaʊ]), where these restrictions do not apply. In this case, the late boundary response [naɪvuː]+[ʃaʊ] is perfectly phonotactically legal in English. Thus, we expect listeners to show less late boundary responses when the middle vowel is lax than when the middle vowel is tense.

Finally, the third possibility (same alien, different color) serves as a control as to whether listeners really listen to the stimuli and segment them; since this response would not be expected under any segmentation strategy, it is counted as an error. In order to avoid the formation of response strategies, we also use sequence pairs that differ in the first two syllables (which reverses the expected responses), and fillers differing in only one or in all three syllables, which all lead to different expected answer patterns (as will be explained in Section 2.3.1).

In order to make the stimuli sound natural and in order to avoid listeners treating the nonsense sequences as a single, trisyllabic word, we recorded them with two primary stresses, one on the first and one on the last syllable (e.g. [ˈnaɪvʊˈʃaʊ]), and manipulated the duration, intensity and pitch contour of all vowels such that they corresponded to segmentally similar real word model phrases

[1] To our knowledge, the only exceptions to this constraint are antiquated pronunciations of words ending in *–y* (e.g. [sɪtɪ] for *city* in older British RP) and dialectal variants of words ending in *–er* (e.g. [evʌ] for *ever*).

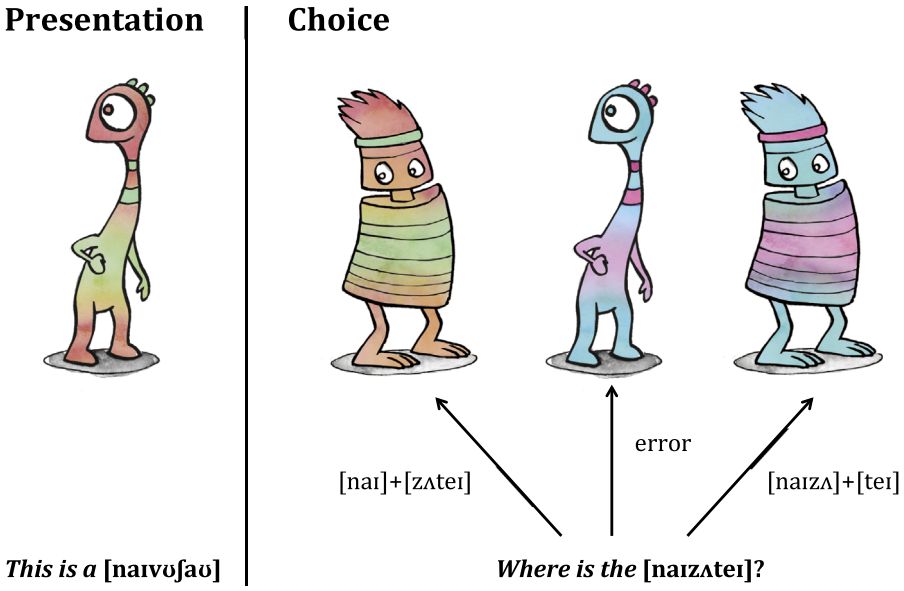

Fig. 1. Example of a test picture set.

recorded alongside with them (e.g. *navy show*). Note that this manipulation eliminated any differences in vowel length between lax and tense vowels, such that vowel quality is the only usable cue to distinguish between them.

These stress cues on their own would favour a late boundary response (e.g. [ˈnaɪvuː]+[ˈʃaʊ]), since English listeners tend to interpret stressed syllables as word onsets (Cutler & Norris, 1988). Thus, any reduction in late boundary responses in the lax vowel condition would show an effect of vowel phonotactics overriding these stress cues, thus confirming the hierarchy established by Mattys et al. (2005), with phonotactic cues being more important than stress cues under good listening conditions.

2.2. Methods

2.2.1. Participants

Sixteen monolingual British English speakers aged 18–29 years participated. They had no history of speech, language or hearing impairment according to self-reported assessment.

2.2.2. Stimuli

2.2.2.1. Auditory stimuli – design. Twenty-four experimental item sets were constructed to test participants' segmentation in two steps (a *presentation phase* and a *choice phase*). In order to assess the role of vowel phonotactics, two versions of each item were created, one with a tense vowel and one with a lax vowel in the middle syllable.

For the presentation phase, 24 three-syllable items (e.g. [naɪvu͟ːʃaʊ] or [naɪvʊ͟ʃaʊ]) were constructed. The second vowel alternated between tense and lax versions of the otherwise phonetically similar vowel pairs ([iː]-[ɪ], [ɑː]-[æ], [uː]-[ʊ], [ɔː]-[ʌ]). For the choice phase, two syllables were changed in each item (e.g. [naɪzɔːteɪ] or [naɪzʌteɪ]), with the same alternations in the second vowel.

