Syllabification, Sonority, and Spoken Word Segmentation: Evidence from Word-Spotting

Jason Bishop and Kristen Toda
(j.bishop@ucla.edu, ktoda@ucla.edu)

1 Introduction[1]

The present study concerns itself with the kinds of linguistic information listeners bring to bear on the problem of word segmentation – i.e., the identification of word boundaries in the continuous speech stream (which contains few explicit boundaries). In particular, we explore the extent to which listeners' on-line segmentation behavior exhibits biases related to syllable structure. We first describe some of the important findings relevant to word segmentation processes, and then present a novel experiment testing listeners' sensitivity to sonority.

1.1 Factors influencing segmentation

Much of what we have learned about how (adult) human listeners segment speech into words has come from the word-spotting task (Cutler & Norris 1988, see also McQueen 1996). In a word-spotting experiment, listeners must identify a word such as *apple* embedded in a nonsense string such as *vuffapple*, and do so as quickly as possible. This task is informative as to listeners' segmentation of the string, since performance depends crucially on a unique parsing of that string. It is known that, when speakers are kind enough to provide explicit phonetic cues to boundaries (Newman et al. 2011), listeners use them. However, listeners perform the task non-randomly even without such cues, indicating that they draw upon non-signal based information as well.

Among the information listeners are known to use is knowledge of their lexicon. One example is the reliance on the statistically dominant word-level prosodic patterns of their language. Using this "Metrical Segmentation Strategy" (MSS; Cutler & Norris 1988) English speaking-listeners will posit word onsets at the onset of strong syllables, since the majority of the English lexicon contains words that begin with strong syllables. In fact, the MSS has been verified as an important segmentation strategy in English and a number of other languages (see McQueen 1996 for a summary). Listeners are also known to exploit their lexical knowledge regarding possible sequences of segments, as their segmentations exhibit the avoidance of phonotactic violations. For example, Dutch listeners are faster to spot *rok*, ("skirt"), in [fim.rɔk] than in [fi.drɔk] because in the first case, the [mr] is an unattested word onset (suggesting a boundary must be present). In addition to sequences, knowledge of possible units is exploited as well. For example, English speakers have difficulty spotting *apple* in *fapple* compared with *vuffapple*, because it would leave a single consonant unparsed, and a single consonant cannot

[1] This paper was presented at (and will appear in the proceedings of) the 47th Meeting of the Chicago Linguistics Society (7-9 April, 2011). The authors are grateful to Arthur Samuel, Robert Daland, and members of the UCLA Phonetics Lab for helpful comments and suggestions.

constitute a word or syllable in the language (the Possible Word Constraint (PWC); Norris et al. 1997). However, it is known that in languages in which this is not a categorical impossibility, listeners can parse out a single consonant (Hanulikova et al. 2010). Thus, on-line segmentation routines seem to be sensitive to various kinds of patterns in the lexicon, at least when they represent dramatic asymmetries (MMS), or what is altogether unattested (phonotactic violations, PWC).

Whereas the basic sensitivity to facts about one's own lexicon is likely language-universal, there is another strategy that is claimed to be language-specific, namely the use of syllable structure. One of the earliest findings in the study of spoken word segmentation was that French, but not English-speaking listeners show segmentation biases that look very much like their off-line syllabification preferences. In their classic findings, Cutler et al. (1986) demonstrated this asymmetry using the sequence monitoring task, in which French listeners were shown to detect "BA" more quickly in *ba.lance* than in *bal.con*, and "BAL" in *bal.con* than in *ba.lance*. Native English speakers showed no such "cross-over" pattern with analogous English stimuli. It has since been argued (summarized in Cutler et al. 2001) that this difference in the use of the syllable reflects the rhythm class of the language. French is syllable-timed, and thus segmentation can exploit syllable structure; English is stress-timed, and so the syllable-based segmentation routine is unavailable.

In the present study, we reexamined this rhythm class hypothesis by probing English-speaking listeners' sensitivity to syllable structure in word-spotting. We chose a property of syllables that studies have not yet explored in on-line segmentation, namely sonority patterns. We chose sonority to test because it is a phonological property known to be active in off-line, explicit segmentations by English-speaking listeners. For example, in syllabification tasks, English-speaking listeners show a strong bias towards placing the consonant in a /…VCV…/ sequence (e.g., *melon*) as a coda to the first syllable rather than an onset to the second one. The opposite pattern is exhibited when the consonant is an obstruent (e.g., Treiman & Danis 1988, and, for a recent large-scale replication, Eddington et al. to appear). Thus, in the experiment presented below, we tested listeners' ability to spot a word like *absent* in (1) versus (2):

(1) *jeemabsent* (/ʤimæbsənt/)
(2) *jeebabsent* (/ʤibæbsənt/)

Note that none of the strategies discussed above makes a clear prediction about how listeners should behave in this case, as neither a word boundary before or after the crucial consonant would produce a phonotactic violation. Further, in both cases, the MSS should bias the listener to choose, correctly, /æ/ as the word-initial syllable nucleus, but the MSS does help the listener make a decision regarding how to parse the preceding consonant. If listeners are sensitive to sonority as they are in off-line syllabification studies, however, they should spot *absent* more quickly in (1), since they will prefer to posit a word boundary that parses the /m/ as a coda, rather than as an onset to the next vowel. Conversely, in (2), listeners' bias will be in the direction of parsing the obstruent /b/ as an onset rather than a coda, requiring a re-parse in order identify the vowel-initial word *absent*. Of course, if listeners make use of segmentation strategies that are only based on their language's rhythm class (i.e., they only attend to stress), their word-spotting, like their sequence monitoring in Cutler and Norris's study, should be unaffected.

