4. Investigating the recognition of embedded words

The presence of words embedded at the onset of longer words has been suggested to challenge certain models of lexical segmentation and identification. As reviewed in the previous two chapters, it has been argued that a lexical segmentation mechanism that uses the pre-offset recognition of words in connected speech to identify word boundaries would be disrupted by these lexical items. If embedded words require post-offset information for longer competitors to be ruled out, pre-offset identification would not be possible and the account of lexical segmentation proposed in sequential models of spoken word recognition would not be able to operate effectively.

Some authors propose that the temporary ambiguity of these words necessitates models of word recognition that incorporate direct competition between lexical units (McQueen, Cutler, Briscoe & Norris, 1995). Lexical competition allows the identification of onset-embedded words since it enables mismatch which rules out longer competitors to boost the activation of the embedded word (McClelland & Elman, 1986; Norris, 1994).

However, in the previous two chapters, two different strands of evidence have been presented that question the validity of this inference from the presence of embedded words to models of spoken word recognition that employ lexical-level competition. Given the theoretical importance of onset-embedded words in distinguishing between alternative accounts of lexical segmentation and spoken word recognition, experimental evidence is required to support this argument.

Two assumptions are involved in this argument from embedded words to lexical competition, both of which make specific predictions regarding the time course of identification of words in connected speech. Consequently, experimental investigations can be used to evaluate whether the conclusion of McQueen et al. (1995) – that lexical competition is a necessary property of models of spoken word recognition – is valid.

The first assumption is that ambiguities created by onset-embedded words are resolved by incorporating direct competition between lexical items. In its strongest form this argument could be interpreted as suggesting that models without competition are incapable of identifying onset-embedded words. This strong form has been ruled out by simulations reported by Content & Sternon (1994) and by the recurrent network models

investigated in the previous chapter. As described in Chapter 3, networks in which the target of the recognition processes is a representation of an entire sequence provide a natural account of the identification of onset-embedded words without incorporating direct, inhibitory links between lexical units. Following contexts that rule out longer competitors are used to rule in embedded words, allowing their identification.

However, it may be more reasonable to reinterpret the argument presented by McQueen et al. as stating that the time course of identification of embedded words in connected speech more closely matches the predictions of lexical competition accounts than models that lack direct competition between lexical items. In previous chapters, lexical competition models were described that predict a short word bias during identification. That is, at the offset of a sequence of phonemes making up an embedded word, competition based models such as TRACE and Shortlist predict greater activation of units representing short words than long words.

Conversely, the recurrent network simulations described in Chapter 3 display probabilistic behaviour in the resolution of ambiguities during recognition. Thus, these networks predict that short embedded words and longer competitors will be equally active following a fragment of speech that matches both a short word and the onset of a longer word (all other things being equal). This difference between the two accounts can be tested in experiments on the time course of identification of words in connected speech. Specifically, lexical competition models predict increased activations for short words in a case where two otherwise equally plausible lexical items are active during identification, while the recurrent network account would predict that there will be no bias towards either short or long words during identification.

In discussing this apparent discrepancy between models with and without direct lexical competition, it is apparent that both models predict that onset-embedded words will be ambiguous with longer competitors during identification. It is only on the basis of complete ambiguity between embedded words and longer competitors that pre-offset identification of onset-embedded words would not be possible.

As described in the review of the acoustic-phonetics literature in Chapter 2, however, differences in segments at word onsets and duration differences between syllables in short and long words may provide acoustic cues to distinguish embedded words from longer competitors. Any pre-lexical acoustic cue that helps distinguish onset-embedded words

from longer lexical items would substantially reduce the ambiguity created by embedded words. These cues would weaken the claim that onset-embedded words produce ambiguities that can only be resolved through the use of lexical competition between word candidates.

From an experimental perspective, acoustic cues to distinguish short and long words would predict that an onset-embedded word (e.g. *cap*) should be activated more strongly by a fragment containing that word than by speech containing a longer lexical item (e.g. *captain*). Conversely, longer words (*captain*) should be activated more strongly by a matching fragment of speech than by a fragment containing an embedded word (*cap*).

Given the conflicting predictions regarding the time course of activation of embedded words in connected speech, the opening section of this chapter will review the relevant experimental literature. This review will focus on whether experiments make comparisons between embedded words and longer competitors that would be required to detect the presence of discriminatory acoustic cues and also whether biases towards short word hypotheses show up in the results of these experiments.