For half of the items, the first two syllables changed, while for the other half, the last two syllables changed. For presentation and test phase stimuli taken together, the experimental items contained the same number of [iː]-[ɪ], [ɑː]-[æ], [uː]-[ʊ], [ɔː]-[ʌ] vowel pairs. A list of all experimental items can be found in Appendix A (Table A1).

All items had a trisyllabic sequence of open syllables (e.g. CV-CV-CV) and contained only British English phones. Care was taken to ensure that no sound appeared twice within the same item, and that none of the individual syllables or syllable combinations corresponded to a real English word.

Furthermore, 16 trisyllabic filler items (e.g. [pɔɪkʌzaɪ]) were constructed according to the same criteria. Since they were not designed to test segmentation, vowels did not alternate in these items. Half of the fillers contained a lax vowel and half of them contained a tense vowel in the second syllable. Items in the choice phase were derived by changing the first syllable ($n=6$), the third syllable ($n=6$), or all syllables ($n=4$). A full list of filler items can be found in Appendix A (Table A2).

Finally, six training sets containing a mixture of real and nonsense words were designed to familiarize participants with the procedure (see Table A3 in Appendix A).

2.2.2.2. Auditory stimuli – recordings. In order to achieve a natural intonation, real adjective-noun phrases containing similar sounds were recorded alongside all experimental and filler items (e.g. the model phrase *navy show* was used as an acoustic template for the pair [naɪvuːʃaʊ] – [naɪvʊʃaʊ]). A full list of these model phrases can be found in Appendix A (Tables A1 and A2).

Items were recorded in two different carrier phrases for presentation (e.g. *This is a* [ˈnaɪvuːˈʃaʊ], *And here's another* [ˈnaɪvuːˈʃaʊ]), and in one carrier phrase for choice phase (e.g. *Where is the* [ˈnaɪzɔːˈteɪ]). The same carrier phrases were used for the lax and tense

Table 1
Example of a recording script.

Phase	Item	Example
Presentation 1	*Model*	This is a navy show
	Lax	This is a [ˈnaɪvʊˈʃaʊ]
	Tense	This is a [ˈnaɪvuːˈʃaʊ]
Presentation 2	*Model*	And here's another navy show
	Tense	And here's another [ˈnaɪvuːˈʃaʊ]
	Lax	And here's another [ˈnaɪvʊˈʃaʊ]
Choice	*Model*	Where is the noisy tie
	Lax	Where is the [ˈnaɪzʌˈteɪ]
	Tense	Where is the [ˈnaɪzɔːˈteɪ]

versions and the model phrase in each experimental set. Experimental sets were recorded one by one. The model phrase was always recorded first and the order of lax and tense stimuli was varied. Table 1 provides an example of a recording script for an experimental set. The training stimuli were recorded in the same carrier sentences, but without model phrases on which to base them.

The third author, a phonetically trained male native speaker of British English, recorded all stimuli at least twice. He produced them fluently without pauses, using a similar intonation for model and nonsense phrases.

Recordings were made in a sound-attenuated booth on a standard PC through CoolEdit 2000, using a Røde NT1-A condenser microphone and an Edirol UA-25 USB pre-amplifier. The sampling rate was 44.1 kHz (mono, 16-bit).

2.2.2.3. Auditory stimuli – manipulations. Praat version 5.2.46 (Boersma & Weenink, 2011) was used for all acoustic manipulations. All three-syllable items were cut out of the carrier phrases and segment boundaries were annotated by hand, and then were subsequently moved to the closest zero crossing automatically. In order to ensure that only the second vowel was different between the lax and the tense version, stimuli were cross-spliced. For half of the experimental items, all segments except for this vowel were taken from the lax version, while for the other half, all other segments were taken from the tense version. Furthermore, the duration, intensity and pitch contour of all vowels was set to the values of the respective model phrases, in order to equalize stress cues and to achieve the most natural intonation possible. Filler items were treated in the same fashion, except that they were not cross-spliced. Before concatenating the stimuli with the respective carrier phrases again, both were scaled to a mean intensity of 70 dB in order to avoid unnatural loudness changes. The third author listened to the final versions and repeated the procedure with different recordings if the result sounded unnatural to him.

2.2.2.4. Visual stimuli. Forty pairs of cartoon alien creatures were selected from the pictures used in Van de Vijver and Baer-Henney (2011). Each pair was filled with two combinations of multiple colours that could not readily be associated with a single English color name. Fig. 1 shows an example of a test picture set.

For training, six pictures depicting the real words used in the auditory stimuli and two distractor pictures were taken from the Snodgrass and Vanderwart (1980) battery of line drawings, and four pictures of unfamiliar objects were selected from Internet webpages. If real color adjectives were used in the auditory stimuli, they were filled with the corresponding color ($n = 8$); if nonsense adjectives were used, they were filled with multiple colours ($n = 4$).

