ALLOPHONIC INFLUENCES ON WORD-SPOTTING EXPERIMENTS

Rachel Smith and Sarah Hawkins
Department of Linguistics, University of Cambridge, U.K.

ABSTRACT
The influence of allophonic detail on word segmentation was examined in two word-spotting experiments. In one, stimuli had a critical allophone which was appropriate or inappropriate for the target word. Responses were faster and more accurate when the allophone was appropriate, but the design confounded allophone and phonological structure. Expt. 2 separated stress and ambisyllabic to remove the confound. Listeners heard items like puzzaath which had been spoken as two words (+Boundary) or one (-Boundary), and with stress on the first or second syllable. Responses were faster and more accurate to +Boundary items, and the difference between ±Boundary items was larger if stress was on syllable 2. Implications are discussed for the relationship between allophones, prosody and linguistic structure in word recognition.

1. INTRODUCTION
Experiments show that word segmentation processes depend heavily on the phonological attributes of the language being listened to. Norris et al. ([1]) propose a Possible Word Constraint (PWC) whereby listeners reject word candidates which leave a non-syllabic residue between the candidate word and a probable word boundary: probable boundaries are those which are cued by metrical, phonotactic or allophonic information, or by silence. Word-spotting experiments typically focus on how listeners use metrical and phonotactic boundary information in segmentation ([2], [3]); allophonic cues have received comparatively little attention.

Allophonic variation within words is important in lexical access, as shown in other paradigms: gating ([4]), lexical decision ([5]), and word identification ([6]). It follows that allophonic variation at word boundaries should be important for segmentation, and may relate in interesting ways to the other sources of boundary information. Some word-spotting studies ([1], [2]) suggest that metrical- and phonotactically-based segmentation are independent of the kind of phonetic detail which can be broadly termed allophonic. But others ([7], [8]) do suggest that segmentation may be guided by phonetic cues to the relevant phonological structures. If this is so, then allophonic information may underly the operation of metrical- and phonotactically-based segmentation processes, rather than merely describing a separate source of boundary information.

We describe two word-spotting experiments that show that allophonic detail can affect segmentation. Expt. 1 broadly confirmed the hypothesis but confounded strong syllable onsets and allophone. Expt. 2 was a more stringent test, distinguishing the effects of metrical strength and allophonic variation due to ambisyllability.

2. EXPERIMENT 1
2.1. Method
2.1.1. Participants
The 30 students from the University of Cambridge, 29 aged 18-30, one aged 47, were all monolingual speakers of British English with no speech or hearing disorder.

2.1.2. Materials
72 monosyllabic target words were each embedded in a pair of nonsense words. Nonsense words were minimal pairs differing only in a single critical allophone (CA). The CA was appropriate (A) for the target word in isolation in one member of the pair, and inappropriate (I) in the other. Onset stimuli (26 bisyllabic pairs) had targets in syllable 2, and CAs at syllable onset, e.g. reap in /gəˈrip/ (A) and /gəˈdɪp/ (I). Coda stimuli (16 trisyllabic pairs) had targets in syllable 1, and CAs at syllable coda, e.g. bell in /bɛlˈpɪsɪm/ (A) and /bɛlˈastɪm/ (I). Morpheme stimuli comprised a target word and a pseudo-morphemic affix: 20 Morpheme Coda pairs (ill in /ˈɪlləs/ (A) and /ˈɪləs/ (I)) and 10 Morpheme Onset pairs (pip in /ˈdɪpsˈpɪp/ (A) and /ˈdɪsˈpɪp/ (I)). Most stimuli had only one embedded word (the target). This criterion was relaxed in 13 items, e.g. where an affix was also a word, like mis- (miss). 72 fillers matched experimental items in phonological structure, but contained no real word.

Stimuli were read by a female phonetician (the second author) from a phonemic transcription which had the CA identified in narrower transcription. They were recorded on high-quality equipment in an IAC booth, and digitized at 16 kHz to a Silicon Graphics machine running xwaves.

2.1.3. Procedure
Participants heard (over headphones) 20 practice items, and then one of two lists, each containing one member (A or I) of each pair, with equal numbers of items from each category. The ISI was 3.5 s. Participants were tested individually in an IAC booth. They were told that they would hear a list of nonsense words and should listen for real English words embedded in the nonsense. They pressed a button each time they detected a real word, and said the word aloud. A PC controlled presentation via a Tucker-Davis DD1 system which recorded button presses, reaction times (RTs) and verbal responses.