4.1. Review of previous experiments

In investigating the time course of activation of words in connected speech, researchers come up against methodological problems caused by the temporal nature of speech. For instance, in order to use reaction time as a dependent measure for a behavioural task, response times need to be measured from an appropriate position in the speech stream. In investigations of visual word recognition the time taken to process a stimulus can uncontroversially be measured from the onset of the visually presented word. However, in spoken word recognition the time that stimulus items take to be presented may differ between different words. Consequently, measurement of response time relative to the onset or the offset of a word may introduce an experimental confound through differences between stimuli in the time at which information needed to make a response becomes available in the speech stream.

This difficulty in controlling both the amount of sensory information available to participants and the amount of time provided for processing of the speech stimuli has led to research focusing separately on these two aspects of the recognition process. For instance, the gating technique provides information regarding the amount of sensory

information required for the identification of words; whereas tasks such as auditory lexical decision or word-spotting provide an indication of the amount of time required for processing stimuli that are assumed to be matched for the rate at which sensory information becomes available.

These issues will be prominent in reviewing previous experimental investigations on the time course of identification of onset-embedded words. Although providing valuable information on the nature of the recognition process for embedded words, there is little data comparing the lexical activation of embedded words and longer competitors at relevant positions in the speech stream. The differential predictions derived from the two classes of computational models and from acoustic-phonetic analysis of spoken words are specific to the activation of lexical hypotheses at particular points in the speech stream. The results of previous experiments may therefore only falsify the predictions of different accounts where they address the processing of stimuli at specific positions in the speech signal.

4.1.1. Gating

Gating is a frequently used task in experimental psycholinguistics in which speech is presented to subjects in fragments or gates of progressively increasing duration. Following each gate, subjects are generally asked to write down their best guess as to the identity of the word (or words) that they can hear, along with a rating reflecting their confidence in the response that they have given. By recording subjects' responses and confidence ratings at gates stepping through a word (usually starting from word onset), the gating task can be used to provide measures reflecting the activation and identification of competing lexical hypotheses as acoustic information accumulates. For an overview of research using the gating task see Grosjean (1996).

Dependent measures provided by gating can be expressed in terms of the amount of sensory information required for activation of a given lexical item. This is usually measured by the *isolation point* of a stimulus – the point at which subjects give a correct response and then do not change their mind at subsequent gates. An alternative statistic derived from gating measures the amount of input that is required for subjects to confidently recognise lexical items. This is measured by *total acceptance point,* or *recognition point,* usually defined as the point at which confidence ratings reach a

predetermined criterion. Both measures are argued to reflect the amount of sensory information required for lexical access.

Note however that on the criteria that were discussed earlier, the standard gating task (with multiple, successive presentations of individual items and un-timed written responses) will not provide any information about the processing time required for lexical access or identification. Indeed this failure to control the amount of processing that can be done on sections of speech may introduce responses biases or otherwise distort the results obtained in gating. For instance, experiments by Cotton and Grosjean (1984) suggest that, through participants perseverating with previous responses, the repeated presentation scheme may produce an overly conservative estimate of the amount of sensory input required for identification. On a more positive note, work by Tyler and Wessels (1985) suggests that spoken responses made under time pressure match reasonably well to those obtained when subjects write their responses.

Experiments using the gating task support accounts in which the recognition of onset-embedded words is delayed until after their acoustic offset. Grosjean (1985) gated through test words measuring isolation points and recognition points for low-frequency monosyllables and frequency-matched bisyllables in minimal sentence contexts. Grosjean found that many monosyllabic words were not isolated or recognised until after their acoustic offset. For instance the isolation point for the word *bun* in the sentence "*I saw the* ***bun*** *in the store*" came at the offset of the following word in the sentence – approximately 150ms after the offset of the word.

Although this result may suggest that acoustic cues to rule out long words are not present in these stimuli, it is unclear whether this conclusion can be drawn in the absence of direct comparisons of matched short and long word stimuli. Although some incorrect responses in the Grosjean study were longer words that contained the target (for example, responding *bunny* for the word *bun*), there was no comparison of responses to stimuli containing this longer word. Therefore, these experiments would be insensitive to effects produced by acoustic differences between short and long words. Similarly, although the presence of long word responses to short word stimuli may be taken as evidence that responses were not entirely biased towards short word responses, it is unclear how to evaluate short word biases without investigation of responses to long words which contain onset-embeddings.