2.2.3. Procedure

Prior to testing, participants were told that they were going to learn names for alien colours and creatures that they should memorize. They were informed that they would be tested on their knowledge and would have to identify the right creature among three alternatives via mouse click. They were told that there would be a short training block with real words, that the experiment would take approximately 15 min, and that they could take small breaks in between. After this brief explanation, they were given the opportunity to ask questions, and written informed consent was obtained. Prior to testing, ethical approval was obtained under the auspices of the UCL Research Ethics Committee.

The experiment was run in a quiet room on a laptop with an external USB mouse, and the presentation and data collection were executed in Python 2.6.6 and Pygame 1.9.1. Auditory stimuli were presented through Sennheiser HD 202 headphones at a fixed, comfortable intensity (ca. 65–70 dB SPL).

The experiment began with six training trials (see Table A3). Each trial consisted of a presentation and a choice phase (see Fig. 2 for a schematic example).

During the presentation phase, a picture (e.g. a red balloon) appeared at a random position on the screen, and .5 s later, the first presentation sentence describing it was played (here, *This is a red balloon*). The picture stayed on the screen for 1.5 s after the end of the auditory stimulus, followed by a blank screen of .5 s. Subsequently, the picture appeared again at a different location on the screen, and the second presentation sentence was played (here, *And that's another red balloon*) with the same timing as for the first presentation.

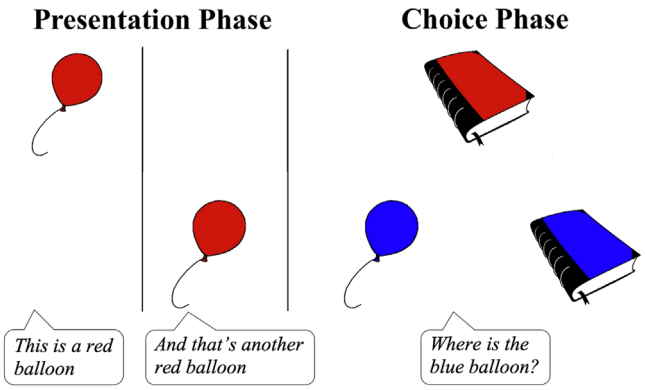

Fig. 2. Schematic example of a training trial.

Table 2
Response and boundary types for experimental items.

Change	Example	Late boundary	Early boundary	Error
Syll. 2+3	[ˈnaɪvʊˈʃaʊ] – [ˈnaɪzʌˈteɪ]	Both different	Same color	Same alien
Syll. 1+2	[ˈʧɛɪgɪˈpɔɪ] – [ˈlaʊθʌˈpɔɪ]	Same alien	Both different	Same color

Finally, during the choice phase, three pictures with different objects and different colours (here, a red book, a blue balloon and a blue book) appeared on the screen in a triangle, with the cursor in the middle. The order of the pictures on the screen was randomized. After .5 s, the choice sentence was played (here, *Where is the blue balloon?*).

After the participant had clicked on one of the objects, a feedback text was displayed on the screen (positive: *Correct!*; negative: *Incorrect! Please try again.*). Training trials were repeated until the participant clicked on the correct object. Test trials had the same structure as training trials, except that no feedback was given, and the next trial started .7 s after the participant's response.

For the test phase, the 24 experimental sets were divided into two lists, each of which contained six sets for which syllables 1 and 2 changed and six sets for which syllables 2 and 3 changed. Half of the participants were presented with the lax versions for the first list and the tense versions for the second list. The opposite was true for the other half of participants. The 40 test items (24 experimental items and 16 filler items) were presented twice in two blocks with different randomized orders for each participant. Participants could take a short break after every 20 trials.

2.3. Results and discussion

2.3.1. Response coding

The responses for experimental items were coded according to the segmentation strategy they revealed. Recall that participants could attribute the second syllable of the three-syllable phrases (e.g. [ˈnaɪvʊˈʃaʊ]) either to the final 'noun' (*early boundary*, here: [naɪ]+[vʊˈʃaʊ]), or to the initial 'adjective' (*late boundary*, here: [ˈnaɪvʊ]+[ʃaʊ]). These two different segmentation strategies lead to different responses during the choice phase. For instance, for the set 'This is a [ˈnaɪvʊˈʃaʊ]'–'Where is the [ˈnaɪzʌˈteɪ]', positing an early boundary after the first syllable ([naɪ]+[vʊˈʃaʊ] and [naɪ]+[zʌˈteɪ]) would lead participants to assume that the adjective stays the same but that the noun is different, and thus choose a different alien creature with the same color during the choice phase. Positing a late boundary after the second syllable ([ˈnaɪvʊ]+[ʃaʊ] and [ˈnaɪzʌ]+[teɪ]) would lead participants to assume that both adjective and noun are different, and thus choose a different alien creature with a different color during the choice phase. If participants chose the third option (the same alien creature with a different color), which did not correspond to any possible segmentation, this was classified as an error. Table 2 shows which response indicates which boundary type, depending on the syllables that were changed between presentation and choice phase.