In addition to the intervocalic consonant's status as a sonorant or obstruent, we also manipulated the preceding vowel to be either tense or lax. It is known that this distinction, like sonority, predicts off-line syllabifications (Eddington et al. to appear). Although it has already

Table 1. Example of a SW and a WS target word in each of the four preceding vowel and sonority conditions.

	SW		WS	
C-Context	Tense	Lax	Tense	Lax
Obstruent	*jee[b]absent*	*ni[b]absent*	*shaw[v]eject*	*shi[v]eject*
Sonorant	*jee[m]absent*	*ni[m]absent*	*zaw[m]eject*	*ze[m]eject*

been found not to influence segmentation in word-spotting (Norris et al. 2001, Newman et al. 2011), we nonetheless included this manipulation because of the possibility that it might interact with sonority.

2 Experiment: Sonority-based biases

2.1 Methods

2.1.1 Stimuli

2.1.1.1 Design

Stimuli for a word-spotting experiment were based on a set of 50 disyllabic English target words: 25 with a strong-weak (SW) and 25 with a weak-strong (WS) stress pattern. For each of the 50 targets, such as *absent*, two different CV contexts were chosen: one in which the vowel was tense (/i, u, ɑ, o, eɪ, aɪ, oɪ/), and one in which it was lax (/ɪ, ɛ, æ, ʌ, ʊ/). This resulted in 100 strings of the form CV+target.

Two versions of each of these 100 strings were then created: one in which an obstruent consonant (/b,d,g,ʤ,θ,f,ʃ,s,z,v/, or /ð/) straddled the boundary between CV and target, and one in which a sonorant (/m, n/ or /l/), straddled the boundary. This resulted in 200 forms like the examples in Table 1. The obstruents used were selected for their lack of relevant allophonic variation (e.g., voiceless stops were not used because listeners are known to use aspiration as a cue to syllabic affiliation in both on- and off-line studies (e.g., Kirk 2001, Eddington et al. to appear). Further, for each string, the selection of consonant was constrained by the need to prevent the occurrence of any additional, especially embedded, words. For example, /ʃ/ was chosen to precede the target word *echo* rather than /g/, because adding /g/ would result in the word *gecko*. Similarly, /m/ rather than /n/ was chosen to precede *echo* in the sonorant condition because an initial /n/ would result in the embedded word *neck*. A necessary exception to the general avoidance of embedded words were English words of a single vowel phoneme, such as /i/ (the name of the letter *e*), /eɪ/ (the name of the letter *a*), /aɪ/ (*eye*), etc.

A full list of the resulting CV+C+target stimuli is listed in the Appendix. In addition to these, a list of 360 filler items was designed in the same way, except that fillers were made to contain no actual English words at all. In other respects they were the same as the targets (i.e., equally divided among the same sonority and stress pattern conditions).

2.1.1.2 Creation

Because the nature of our main question involved the effect of sonorant versus obstruent consonants, it was necessarily the case that the target words would follow different consonants. For this reason, it was not possible to perform the cross-splicing often used in word-spotting ex–

Table 2. Example of the four prosodic carriers used to produce the CVC (in bold) contexts for the test stimuli. There are two prosodic contexts (SW and WS stress pattern) and four CVC types, based on crossing vowel tenseness with consonant type (sonorant vs. obstruent). These initial CVCs were then excised and spliced onto target words with the same stress pattern.

	CV+C Contexts			
	jeem	*jim*	*foosh*	*fush*
To be appended to:				
SW targets	/ˌ**ʤim**ˈboʊboʊ/	/ˌ**ʤɪm**ˈboʊboʊ/	/ˌ**fuʃ**ˈtoʊtoʊ/	/ˌ**fʊʃ**ˈtoʊtoʊ/
WS targets	/ˌ**ʤim**bəˈboʊ/	/ˌ**ʤɪm**bəˈboʊ/	/ˌ**fuʃ**təˈtoʊ/	/ˌ**fʊʃ**təˈtoʊ/

periments to control for differences in the productions of targets across conditions. Therefore, we took special care to create stimuli that (a) would lack any potential speaker-encoded cues to intended syllabifications (prosodic or coarticulatory), and (b) would be unlikely to differ across conditions in terms of duration or intelligibility. After producing stimuli carefully, it would then be possible to "control" for any remaining differences, at least the durational ones, by adding them to our statistical model.