A gating experiment carried out by Bard and colleagues (Bard, Shillcock & Altmann, 1988) evaluated the extent of delayed recognition for more naturalistic speech stimuli. When gating through samples of connected speech a word at a time they found that listeners typically failed to identify some 20% of words (mostly closed-class items) until after their acoustic offset. However, since stimuli were presented to participants a word at a time in this study, speech was explicitly segmented during presentation. Consequently, questions regarding the ambiguity of words embedded at the onset of longer words are not addressed. Furthermore, since short word stimuli would be cut off at their acoustic offset (making the length of the word explicit), the results cannot evaluate short biases during lexical identification.

Gating experiments have provided valuable information on the relationship between the acoustic signal and lexical access in connected speech. However, there has, thus far, been no systematic investigation of the recognition of embedded words in connected speech where competition between monosyllables and the longer words in which they are embedded has been controlled for. Consequently, the utility of the acoustic cues to word length that was described in Chapter 2 remains unclear. Similarly, without direct comparisons of responses to short and long target words, it is unclear whether short word biases are present in these experiments. Further gating experiments with materials designed to investigate the recognition of onset embedded words are therefore necessary.

4.1.2. Word-spotting and auditory lexical decision

The word-spotting task, as reviewed by McQueen (1996), appears tailor-made for the investigation of lexical segmentation. The task requires subjects to listen to (usually bisyllabic) nonword strings such as /mɪntəf/ and press a button if they detect a monosyllabic word (in this case *mint*) embedded at either the onset or the offset of the string. Since subjects are not told the identity of the word that they are trying to detect, the task resembles the problem that listeners face in trying to segment words in connected speech – where lexical items will be embedded in a longer stream of speech.

However, despite this resemblance between the word-spotting task and recognition in connected speech, the slow reaction times and high error rates suggest that this task is a difficult one for subjects to carry out. In some cases as many as 70% of trials result in an error where subjects fail to make a response to a stimulus containing an embedded word. Perhaps a more appropriate interpretation of this task is as a go/no-go version of the

lexical decision task[1] in which subjects must decide the lexical status of all possible segmentations of a bisyllabic non-word into two words. In some cases, participants are told the location of the embedded word beforehand (i.e. whether the word is at the start or end of the nonword) thereby reducing the number of alternative segmentations to consider. However, making a response in word-spotting task is still likely to require multiple segmentations of the stimulus, followed by a lexical decision on each potential segmentation.

A necessary assumption in order to interpret reaction time data obtained with word-spotting is that it is the process of segmentation that produces differences in reaction time for different conditions. However, since comparisons of auditory lexical decision responses can be affected by choosing an inappropriate position from which to measure response times (see Goldinger, 1996b for further discussion) there may be an additional confound from the lexical decision component of the task. Furthermore, the task only provides simple measures of processing time (RT and error rate) and therefore ignores the temporal properties of the stimulus that subjects are responding to. This may make it difficult to interpret results in terms of properties of specific sections of the stimuli.

Experiments carried out by Cutler and Norris (1988) used the word-spotting task to investigate whether the lexical stress of a subsequent syllable affected the detection of a word embedded at the onset of a bisyllabic non-word. For CVCC words (such as *mint*) detection latencies were slower where the word was embedded in a bisyllable with a stressed first and second syllable (stimuli such as /mɪntɛɪv/) than in words with a weak or unstressed second syllable (stimuli such as /mɪntəf/). Cutler and Norris interpret this effect as indicating that stressed syllables are used as a cue to the segmentation of connected speech and consequently that stimuli such as /mɪntɛɪv/ are segmented into words as [mɪn][tɛɪv], producing slower latencies for the detection of the word *mint*.