Recall that English stress rules would disfavour an early boundary, since the noun would have atypical final stress (here, [zʌˈteɪ]). However, for the lax versions, a late boundary would clash with the lax vowel constraint, since the adjective would end in a lax vowel (here, [ˈnaɪzʌ]). We thus expect fewer late boundary responses for the items with tense vowels than for the items with lax vowels in the second syllable.

For filler items, only one response was correct, regardless of segmentation strategy, as summarized in Table 3.

2.3.2. Analysis

We calculated the percentage of late boundary responses over all valid trials, that is, we divided the number of late boundary responses by the total number of late and early boundary responses. Since the outcome data were in the form of proportions, they were analysed with non-parametric Wilcoxon signed rank tests and with mixed effect logistic regression using the software R version 2.15.2 (R Core Team, 2012) and the package lme4, version 0.999375-39 (Bates, Maechler, & Bolker, 2011). Median filler accuracy

Table 3
Correct responses for filler items.

Change	Example	Correct response
Syllable 1	[ˈlɔɪnɪˈdʒaɪ] – [ˈtaʊnɪˈdʒaɪ]	Same alien
Syllable 3	[ˈfaʊmʊˈteɪ] – [ˈfaʊmʊˈdɔɪ]	Same color
All syllables	[ˈfaʊgʊˈpɔɪ] – [ˈfɔɪmɪˈθuː]	Both different

was 87.5% (IQR 25.8), well above chance level (33.3%) in a Wilcoxon test ($V = 136$, $p < .001$), showing that participants had no difficulties with the task.

For the experimental items, errors (7.7% of trials) were excluded from analysis. The proportion of late boundary responses in the remaining trials was analysed using a mixed effect logistic regression model with random intercepts[2] for Participant and Item, and Vowel Type (tense vs. lax) as a fixed effect. This analysis revealed a significant effect of Vowel Type ($z = 2.844$, $p = .004$). Fig. 3 shows the percentage of late boundary responses in both vowel conditions for each individual participant. As expected, there were more late boundary responses in the tense condition (median: 59.2%, Interquartile Range [IQR]: 30.3) than in the lax condition (median: 53.2%, IQR: 22.8). Despite considerable variability in the overall percentage of late boundary responses, this relation held for 13 out of 16 participants.

In conclusion, this first experiment shows that in absence of lexical cues, the lax vowel constraint has a small but consistent effect on participants' segmentation of nonsense phrases, despite the fact that it led to the segmentation of words with the atypical final stress pattern (e.g. [zʌˈteɪ] for [ˈnaɪzʌˈteɪ]). Thus, English adults not only use consonant clusters (Mattys et al., 2005), but also vowel phonotactics as cues for speech segmentation in their native language.

3. Experiment 2

3.1. Rationale

The second experiment tests whether listeners can also use the lax vowel constraint in less ideal conditions. Phonotactics are generally characterized as weak segmentation cues; consonant cluster cues, for example, tend to be overridden by stronger stress cues when listening conditions are degraded by background noise (Mattys et al., 2005). In order to test whether vowel cues are similarly affected by adverse listening conditions, or whether they are more robust to noise, we compare participants' use of the lax vowel constraint in quiet and in noise.

3.2. Methods

3.2.1. Participants

Twenty-four monolingual British English speakers aged 18–36 years participated after giving written informed consent. According to self-reports, they had no history of speech, language or hearing impairment. None of them had taken part in Experiment 1. Some participants were recruited from UCL Speech Sciences courses and the UCL Psychology Subject Pool and received £5 in return.

3.2.2. Stimuli

The auditory stimuli to be used in quiet and the visual stimuli were the same as in Experiment 1. For the auditory stimuli to be used in noise, speech-shaped noise was added at 0 dB signal-to-noise ratio to all test sets and to three training sets used in Experiment 1.

3.2.3. Procedure

The procedure was the same as in Experiment 1, except that (a) the experiment was run in a sound-attenuated booth, (b) the second stimulus presentation ('*And here's another* ….') was omitted and (c) participants heard each test set once in quiet and once in noise. Half of the participants started with the block in quiet, and received the same training as in Experiment 1. Before starting the second block in noise, this group repeated three training items in noise. The other half started with the block in noise, and received half of the training in noise. Before starting the second block in quiet, this group repeated three training items in quiet. Participants were informed when stimuli would be presented in noise.