To this end, broad phonemic transcriptions of the stimulus materials were read and recorded by a 23 year-old female, a native speaker of American English from Southern California (where /ɔ/ and /ɑ/ have merged, and the remaining vowel, /ɑ/, patterns as tense). Recording took place in a sound-attenuating booth, digitized at 22.05 kHz. The CV syllable contexts (e.g., *jee*) were recorded separately from the target strings (e.g., *absent*). However, to create intervocalic consonants that would be phonetically ambiguous with respect to syllabic affiliation, those consonants were recorded as both codas to the initial CVs, and as onsets to the vowel initial targets. Thus, for example, CV+C strings like /ʤim/ and C+target strings /mæbsənt/ were produced separately so that, at a later date, they could be spliced together to create the stimulus item *jeemabsent*. The C+target strings were recorded by the speaker in isolation, but, in order to produce CV+Cs that would be prosodically appropriate for the targets onto which they would be spliced, all CV+Cs were produced in carrier strings, as shown in Table 2. Each carrier string had the same stress pattern as the relevant target word. Carrier strings were of the form __ˈX oʊXoʊ or __XəˈXoʊ , where X was a stop or fricative that matched the final C in the CV+C in place and voicing. Whether X was a stop or fricative depended on the identity of the final consonant in the CV+C; when the final C in the CV+C was a fricative, X was a stop; when the final C of the CV+C syllable was a stop, X was a fricative. This was done to prevent assimilation or excessive reduction of stops. Note also that all of the final Cs in the CV+Cs should have been treated as codas by the speaker, given their pre-consonantal position in the carrier string. Thus, any coarticulatory information in the vowel should have been that of a tautosyllabic coda consonant.

Yet another necessary control also involved prosody, namely that of the targets words. Because we wished to compare reaction times across sonority conditions, it was necessary to control for the overall prosody of targets across those conditions, especially durational differences. This was done as follows. The speaker produced the two versions of C+target strings together in a paired sequence, as similarly as possible. For example, *mabsent* and *babsent* were always produced one after another. Five to fifteen repetitions of each such pair of C+target strings

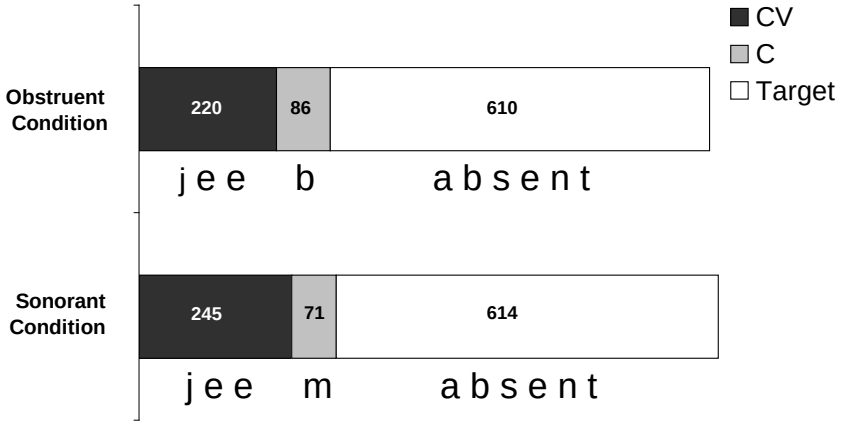

Figure 1. Mean durations (in milliseconds) for the initial CV contexts, intervocalic consonants, and target words in each of the two sonority conditions.

strings were read by the speaker. From this set of repetitions, the two most similar sonorant-initial and obstruent-initial productions were chosen by the experimenters, where "similar" meant that they matched as closely as possible in terms of (a) duration, (b) the shape of the f0 contour, and (c) the subjective judgment of the authors as to the relative prominence of the stressed and unstressed syllables. The CV+C syllables were also produced in this manner (but, again, separately, and in carrier strings).

Finally, the CV+C strings and the C+target strings were used to create the actual stimuli for the word-spotting task by splicing the relevant pairs together using a waveform editor. This was done by removing the final C in the CV+C context and splicing in the C+target string such that, for example, the /m/ recorded as an onset in *mabsent* was now the intervocalic consonant in *jeemabsent*. Extraction and splicing was done at zero-crossings. All CV+Cs and C+targets were combined in this way, resulting in 200 stimulus items: 50 target words, each appearing in 4 conditions. The 360 filler items were similarly recorded, although the speaker read the entire initial syllable and disyllabic non-word intact (e.g., /kuˈvimɛp/), as segmentations of fillers were not of interest.

In order to determine how successful the above methods were in producing stimuli that were similar in the crucial ways across conditions, we measured the durations of the initial CVs, intervocalic Cs, and targets for all stimulus items. Mean values for these durations are shown in Figure 1. As can be seen in the figure, a high level of durational similarity was achieved for target words across sonority conditions, a mean difference of only 4 ms. (There was, of course, no difference across the vowel tenseness conditions, since these were the exact same productions). Differences between conditions for the CV and the intervocalic consonants themselves were somewhat larger: initial CVs were 25 ms shorter, and intervocalic Cs 15ms longer, in the obstruent condition compared with the sonorant condition.