2.2. Results and Discussion
2.2.1. Results
RTs were measured from target offset. Missing, or wrong verbal, responses were treated as errors. Errors and RTs
were each submitted to two ANOVAs, one with subjects (F1) and the other with targets (F2) as the repeated factor. Category (Onset, Coda, Morpheme) and Allophone (A, I) were the other factors. Only F1 is reported if effects were significant in both subject and target ANOVAs.

Table 1 shows an interaction between Allophone and Category (Errors, F1(2,58) = 17.5, p<0.001; RTs, F1(2,58) = 6.2, p<0.005). Responses were slower and less accurate to I allophones in Onset and Morpheme stimuli, but Coda stimuli showed no difference between A and I. Coda targets were responded to more slowly and less accurately overall than the other two categories.

2.2.2. Discussion

Allophonic appropriateness thus can affect segmentation in word-spotting. If the A allophone is in the target syllable onset, its influence on target detection is very large, while in the target syllable coda, its effect is smaller or negligible. An I allophone at word onset may be more disruptive than one at the end of a word after evidence for the correct word has accrued, as continuous-evaluation theories predict. Also, Coda targets were the first syllable of a trisyllabic foot: this probably shortened them and altered the timing relationships between their segments, so that both A and I Coda stimuli might have contained inappropriate versions of their target words, relative to Onset stimuli where targets formed the sole syllable of a foot, as they would if uttered in isolation.

The largest effects confounded allophones with phonological structure: targets in I Onset stimuli like /pa'dripl/ had to be extracted from syllables with complex onsets; stress was not varied, as target syllables were always strong and nonsense contexts weak. Non-parametric tests on subsets of the data suggest allophones were important despite the confounds. Words were harder to spot when the CA in onset clusters was devoiced, and allophones affected Morpheme Coda pairs which did not confound CA with clusters.

The stimuli also represented an unsystematic approach to the issue of what makes an allophone appropriate. For example, we assumed that Coda /s/ would be better in /'bɛlpəsɪm/ than /'bɛləsɪm/ because intervocalic /s/ has features of the clear allophone. Yet in connected speech, bell can be followed by a vowel, so intervocalic /s/ cannot be automatically wrong as a syllable coda.

Expt. 2 attempted to address the weaknesses of Expt. 1, using the ProSynth (PS) model ([9]) to systematise the relation between metrical structure and phonetic detail.

### Table 1: Mean percentage error rates and RT (ms) to stimuli in Expt. 1, by Category and Allophone.

<table>
<thead>
<tr>
<th>Category</th>
<th>Percentage</th>
<th>Error</th>
<th>RT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Onset</td>
<td>15%</td>
<td>667</td>
<td>58%</td>
</tr>
<tr>
<td>Coda</td>
<td>38%</td>
<td>1050</td>
<td>47%</td>
</tr>
<tr>
<td>Morpheme</td>
<td>20%</td>
<td>768</td>
<td>42%</td>
</tr>
<tr>
<td>Total</td>
<td>24%</td>
<td>828</td>
<td>49%</td>
</tr>
</tbody>
</table>

3. EXPERIMENT 2

The PS model provides systematic principles to explain allophonic differences in identical phoneme sequences. For example, the differences in the realisation of the first four phonemes of mistakes and mistimes are predictable from the first syllables’ phonological weight and metrical strength. The weight difference depends on whether or not the cluster /st/ is ambisyllabic, which is determined in turn by the productivity of the morpheme mis (productive for mistimes but not mistakes). Thus weight and strength predict the allophonic differences, but it is the allophonic differences that signal the strength and weight values.

The PS model uses maximal ambisyllabicity within words and minimal ambisyllabicity between words. This basic principle is sensitive to the constituency of the Coda and Onset of adjacent syllables and to the stress pattern. Expt. 2 studied just one type of word join, where the Coda of word 1 is a single consonant and the Onset of word 2 is empty (e.g. rash oath). For such cases, the PS model states that if σ1 is stressed (RASH oath) its Coda is ambisyllabic. If σ2 is stressed (rash OATH), the Coda of σ1 is not ambisyllabic. In a similar nonsense word like puZoath, with no internal word boundary, the /Z/ is ambisyllabic whether σ1 or σ2 is stressed.

Table 2: Stress and Boundary conditions for Expt. 2.