3.3. Results and discussion

Responses were coded as in Experiment 1. Median filler accuracy was significantly higher in quiet (93.8%, IQR 31.3) than in noise (71.9%, IQR 32.8) in a paired Wilcoxon test ($V = 267$, $p < .001$), showing that noise affected participants' performance. However,

[2] In these and all subsequent mixed effect analyses, more complex models with random slopes for Participants and/or Items were also tested, but abandoned because they did not yield a better fit.

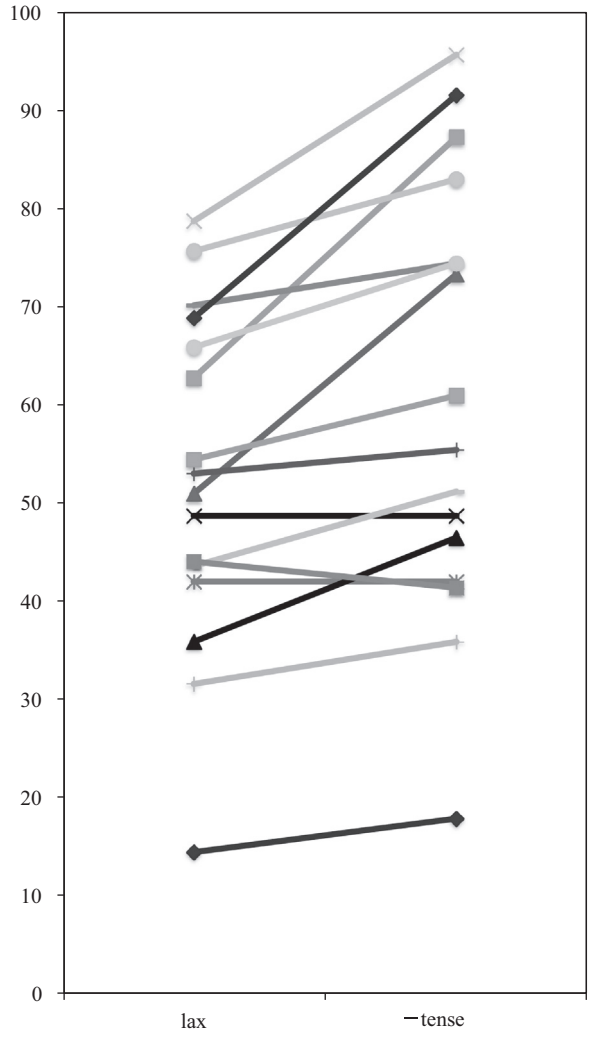

Fig. 3. Percentage of late boundary responses by participant in Experiment 1.

participants were well above chance level in both blocks in Bonferroni-corrected Wilcoxon tests (quiet: $V=296$, $p<.001$; noise: $V=286$, $p<.001$), indicating that they could do the task both in quiet and in noise.

As in Experiment 1, errors (13.6% of trials) were excluded for the experimental items. The proportion of late boundary responses in the remaining trials was analysed using a mixed effect logistic regression model with random intercepts for Participant and Item, and with the fixed effects Vowel Type (tense vs. lax) and Listening condition (quiet vs. noise). This analysis revealed a significant effect of Listening Condition ($z=2.12$, $p=.034$) and an interaction between Listening Condition and Vowel Type ($z=3.15$, $p=.002$). In order to explore this interaction further, responses in quiet and in noise were evaluated in two separate models with random effects for Participant and Items and the fixed effect Vowel Type. These analyses revealed that in quiet, there were significantly more late boundary responses in the tense condition than in the lax condition ($z=3.09$, Bonferroni-corrected $p=.004$), whereas there was no difference between the two conditions in noise ($z=1.41$, Bonferroni-corrected $p>.1$). Fig. 4 shows the percentage of late boundary responses by participant in the four different vowel and listening conditions.

In conclusion, this experiment replicates the effect of the lax vowel constraint on segmentation in quiet that we found in Experiment 1, but does not provide any evidence for an influence of this cue in noisy listening conditions.

4. General discussion

In contrast with prior studies using word spotting tasks (Norris et al., 2001; Newman et al., 2011), we found a small but robust effect of vowel phonotactics (more specifically, the English lax vowel constraint) on nonsense word segmentation in quiet during a phrase-picture matching task in both experiments. Thus, it seems that if lexicality as a factor is removed, vowel phonotactics are important for speech segmentation. Another difference between our task and the more traditional word spotting tasks cited above that