Stimuli were divided into blocks such that a target word occurred once in each block, in a different condition; blocks were counterbalanced with respect to tenseness and sonority conditions. Fillers were distributed among the blocks.

2.1.2 Participants

Participants were 33 native speakers of American English, mostly undergraduate students from California; each received either course credit or monetary compensation.

2.1.3 Procedure

Participants served as listeners in a word-spotting experiment. A within-subject design was used, with all participants hearing all targets in all conditions. Participants did this over two experimental sessions, separated by approximately one week. In each session, the participant heard each target word twice, once in each sonority condition and once in each vowel tenseness condition, separated by blocks. Orderings of the blocks were counterbalanced across participants, and all items (stimuli together with fillers) were randomized within each block for each participant. Stimuli were presented with an inter-stimulus interval of 3 seconds by a MATLAB script, which also recorded the reaction times (RTs) to a button push. In addition to RTs, participants' verbal responses (to indicate the word they had just heard) were also recorded for each item and saved as wav files for later assessment of accuracy.

Testing took place in a sound-attenuating booth; each of the two experimental sessions lasted approximately 20 minutes. Participants were told that they would hear a list of nonsense strings of speech produced by an English speaker, and that, in some of them, a real English word would be embedded. Their task, they were told, was to indicate when they heard a real English word by clicking a computer key as soon as they knew what the word was; immediately after pushing the key, they were to then say that word aloud. If they heard no word in a string, they were to give no response, and to simply wait for the next trial. A brief practice session was used to familiarize participants.

2.2 Results

Data were dropped from two participants who did not follow instructions (one did not consistently give a verbal response, one participant pressed the response key for every item presented, including non-word fillers), and two additional participants were dropped because they did not return for their second session. This left 29 participants whose accuracy and RT were analyzed. A response was considered an error if (a) the key was not pressed, (b) no verbal response was given, (c) a verbal response other than the intended target was given, or (d) the key press was made more than 2 seconds after target offset. From the pool of accurate responses, the RT analysis reported below was limited to those within 2 standard deviations of the mean RT.

The mean RT and accuracy data for the four conditions are shown in Table 3, separately for the SW and WS targets. In those averages, the following observations can be made. First, preceding vowel tenseness is a consistent predictor of neither RTs nor accuracy, consistent with previous research (Norris et al. 2001, Newman et al. (2011). Second, there is a quite consistent relationship between sonority and word-spotting performance: words like *absent* are more quickly and more accurately identified in *jeemabsent* than in *jeebabsent*. Another difference, one which was not a specific focus, is between stress pattern groups: target words with a SW pattern were generally recognized more slowly and somewhat less accurately than those with a WS stress pattern, regardless of the experimental manipulations. This pattern is highly unexpected given previous work supporting the MSS, but has a reasonable explanation (discussed below).

We attempted to better understand the above RT and accuracy patterns by modeling them as a function of our manipulations, and properties of the stimuli themselves using mixed-effects regression. In the (linear) regression model of RTs, we modeled the outcome *RT* using *participant* and *item* as random effects. The following fixed-effects predictors were also included in the model. Linguistic predictors were: preceding *vowel tenseness* (Tense/Lax), *sonority* of the intervocalic consonant (sonorant/obstruent), *stress* (SW/WS); stimulus-level predictors were: *target duration*, *duration of the intervocalic consonant*, *duration of the preceding CV*, and the

Table 3. Average RT (in ms, measured from target offset) and accuracy for strong-weak and weak-strong target words in the four experimental conditions.

	Strong-Weak Targets	
	Tense Vowel Context	*Lax Vowel Context*
Obstruent C Context		
RT	403	414
Accuracy	87.6%	89.6%
Example	/ʤib/absent	/nɪb/absent
Sonorant C Context		
RT	383	392
Accuracy	93.3%	91.9%
Example	/ʤim/absent	/nɪm/absent

	Weak-Strong Targets	
	Tense Vowel Context	*Lax Vowel Context*
Obstruent C Context		
RT	363	347
Accuracy	90.3%	92.0%
Example	/ʃɑv/eject	/ʃɪv/eject
Sonorant C Context		
RT	345	350
Accuracy	94.3%	94.7%
Example	/zɑm/eject	/zɛm/eject

presentation of the stimulus (1st, 2nd, 3rd or 4th time the participant had been presented with that particular target word during the course of the experiment). Among these fixed-effect parameters, the three linguistic variables were all permitted to interact with one another, and with *presentation*. Model comparison (using log-likelihood ratio tests) was used to remove non-contributing parameters. In order to test whether the predictors of participants' RTs were also predictors of participants' accuracy, the same predictors were included in a second round of modeling, (this time using logistic regression) of the binary outcome variable *accuracy* (correct/incorrect response). We present the resulting best-fit models of RTs and accuracy separately below.

Table 4. Estimate, standard error, t- and p-values from the model of reaction times.