This finding suggests that information coming in after the offset of a word assists recognition. However, with respect to potential acoustic cues to word length or word boundaries, this study is unable to provide any information about what sections of the speech stream played a role in aiding recognition. For instance there is a potential

---

[1] Thanks to Billi Randall for discussion of this interpretation of the word-spotting task.

confounding factor in the stimuli, whereby items that have a stressed second syllable and start with a voiceless stop will be aspirated in stressed syllables. This is of particular relevance, since Christie (1974) reports that the aspiration of voiceless stops in syllable initial position provides a strong cue to the detection of word boundaries. Thus it is unclear whether effects reported by Cutler and Norris (1988) as being caused by metrical stress of syllables of these stimuli reflect the operation of a metrical segmentation mechanism or instead arise through the introduction of allophonic variation that provides an acoustic cue to a word boundary.

Other studies have used word-spotting to show that competition from other lexical items has an inhibitory effect on the identification of monosyllabic words in longer strings. For example, studies by McQueen, Norris and Cutler (1994) in English and Vroomen, van Zon and de Gelder (1996) in Dutch, show that response times were significantly slower for detecting the word *mess* in the sequence /dəmɛs/ which is part of the word *domestic* than in the matched sequence /nəmɛs/ which can not be continued to form a word. This effect, they suggest, results from competition between the word *domestic* and the embedded word *mess*. Similar results were also obtained for words embedded at the onset of a lexical item; participants found it harder to detect the word *sack* in the sequence /sækrəf/ (part of the word *sacrifice*) than in the non-word sequence /sækrək/ – though this effect was only apparent by increased error rates, not by slower response times.

These findings provide evidence for effects of competition between lexical candidates that do not share word boundaries. This result is therefore cited in support of models of lexical segmentation that incorporate inhibitory connections between lexical items. However, effects of the lexical status of continuations of embedded words would also be predicted by the recurrent network models described in the previous chapter (see Figure 3.5 for example). It is therefore unclear that these results can mediate between different theories of lexical segmentation.

Furthermore, these inhibitory effects of competition have not been replicated in experiments using lexical decision. For instance, in experiments reported by Luce and Lyons (1999) it was found that lexical decision latencies to words that contained an embedded word were faster than to matched words that did not contain an onset-embedded word. Although these results suggest that onset-embedded words are activated

during recognition – since the activation of embedded words facilitates lexical decision responses – these findings are contrary to the predictions of competition based models.

Word-spotting and lexical decision tasks have provided evidence that onset-embedded words are activated during the perception of longer words and that following context can influence their identification. However, such results fall short of the systematic comparison of the activation of onset-embedded words and longer competitors that would be required to arbitrate between lexical competition and recurrent network accounts. Furthermore, since these lexical decision and word-spotting experiments used short, bisyllabic stimuli they may underestimate the role of acoustic differences between embedded words and longer competitors during identification. Acoustic cues to word boundaries, such as the duration differences that were described in Chapter 2, may require more extended spoken contexts in order to be processed effectively. It is therefore possible that effects of these acoustic cues will only be apparent where full sentences can be used as experimental stimuli.

4.1.3. Cross-modal priming

One method of assessing the activation of competing interpretations of words within spoken sentences is through the cross-modal priming of lexical decision responses. As pioneered by Swinney, Onifer, Prather and Hirshkowitz (1979) this task has been used to assess the on-line activation of the different meanings of homophonous words like *bank*. By comparing lexical decision RTs to words that are related to the different meanings of *bank* (for example the target words RIVER and MONEY) following either the ambiguous test word or an unrelated control prime, Swinney was able to assess the degree to which different meanings were activated. Comparing the priming effect obtained at different positions in the speech stream allows investigation of the time course with which contextually appropriate meanings of homophonous words are selected.

In the literature on spoken word recognition, cross-modal priming has also been used to investigate the lexical access process for words that have clear meanings, but which may be temporarily ambiguous in the speech stream, such as cohort competitors like *cabin* and *cabbage* (Gaskell & Marslen-Wilson, in press; Marslen-Wilson, 1990; Zwitserlood, 1989; Zwitserlood & Schriefers, 1995). In priming experiments target words are commonly semantically and/or associatively related to different meanings of the target (Moss, Ostrin, Tyler & Marslen-Wilson, 1995; Shelton & Martin, 1992; Tanenhaus, Burgess &

Seidenberg, 1988). Where alternative interpretations are orthographically distinct lexical items it is possible simply to repeat the prime word as the target – see Tabossi (1996) and Zwitserlood (1996) for a review of different variants of the cross-modal priming task and Gaskell & Marslen-Wilson (submitted) for a direct comparison of cross-modal semantic and repetition priming of cohort competitors.