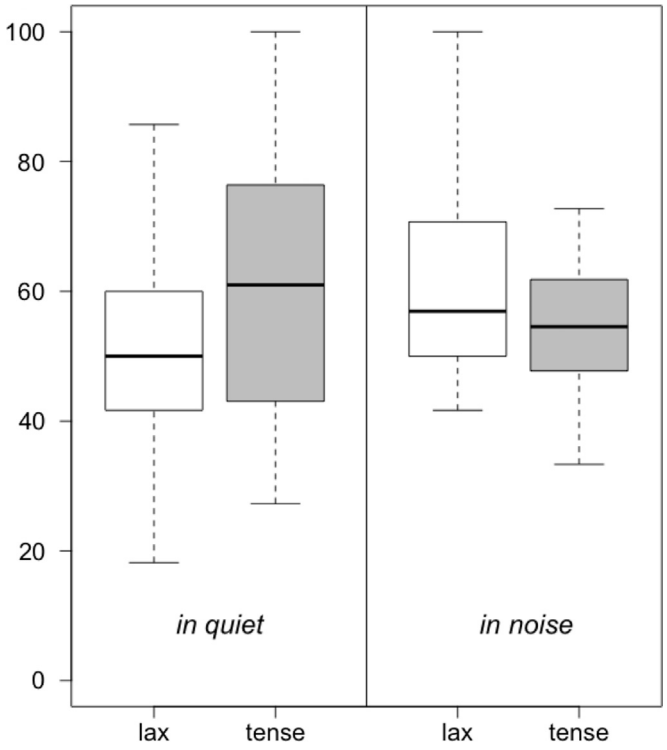

Fig. 4. Boxplots for percentages of late boundary responses by listening condition and vowel type in Experiment 2. Whiskers extend to ±1.5 IQR.

may have played a role is the strong memory component involved in this task, where participants have to remember six nonsense syllables across presentation and choice phases.

It is important to note that there was quite a large amount of individual variability even in the tense vowel condition, where one could expect a strong and uniform stress-driven effect. This may be due to the fact that we included equal numbers of stress-initial and stress-final items in the training phase, which may have overridden participants' natural preference for stress-initial words. Interestingly, however, the effect of vowel quality held well for individual participants despite this variability, showing that the influence of the lax vowel constraint does not depend on the participants' "baseline" response rate.

Thus, contrary to what a very strong version of the hypothesis on the division of labour between consonants and vowels would predict (Nespor et al., 2003), vowel phonotactics play a role in English listeners' speech segmentation. These results tie in well with results from other languages, which have different constraints on vowels that are exploited for segmentation purposes; for instance, vowel harmony – the requirement that all vowels within a word share certain features – has been shown to influence listeners' segmentation of nonsense words in both Finnish (Suomi, McQueen, & Cutler, 1997) and Turkish (Kabak, Maniwa, & Kazanina, 2010).

The second experiment, in which we also found an effect of the lax vowel constraint in quiet, but not in noise, provides further evidence that vowels behave similarly to consonants during native speech segmentation. As for consonant clusters, examined in Mattys et al. (2005), this cue proved to be stronger than stress in quiet, leading to a segmentation solution that accomodates atypical stress-final words, though is vulnerable to background noise. Assessing the role of consonant and vowel phonotactics further in different types of adverse listening conditions would also be of clinical relevance, since it may contribute to understanding how listeners with hearing impairment segment speech. In contrast to stress, which is relatively well perceived and used by listeners with cochlear implants (Spitzer, Liss, Spahr, Dorman, & Lansford, 2009), to our knowledge phonotactic cues have not yet been examined in this population.

Furthermore, the fact that the lax vowel constraint only seems to be active when no lexical cues are present hints to the possibility that it may be playing a more important role in language acquisition than in adult language processing. Although it contains a strong working memory component, our phrase-picture matching task is independent of vocabulary skills and more child-friendly than the word spotting and priming tasks used in previous studies (Mattys et al., 2005; Newman et al., 2011; Norris et al., 2001). Pilot data indicate that it can be used with school-age children, and potentially also with clinical populations such as paediatric cochlear implant users.

Acknowledgements

This work was funded by the European Commission (Marie Curie Actions FP7-PEOPLE-2009-IEF 253782) and by the British Academy Rounds 2009 and 2010. We thank Veronika Smiskova, Helai Mussa, Emily Whitters and Rachel Nicholson for help with

recruiting and running participants for these and previous versions of the experiments. We are grateful to Steve Nevard for help with recording the auditory stimuli, to Anita Wagner, Paul Iverson and Volker Dellwo for help with Praat scripting and to Noor Ali for colouring the visual stimuli.

Appendix A

See Tables A1–A3.

Table A1
Experimental Items.