	β	*SE*	*t*	*p*
(Intercept)	1012.15	30.50	33.18	< .0001
Presentation	-46.85	1.89	-24.84	< .0001
Target Duration	-765.52	33.74	-22.69	< .0001
Sonority (+son)	-18.81	7.25	-2.59	< .01
Stress (SW)	-52.27	8.59	-6.09	< .0001

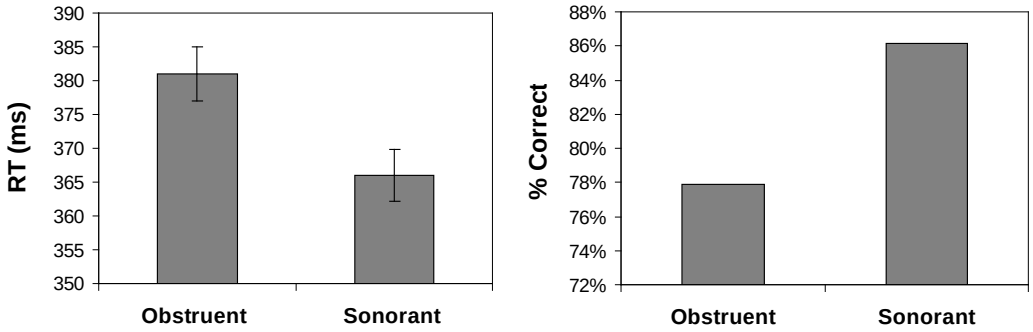

Figure 2. Main effects for sonority on reaction times (*left*, mean RT and standard errors) and accuracy (*right*, overall percent correct).

2.2.1 Reaction time

The results of the best fit model of the RT data contained the following fixed effects: *presentation*, *target duration*, *sonority*, and *stress*; we group the description of the model's output, shown in Table 4, by predictor type.

2.2.1.1 Stimulus-based predictors

The model showed a significant effect for *presentation*, having an inverse relationship with RTs ($p<.0001$), indicating that, unsurprisingly, listeners were faster to spot a target word with each successive time they encountered it in the experiment. The effect of *target duration* on RTs was also significant ($p<.0001$), such that longer target words were associated with faster spotting.

2.2.1.2 Linguistic predictors

The factors that were of primary interest to us were linguistic factors relating to the sonority of the intervocalic consonant and the tenseness of the vowel preceding that consonant. The best fitting model was one which did not include preceding *vowel tenseness*, indicating that this factor had no influence on segmentations. There was, however, a significant effect for *sonority*, in the direction of targets being spotted more quickly when they followed sonorants ($p <.01$). On average (collapsing across the two vocalic conditions) target words were identified 15 ms faster when they followed a sonorant rather than obstruent consonant (Figure 2). Although there was no interaction with *presentation* (indicating that sonority's effect was robust even within a compressed RT range), the advantage was numerically largest for first (22 ms) and second (24 ms) encounters of targets.

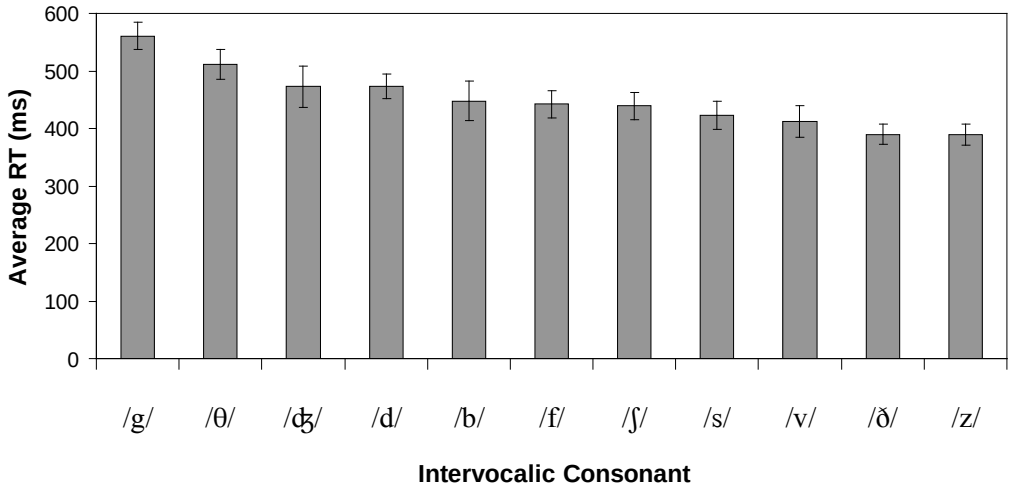

Figure 3. Mean reaction times to targets, grouped by intervocalic consonant. Error bars show standard error.

We also explored the possibility of a more gradient influence of sonority. That is, we asked whether the sonority scale was a predictor of segmentations *within* the obstruent category. If it were, spotting *absent* should be easier in a string like *jeevabsent* than in *jeebabsent*, since /v/ is a high-sonority obstruent relative to /b/ (given any definition of sonority that we know of). Our stimuli were not designed to address this question, as different targets (with different lexical frequencies and very different durations) were paired with different obstruent consonants (i.e., we would be comparing *jee[b]absent* versus *vu[ʃ]echo* vs. *chi[v]onion*, and so on). However, because we had a number of words per obstruent, and a range of obstruents, we explored whether there was any advantage for targets following more sonorant obstruents. We therefore binned target words by obstruent and calculated mean RTs for each group of targets, limited to first presentation responses only.