<table>
<thead>
<tr>
<th>+ Boundary</th>
<th>- Boundary</th>
</tr>
</thead>
<tbody>
<tr>
<td>σ1 Stress</td>
<td>PUZ # oath (A)</td>
</tr>
<tr>
<td>σ2 Stress</td>
<td>puZ # OATH (C)</td>
</tr>
</tbody>
</table>

Applied to word-spotting, this leads to the following predictions: (1) all allophonic cues will distinguish between one- and two-word utterances with stress on σ2 (puzOATH vs. puZ # OATH) but no such cues will distinguish one- and two-word utterances with stress on σ1 (PUZoath vs. PUZ # oath). Thus, listeners should find it harder to spot oath in condition D of Table 2 than in C, but oath should not be harder to spot in B than in A. (2) the extent to which allophonic detail cues word boundaries will vary with the type of consonant preceding the target word. Some consonants, like voiceless stops, have clearly distinct Onset and Coda allophones, others less so; the former may produce larger differences between conditions C and D. (3) overall, stressed targets should be easier to spot than unstressed ones: they are more intelligible, at least in isolation [10].

3.1. Method

3.1.1. Participants

The 32 students from the University of Cambridge, aged 19-27, were all monolingual speakers of British English with no speech or hearing disorder.

3.1.2. Materials

64 target words (32 monosyllables, 32 bisyllables) were each embedded in a nonsense word. All target words began with a vowel. Nonsense contexts always preceded
the target word, and were C1VC2 syllables whose vowel was one of /æ eɪ ou i ʌ / . These vowels were chosen since in English they may appear as the first syllable of a Weak-Strong word while retaining the quality they have in stressed syllables. C2 fell into one of two groups, those whose onset and coda allomorphs are very distinct (/lptk/) and those where they are less so (fricatives, /bdg/). Stress and word boundary location were fully crossed as in Table 2. 64 fillers matched the experimental items in phonological structure but contained no real word.

Recording followed the procedure of Expt. 1, except that items appeared in a sentence (e.g. He said it was the PUZoath after all) to help the speaker produce the stress and boundary placement naturally, without stating the particular allomorphs via transcription. The sentence context was read silently, then the critical item spoken.

3.1.3. Procedure

Participants heard 20 practice items, then one of 4 lists, each containing one member of each Stress/Boundary quadruplet, and equal numbers of monosyllabic and bisyllabic targets from each group of preceding consonant types. The procedure was otherwise identical to Expt. 1.

3.2. Results and Discussion

The data were pre-processed as in Expt. 1. RTs were log-transformed because of high variability. Errors and log-transformed RTs were each submitted to two ANOVAs (Stress x Boundary x Preceding Consonant x Number of Syllables in Target), with subjects (F1) and targets (F2) as the repeated factor. For effects that were significant in both subject and target ANOVAs, only F1 is reported. T-tests are one-tailed.

Table 3: Mean percentage error rates and RT (ms) for Stress and Boundary conditions of Expt. 2.

<table>
<thead>
<tr>
<th></th>
<th>+Boundary</th>
<th>-Boundary</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Error RT</td>
<td>Error RT</td>
<td>Error RT</td>
</tr>
<tr>
<td>σ1 Stress</td>
<td>25% 607 (A)</td>
<td>25% 644 (B)</td>
<td>25% 625</td>
</tr>
<tr>
<td>σ2 Stress</td>
<td>22% 575 (C)</td>
<td>32% 637 (D)</td>
<td>27% 606</td>
</tr>
<tr>
<td>Total</td>
<td>23% 591</td>
<td>29% 641</td>
<td>26% 615</td>
</tr>
</tbody>
</table>

As the Totals of Table 3 show, the presence of a boundary produced more accurate and faster responses (Boundary: Errors, F1 (1,31) = 9.9, p<0.005; RTs, F1 (1,31) = 21.3, p=0.0001) whereas stress affected only speed of response (Stress: F1 (1,31) = 5.7, p<0.05), not accuracy. As predicted, the difference between +Boundary items was greater when σ2 was stressed: words were spotted more accurately in condition C than D, while there was no difference between A and B (Stress x Boundary: F1 = 10.0, p<0.005). RTs show a roughly similar pattern, although it just misses significance by Subjects (F1 (1,31) = 3.6, p=0.067; F2 (1,63) = 5.9, p<0.02): RTs are much faster in condition C than D, and somewhat faster in A than B (t(127) = 1.7; p<0.05). C and A did not differ (t(127) = 1.3; p=0.18), suggesting that weak stress alone does not slow target detection if a word boundary is allophonically marked.