Presentation phase		Choice phase		Changed syllables
Item in IPA	**Model phrase**	**Item in IPA**	**Model phrase**	
ʧeɪg(iː/ɪ)pɔɪ	shaky pie	laʊθ(ɑː/æ)pɔɪ	Lousy pie	S1, S2
dɔɪz(ɔː/ʌ)ʃaʊ	dicey show	naɪθ(ɑː/æ)ʃaʊ	Mousy show	S1, S2
gɑːf(uː/ʊ)ðɔ	goofy jaw	θuːn(ɑː/æ)ðɔː	Thorny jaw	S1, S2
teɪd(ɑː/æ)mɔɪ	tidy May	nɑɪð(uː/)mɔɪ	Navy May	S1, S2
lɔɪz(ɔː/ʌ)θəʊ	lazy thigh	faʊr(iː/ɪ)θəʊ	Foamy thigh	S1, S2
taʊθ(ɑː/æ)ʃɔɪ	toothy show	giːv(ɔː/ʌ)ʃɔɪ	Beefy show	S1, S2
fɔɪn(ɑː/æ)zaɪ	phoney thigh	laʊθ(iː/ɪ)zaɪ	Lacy thigh	S1, S2
nɔɪd(ɑː/æ)zaʊ	meaty vow	vəʊð(iː/ɪ)zaʊ	Hazy vow	S1, S2
gaʊð(ɑː/æ)vɔɪ	cosy vow	ʧeɪr(iː/ɪ)vɔɪ	Shiny vow	S1, S2
jiːθ(uː/)dʒaʊ	leafy joy	zɔːð(iː/)dʒaʊ	Saucy joy	S1, S2
ʧɔɪg(iː/ɪ)jaʊ	shaky row	kaɪð(uː/ʊ)jaʊ	Cagey row	S1, S2
zɔɪg(iː/ɪ)θaʊ	soapy thigh	ʃɔɪd(ɑː/æ)θaʊ	Shady thigh	S1, S2
naɪv(uː/ʊ)ʃaʊ	navy show	naɪz(ɔː/ʌ)teɪ	Noisy tie	S2, S3
taʊd(ɑː/æ)dʒaɪ	tidy joy	taʊr(iː/ɪ)pɔɪ	Tiny pie	S2, S3
jɑːf(uː/ʊ)kaɪ	leafy cow	jɑːg(iː/ɪ)mɔɪ	Leaky May	S2, S3
θeɪn(ɑː/æ)vəʊ	tiny vow	θeɪg(iː/ɪ)lɔɪ	Shaky lie	S2, S3
mɔɪθ(uː/ʊ)faʊ	mousy foe	mɔɪg(iː/ɪ)ʧəʊ	Mighty show	S2, S3
θuːd(ɑː/æ)ʃɔɪ	seedy show	θuːr(iː/ɪ)gaʊ	Thorny guy	S2, S3
taʊj(iː/ɪ)fɔɪ	tiny foe	taʊg(ɑː/æ)zaɪ	Tidy joy	S2, S3
laʊz(ɔː/)θəʊ	lousy thigh	laʊð(ɑː/æ)nɔɪ	Lacy neigh	S2, S3
zeɪg(ɑː/æ)dʒaʊ	soapy joy	zeɪr(iː/ɪ)ʧɔɪ	Shiny shoe	S2, S3
jiːv(ɔː/ʌ)zaʊ	leafy sow	jiːg(ɑː/æ)θɔɪ	Leaky thigh	S2, S3
θaʊg(iː/ɪ)vɔɪ	shaky foe	θaʊn(ɑː/æ)ʧaɪ	Shiny show	S2, S3
dɔɪð(uː/ʊ)vɑː	dozy fee	dɔɪz(ɔː/ʌ)θiː	Dicey sea	S2, S3

Both tense and lax vowel versions are given in brackets.

Table A2
Filler items.

Presentation phase		Choice phase		Changed syllables
Item in IPA	**Model phrase**	**Item in IPA**	**Model phrase**	
lɔɪnɪdʒaɪ	rainy joy	taʊnɪdʒaɪ	Tiny joy	S1
ʧeɪgɪːzɔː	shaky saw	laʊgiːzɔː	Leaky saw	S1
pɔɪkæfaʊ	poky foe	ʧeɪkæfaʊ	Shaky foe	S1
teɪnɑːθəʊ	tiny thigh	ʃaʊnɑːθəʊ	Shiny thigh	S1
pɔɪfuːθiː	beefy sea	gaʊfuːθi	Goofy sea	S1
riːmʌʧeɪ	roomy jay	gɑːmʌʧeɪ	Corny jay	S1
dɔɪðɔːvəʊ	dozy vow	dɔɪðɔːdʒaɪ	Dozy joy	S3
faʊmʊteɪ	foamy tie	faʊmʊdɔɪ	Foamy day	S3
gɑːfuːlɔɪ	goofy lie	gɑːfuːθeɪ	Goofy thigh	S3
pɔɪkʌʧeɪ	poky joy	pɔɪkʌzaɪ	Poky jay	S3
kaɪðiːʃɔɪ	cosy show	kaɪðiːtaʊ	Cosy tie	S3
nɔɪvæθəʊ	navy thigh	nɔɪvætʃeɪ	Navy jay	S3
lɔɪðːfaʊ	lousy foe	dɑːzɔːkaɪ	Dozy guy	all
ʃaʊgʊpɔɪ	shaky pie	fɔɪmɪθuː	Foamy shoe	all
ðɔːriːdʒaʊ	thorny jay	gɑːðuːʧəʊ	Gauzy show	all
mɔɪtɪgaʊ	mighty guy	naɪzʌθeɪ	Noisy thigh	all

Table A3
Training sets.