These means are shown in Figure 3, ordered according to average RT for that group of targets. To test whether there were any significant differences among these groups, a series of linear models of the same sort as above were used. In each model, the outcome variable *RT* was modeled as a function of the random effects *participant* and *item*, and the fixed effects *intervocalic consonant* and *target duration*. Each model carried out pairwise comparisons for one consonant group against all others (and significance levels were adjusted for the number of comparisons made). This set of models indicated that, when target duration was taken into account, the only noteworthy difference that occurred was a marginally significant one between two items on opposite ends of the sonority scale: /v/ versus /g/. (We suppress the full table of results, and instead report statistics for just /g/ relative to /v/: β = 144.9, p=.0996). Thus, while a common version of the sonority scale might have been used to predict a ranking for ease of segmentation as in (3), statistical results supported only that in (4), and indeed just marginally.

(3) /d/, /b/, /g/ < /ʤ/ < /θ/, /f/, /ʃ/, /s/ < /z/, /ð/, /v/

(4) /g/ < /v/

Table 4. Estimate, standard error, z- and p-values from the model of accuracy.

	β	*SE*	*z*	*p*
(Intercept)	-1.96	0.70	-2.79	**< 0.01**
Presentation	0.83	0.06	14.71	**< .0001**
Target duration	4.89	1.10	4.45	**< .0001**
Sonority (+son)	0.61	0.11	5.55	**< .0001**

However, given the lack of control over the target words in each group, we might have expected this to come out much more randomly than it did. Figure 3 shows that, at least numerically, the sonority scale was a remarkably good predictor of RTs. Further study would be needed to confirm this trend.

The final linguistic variable in the model was not one relating to a hypothesis of interest here, but is nonetheless important to note. The highly significant effect for stress in the model (p <.0001), indicated that target words that had a WS stress pattern (e.g., *eject*, *occur*) were associated with slower RTs than targets with a SW pattern (*echo*, *absent*). This is consistent with previous research, and offers further support for the MSS. Note that this finding seems to be at odds with the overall averages in Table 3. This is likely because, overall, WS targets were longer than SW target words (and longer target duration was associated with shorter RT); our model takes this into account and factors out individual target durations, revealing the expected asymmetry.

2.2.2 Accuracy

The best fit model of listeners' word-spotting accuracy contained the same predictors as the model of RTs, except for *stress*, which did not predict accuracy. As the brief description of results indicates, the accuracy results generally mirror the RT results.

2.2.2.1 Stimulus Variables

The model indicated a significant effect for *presentation* (p < .0001), subsequent presentations of a target word were associated with higher accuracy. The effect for *target duration* was also significant, such that longer durations were associated with higher accuracy (p <.0001).

2.2.2.2 Linguistic Variables

As in the RT data, sonority was found to contribute to improved word-spotting (p>.0001), as can be seen in Figure 2, above. Notably, both *stress* and preceding *vowel tenseness* were absent from the model, indicating that they were not relevant to predicting accuracy. Due to lack of space, we do not pursue the possibility of any gradient sonority effects on accuracy.

3 Discussion and Conclusion

In the experiment presented above, we tested whether English-speaking listeners showed a bias that could be understood in terms of syllable structure. The results showed very clearly that such a bias exists: when positing word boundaries, listeners do not treat intervocalic consonants all equivalently. Rather, they prefer to parse sonorants with the preceding vowel (jeem.absent), and

obstruents with the following vowel (jee.babsent), making vowel-initial words more difficult to spot in the latter case. Thus, on-line segmentations by English-speakers show the same preference that off-line syllabifications show, which also reflect cross-linguistic preferences for syllable structure (e.g., Clements 1990). This is inconsistent with the claim that native language rhythm class dictates the utilization of such a strategy, since listeners of English—the prototypically stress-timed language—seem to be guided by syllable structure as well. Indeed, the results may be interpreted as adding to the uncertainly regarding the validity of the rhythm class distinction (see Arvaniti 2012 for a recent discussion). Finally, the present study also revealed a trend that suggested sonority's effect was gradient, as the sonority scale was a predictor of RTs within the obstruent category. This finding, though interesting, requires further testing (with more statistical power) to be verified.

Worth noting is that it may be the case that these sonority-based preferences can be explained by the same mechanisms as other findings from word-spotting studies. As discussed in Section 1.1, the various segmentation strategies that listeners have been found to employ seem to be reliably related in some way to lexical statistics. We did not consult any corpora to determine what the facts of the English lexicon are regarding the sonority of consonants in simple onsets and simple codas, the forms relevant to our findings. If it turned out that the sonority scale actually fell out completely from such statistics, then of course the sonority-based preference can actually be characterized as another lexical one (although the cross-linguistic tendency would still not be explained). That being said, it has been shown that gradient preferences, indeed ones having to do with sonority, can emerge indirectly from lexical statistics, given that certain other kinds of information are available to allow for generalization (Daland et al. 2011).