Although stress seems to have a less powerful effect than allophonically-defined boundaries, its importance may lie more in interactions with other factors. As the first two rows of Table 4 show, words preceded by (/lptk/) were spotted more accurately if they had σ1 Stress than σ2 Stress; words preceded by (/Fric /bdg/) showed a smaller difference in the opposite direction (Stress x Consonant: Errors, F1 (1,31) = 5.25, p<0.05). When stress is on σ1, the distinctive coda allomorphs of (/lptk/) provide good evidence that the first syllable has ended, which should help listeners predict where a word might begin. Table 4 further shows that, as predicted, (/lptk/) produce larger differences between conditions C and D than (/Fric /bdg/): (RT: t(63) = 1.7, p<0.05). When stress is on σ2, an incorrect onset has a particularly disruptive effect if it is distinctive, as an aspirated stop or clear /l/ tends to be.

There were no other significant interactions. Consistent with the literature, monosyllabic targets were spotted less accurately and slower than bisyllabic ones.

4. GENERAL DISCUSSION

Our results show that, in at least some cases, there are reliable allophonic cues to word boundaries. This accords with e.g. [11] where listeners detected boundary placement at better than chance levels. But allophonic cues affect segmentation more in some contexts, like the stress-final conditions of Expt. 2, than in others, like the initial-stress conditions.

The results broadly support the insight of [1], [2], [3] that word boundaries are identified on the basis of phonological structures which do not map on to a language’s words in a one-to-one fashion. Our approach differs from that embodied in the PWC rather because we treat metrical, phonotactic and allophonic information sources as inextricably related by a prosodic hierarchy. The PWC treats them as essentially separate, if yoked together by the fact that they all feed the same elegant general constraint. Moreover, the PWC approach seems to imply a certain informational poverty of the acoustic signal, which needs supplementing with knowledge of
general distributional regularities. This contrasts with our assumption that the signal is informationally rich because linguistic structure is realised as systematic variation in its fine phonetic detail (cf. [9]). Of course, any of the cues proposed requires information to be identified from the acoustic signal, be it strong syllables, tautosyllabically illegal phoneme strings, or allophones. Equally, use of any of the cues needs knowledge of statistical regularity (eg of the distributions of allophones). The difference between the approaches is mainly in the extent to which subtle acoustic variation is seen as information or noise.

Treating subtle, systematic acoustic variation as informative has interesting consequences for theories of spoken word recognition. For instance, our results suggest that listeners are sensitive to morphologically conditioned allophonic variation. Such knowledge could assist lexical access by constraining sets of word candidates and speeding the recognition of the correct candidate (e.g. by lowering the activation of mistake and mysterious when a light syllable /mts/ is encountered, and conversely lowering the activation of prefixed words like mistime when /mts/ from a heavy syllable is heard). This might contribute to "affix-stripping" of the kind proposed by [12]. In other cases, morphologically-conditioned allophonic detail might allow grammatical and related semantic processing to be initiated prior to full recognition of a word (cf. [13]).

Allophonic segmentation cues may also be relevant to discourse processing. Consider the stress-initial case of Expt. 2, where allophonic boundary cues did not affect the accuracy of responses. These stimuli are comparable to Adjective + Noun phrases of English (e.g. rash oath). In real discourse, when the adjective in such phrases receives accent, the following noun is usually in some sense "old information" ([10]). Bard (this volume) argues that in such cases of relatively unintelligible old information, lexical access is more difficult, but listeners are simultaneously better able to integrate the word into a discourse context and form a representation of speaker intention. Following this line of reasoning, even the differences in the informational value of segmentation cues may, if systematic, be informative to listeners.

Further research could examine whether allophonic cues take priority over metrical or phonotactic cues. We suggest that allophonic detail may underly these other kinds of information: the VOT of a stop, for instance, helps define the degree of stress of the syllable it begins. Infants' earliest segmentation seems to be based around gross metrical cues ([14], [15]). Nevertheless, if our argument is valid, then infants' use of metrical cues may still reflect sensitivity to allophonic detail; they may simply not have systematised representations of all aspects of allophonic variation.

Our results are best accommodated by models of spoken word recognition that assume words to be recognised by matching a fine-grained auditory representation of the acoustic signal directly against the lexicon. Allophonic cues can be weak and subtle, if considered in isolation, but can be powerful if integrated over the syllable and longer domains.

5. ACKNOWLEDGEMENTS

Thanks to Noël Nguyen for carrying out the Expt. 1 ANOVAs. Correspondence: Rachel Smith, Department of Linguistics, Cambridge University, Sidgwick Avenue, Cambridge CB3 9DA, UK; rhs20@hermes.cam.ac.uk.

6. REFERENCES