Set number	Presentation phase	Choice phase
1	red balloon	Blue balloon
2	yellow chair	Yellow shoe
3	pink [dɑːˈlaʊ]	Pink [kaɪˈlaʊ]
4	[ˈnaɪlɔɪ] train	[ˈnaɪθəʊ] house
5	[ˈʃɔɪvəʊ] car	[ˈʃɔɪriː] car
6	green [zəːˈgiː]	Black [vuːˈgiː]

References

Bates, D., Maechler, M., & Bolker, B. (2011). *lme4: Linear mixed-effects models using S4 classes*. R package version 0.999375-39. ⟨http://CRAN.R-project.org/package=lme4⟩.

Boersma, P. & Weenink, D. (2011). Praat: doing phonetics by computer [Computer program] http://www.praat.org/.

Creel, S. C., Aslin, R. N., & Tanenhaus, M. K. (2006). Acquiring an artificial lexicon: Segment type and order information in early lexical entries. *Journal of Memory and Language*, *54*, 1–19.

Cutler, A., & Norris, D. (1988). The role of strong syllables in segmentation for lexical access. *Journal of Experimental Psychology: Human Perception and Performance*, *14*, 113–121. Delle

Luche, C., Poltrock, S., Goslin, J., New, B., Floccia, C., & Nazzi, T. (2014). Differential processing of consonants and vowels in the auditory modality: A cross-linguistic study. *Journal of Memory and Language*, *72*, 1–15.

Kabak, B., Maniwa, K., & Kazanina, N. (2010). Listeners use vowel harmony and word-final stress to spot nonsense words: A study of Turkish and French. *Laboratory Phonology*, *1*, 1. Klatt,

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

Mattys, S. L., Davis, M. H., Bradlow, A. R., & Scott, S. K. (2012). Speech recognition in adverse conditions: A review. *Language and Cognitive Processes*, *27*(7–8), 953–978.

Mattys, S. L., White, L., & Melhorn, J. F. (2005). Integration of multiple speech segmentation cues: A hierarchical framework. *Journal of Experimental Psychology: General*, *134*(4), 477–500.

McQueen, J. M. (1998). Segmentation of continuous speech using phonotactics. *Journal of Memory and Language*, *39*, 21–46.

Nespor, M., Pena, M., & Mehler, J. (2003). On the different roles of vowels and consonants in speech processing and language acquisition. *Lingue e Linguaggio*, *2*, 203–229. Newman, R. S., Sawusch, J. R., & Wunnenberg, T. (2011). Cues and cue interactions in segmenting words in fluent speech. *Journal of Memory and Language*, *64*(4), 460–476. Newport, E. L., & Aslin, R. N. (2004). Learning at a distance I. Statistical learning of non-adjacent dependencies. *Cognitive Psychology*, *48*(2), 127–162.

Norris, D., McQueen, J. M., Cutler, A., Butterfield, S., & Kearns, R. (2001). Language-universal constraints on speech segmentation. *Language and Cognitive Processes*, *16*, 637–660. R Core Team (2012). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. ⟨http://www.R-project.org/⟩.

Saffran, J. R., Newport, E. L., & Aslin, R. N. (1996). Word segmentation: The role of distributional cues. *Journal of Memory and Language*, *35*, 606–621.

Sharp, D. J., Scott, S. K., Cutler, A., & Wise, R. J. S. (2005). Lexical retrieval constrained by sound structure: The role of the left inferior frontal gyrus. *Brain & Language*, *92*, 309–319.

Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning and Memory*, *6*(2), 174–215.

Spitzer, S., Liss, J., Spahr, T., Dorman, M., & Lansford, K. (2009). The use of fundamental frequency for lexical segmentation in listeners with cochlear implants. *The Journal of the Acoustical Society of America*, *125*(6), EL236–EL241.

Suomi, K., McQueen, J. M., & Cutler, A. (1997). Vowel harmony and speech segmentation in Finnish. *Journal of Memory and Language*, *36*, 422–444.

Van de Vijver, R., & Baer-Henney, D. (2011). Acquisition of voicing and vowel alternations in German. In N. Danis, K. Mesh, & H. Sung (Eds.), *Proceedings of BUCLD 35*. Somerville: Cascadilla Press.

Van Ooijen, B. (1996). Vowel mutability and lexical selection in English: Evidence from a word reconstruction task. *Memory & Cognition*, *24*, 573–583.