To our knowledge, such gradient phonotactic knowledge has not been shown to influence word-spotting, a fact that seems to distinguish on-line segmentation behavior from off-line acceptability judging behavior (see Albright 2009 for a recent review). One complication is that another, quite categorical, phonotactic violation did *not* affect listeners in our study. That is, we did not find any tendency for listeners to avoid parses that would leave as residue a preceding CV sequence with a lax vowel. Norris et al. (2001) argue that this result, which they also found, indicates that the Possible Word Constraint is actually more like a "Possible Syllable Constraint" (see also Cutler et al. 2001). This construal is somewhat awkward, however, since in languages like Slovak, there is evidence that the constraint is in fact about possible wordforms, not syllables (Hanulikov et al. 2010). Newman et al. (2011) suggest the apparent lack of penalty for such residues with open lax vowels may indicate that on-line segmentation does not make use of lexical statistics afterall, and that such information may only become available post-perceptually. They also note the possibility, however, that prohibitions on ill-formed residues may be less crucial than those on ill-formed target onsets, echoing earlier claims about the primacy of onsets over codas in segmentation (van der Lugt 2001, Dumay, Frauenfelder, & Content 2002).

We think it is a reasonable hypothesis that the kinds of information that listeners are able to tap into on-line versus off-line differs, and that overlap will occur for information that is maximally internalized. We might assume that grammatical preferences relating to sonority — which may emerge from lexical statistics— will be utilized both on- and off-line. Similarly, knowledge of what is completely unattested may be used by both processes. However, even when segmentation routines apply constraints based on such knowledge, they do not assign violations equally at all points in the string; rather, they are applied primarily to possible word onsets. Thus, in word-spotting experiments where the residue precedes rather than follows the target word, only the worst violations incurred by the residue will be detectable, if at all.

Possibly, this might reflect an overall strategy that not all listeners employ equally well on-line. For example, Yu et al. (2011) found that working memory capacity distinguished listeners' sensitivity to phonotactic context, even in an off-line task. In that study, the authors found that listeners with lower working memory capacity (and/or fewer "autistic" traits) were actually more sensitive to phonotactics. Possibly, then, strategies for investigating the on-line use of more gradient well-formedness may include an individual differences approach – a question for future research.

Appendix: Stimuli

Obstruent		**(SW Targets)**	**Sonorant**	
Tense	**Lax**		**Tense**	**Lax**
/dɑv/over	/mɪv/over		/dɑm/over	/mɪm/over
/doɪg/image	/dɛg/image		/doɪm/image	/dɛm/image
/foɪz/under	/fɛz/under		/foɪm/under	/fɪm/under
/fuð/album	/vʊð/album		/fun/album	/vʊn/album
/fuʃ/echo	/fʊʃ/echo		/fum/echo	/fʊm/echo
/keɪθ/angle	/kɛθ/angle		/keɪl/angle	/kɛl/angle
/kus/empty	/kʊs/empty		/kum/empty	/kʊm/empty
/loɪb/expect	/lʌb/expect		/loɪm/expect	/lʌm/expect
/muʃ/average	/mɪʃ/average		/mul/average	/gɛl/average
/neɪs/effort	/nɛs/effort		/neɪn/effort	/nɛn/effort
/θeɪf/enter	/vɛf/enter		/zin/enter	/næn/enter
/θeɪf/extra	/θɛf/extra		/θeɪm/extra	/θɛm/extra
/poɪd/uncle	/dʊd/uncle		/poɪn/uncle	/dʊn/uncle
/ʧɑθ/ancient	/ʧʌθ/ancient		/teɪn/ancient	/tʊn/ancient
/ʧiv/onion	/ʃɪv/onion		/plul/onion	/plɛl/onion
/vif/anxious	/vʌf/anxious		/vin/anxious	/vɛn/anxious
/vub/other	/vʊb/other		/vul/other	/vʌl/other
/vuʤ/ever	/vɛʤ/ever		/vum/ever	/vɛm/ever
/vuf/expert	/væf/expert		/vɑm/expert	/væm/expert
/vuθ/either	/gʊθ/either		/vul/either	/gʊl/either
/vuʃ/evict	/vʊʃ/evict		/vum/evict	/vʊm/evict
/vus/ugly	/vʊs/ugly		/vun/ugly	/vʊn/ugly
/zeɪf/oven	/zɛf/oven		/zeɪn/oven	/sɛn/oven
/zeɪv/action	/zæv/action		/zɑm/action	/zæm/action
/zif/elect	/zɪf/elect		/zin/elect	/zɪn/elect
/zoɪg/ankle	/zʌg/ankle		/zoɪn/ankle	/zʌn/ankle
/ʤib/absent	/nɪb/absent		/ʤim/absent	/nɪm/absent

Obstruent		**(WS Targets)**	**Sonorant**	
Tense	**Lax**		**Tense**	**Lax**
/bus/against	/bʊs/against		/fun/against	/fʊn/against
/ʤaɪf/accept	/ʤæf/accept		/ʤaɪm/accept	/gæm/accept
/doɪg/allow	/dʊg/allow		/doɪm/allow	/dɛm/allow

/faɪʤ/attach	bʊʤ/attach	/faɪm/attach	/fæm/attach
/fug/astound	/fɛg/astound	/fum/astound	/fɛm/astound
/fuʃ/occur	/fʊʃ/occur	/fun/occur	/fʊn/occur
/gib/immerse	/gɪb/immerse	/θul/immerse	/dæl/immerse
/giʤ/annoy	/gʌʤ/annoy	/zin/annoy	/zʊn/annoy
/gig/enough	/nʌg/enough	/blim/enough	/nɪm/enough
/giz/erect	/gɛz/erect	/gin/erect	/gɪn/erect
/keɪθ/among	/kɛθ/among	/bul/among	/ʧʊl/among
/θeɪʃ/adopt	/θɛʃ/adopt	/θeɪn/adopt	/θɛn/adopt
/θeɪʃ/exist	/θʌʃ/exist	/θeɪn/exist	/θʌn/exist
/ʃɑv/eject	/ʃɪv/eject	/zɑm/eject	/zɛm/eject
/ʧeɪð/exempt	/ʧɪð/exempt	/ʧeɪm/exempt	/kɛm/exempt
/teɪθ/applaud	/tæθ/applaud	/teɪn/applaud	/gɛn/applaud
/vif/agree	/vɪf/agree	/vin/agree	/vɪn/agree
/voʊd/avenge	/vɪd/avenge	/voʊn/avenge	/vɪn/avenge
/vuʤ/effect	/vʊʤ/effect	/vul/effect	/vʊl/effect
/vuθ/achieve	/væθ/achieve	/vul/achieve	/fæl/achieve
/vuʃ/object	/vʊʃ/object	/vul/object	/vʊl/object
/zeɪf/obsessed	/zɛf/obsessed	/zeɪm/obsessed	/zɛm/obsessed
/zoɪg/exert	/zʊg/exert	/zoɪn/exert	/zʊn/exert

References

Albright, A. 2009. Feature-based generalisation as a source of gradient acceptability. *Phonology* 26. 9-41.

Arvaniti, A. 2012. The usefulness of metrics in the quantification of speech rhythm. *Journal of Phonetics* 40, 351-373.

Clements, G. 1990. The role of the sonority cycle in core syllabification. In J. Kingston & M. Beckman (Eds.), *Papers in laboratory phonology 1: Between the grammar and physics of speech*, 283–333. New York: Cambridge University Press.

Cutler, A., J. Mehler, D. Norris, & J. Segui. 1986. The syllable's differing role in the segmentation of French and English. *Journal of Memory and Language* 85, 385-400.

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

Cutler, A., J. McQueen, D. Norris, & A. Somejuan. 2001. The roll of the silly ball. In E. Dupoux (Ed.), *Language, brain and cognitive development: Essays in honor of Jacques Mehler*, 181-194. Cambridge, MA: MIT Press.

Daland, R., B. Hayes, J. White, M. Garellek, A. Davis, & I. Normann. 2011. Explaining sonority projection effects. *Phonology* 28, 197-234.

Dumay, N., U. Frauenfelder, & A. Content. 2002. The role of the syllable in lexical segmentation in French: Word-spotting data. *Brain and Language* 81, 144-161.

Eddington, D., R. Treiman, D. Elzinga. to appear. Syllabification of American English: Evidence from a large-scale experiment (Part I). *Journal of Quantitative Linguistics*.

Hanulikova, A., J. McQueen, & H. Mitterer. 2010. Possible words and fixed stress in the segmentation of Slovak speech. *Quarterly Journal of Experimental Psychology* 63, 555 -579.

Kirk, C. 2001. *Phonological constraints on the segmentation of continuous speech*. Ph.D. Thesis, University of Massachusetts Amherst.

McQueen, J. 1996. Word Spotting. *Language and Cognitive Processes* 11, 695-699.

McQueen, J. 1998. Segmentation of continuous speech using phonotactics. *Journal of Memory and Language* 39, 21-46.

Norris, D., J. McQueen, A. Cutler, & S. Butterfield. 1997. The Possible Word Constraint in the segmentation of continuous speech. *Cognitive Psychology* 34, 191-243.

Norris, D., J. McQueen, A. Cutler, S. Butterfield, & R. Kearns. 2001. Language-universal constraints on speech segmentation. *Language and Cognitive Processes* 16, 637-660.

Treiman, R. & C. Danis. 1988. Syllabification of intervocalic consonants. *Journal of Memory and Language* 27, 87-104.

van der Lugt, A. 2001. The use of sequential probabilities in the segmentation of speech. *Perception & Psychophysics* 63, 811–823.

Yu, A., J. Grove, M. Martinović, M. Sonderegger. 2011. Effects of working memory capacity and "autistic" traits on phonotactic effects in speech perception. *Proceedings of the XVIIth International Congress of Phonetic Sciences*, 2236-2239.