Experiments using cross-modal priming have demonstrated that words embedded at the offset of other words are activated during the recognition of connected speech. This finding has been inferred from the significant priming of words related to these offset-embedded words. For instance, Shillcock (1990) demonstrated significant priming of the target word RIB by sentences containing the word *trombone* (via the embedded word *bone*). This has been confirmed using single word presentations (Luce & Cluff, 1998; Vroomen & de Gelder, 1997) - though experiments using repetition priming have failed to replicate this finding (Marslen-Wilson et al., 1994). These results have been taken as evidence that non-aligned lexical hypotheses are activated during connected speech – a finding that would challenge accounts of lexical identification (such as sequential recognition accounts) in which only words that start at a known word onset are activated during recognition.

None of these experiments, however, measured the activation of words related to the longer word as well as the embedded word. Consequently, it is unclear whether significant priming indicates that the perceptual system fails to distinguish between the embedded word and its longer competitor (a finding no current account of spoken word recognition would predict) or merely that participants in these experiments become aware of the relationship between prime and target by some post-access strategic process (Shelton & Martin, 1992).

One study that did compare activations of both appropriate and inappropriate segmentations of potentially ambiguous stimuli was carried out by Gow and Gordon (1995). They compared the priming of associatively related targets (FLOWER or KISS) from phonemically identical sequences such as *tulips* and *two lips*. Results demonstrated priming of the target KISS from two word stimuli (*two lips*) though not from single word stimuli (*tulips*). Conversely, targets (such as FLOWER) related to the long word were primed by both single word (*tulips*) and two word (*two lips*) stimuli.

By comparison with the results of Shillcock (1990), this failure to find priming between the offset-embedded word in the auditory prime *tulips* and the target KISS is surprising. Most accounts of lexical segmentation predict that stimuli in which both syllables are words would produce more mis-segmentations than words such as *trombone* used in the Shillcock study. One interpretation of the discrepancy between these findings is that in experiments where both appropriate and inappropriate interpretations of the prime stimuli are probed (as was the case for the Gow and Gordon study), priming effects are more resistant to effects of strategic expectations by participants.

The failure to observe priming of offset-embedded words in the Gow and Gordon (1995) study is interpreted as evidence for sensitivity to acoustic cues that mark word onsets. However, since prime sentences in their experiments continued after the presentation of the visual targets, it is unclear whether information in the 700ms of following context that participants heard whilst making a response may also play a role in the priming effects obtained in these experiments. By allowing prime stimuli to continue after the presentation of the target, these studies fail to control the amount of information in the speech stream that can be processed by participants. Hence conclusions regarding the importance of one particular section of the speech stream may be questioned. Furthermore, since these experiments only measured the activation of words embedded at the offset of a longer word, they do not provide constraining evidence regarding the important issue introduced earlier in the chapter of whether the recognition system is biased towards short word hypotheses in the identification of onset-embedded words.

Experiments by Tabossi, Burani & Scott (1995) in Italian also investigated ambiguities created by words being embedded at the onset of longer words. They found equal priming for associates of the word *visite* (visit) from sequences containing that word and from sequences where *visite* was formed by sections of two adjacent words (as in the sequence *visi tediati* (faces bored)). This effect was also found in a subsequent experiment where an allophonic cue to the presence of a word boundary was present in these two word stimuli. These results confirm the findings of Gow and Gordon (1995) that lexical items made by combining two adjacent words are accessed in connected speech.

The results obtained by Tabossi and colleagues suggest that whatever acoustic cues to word boundaries may be present in Italian – and they may be more weakly marked than in English (Bertinetto, 1981) – do not allow the system to distinguish words created from concatenating two adjacent lexical items from a single lexical item. Priming of meanings

related to words created from concatenated lexical items was even obtained in the case where an allophonic cue to a word boundary was present in their stimuli. However since, as in other studies, Tabossi et al. (1995) did not investigate the identification of short words that are embedded in longer lexical items – such as *visi* (faces) in *visite* (visit) it is unclear if alternative explanations of these results based on strategic effects can be ruled out.

Several studies have demonstrated ambiguity created by the absence of explicitly marked word boundaries in connected speech. However, the only study (Gow & Gordon, 1995) to measure the severity of this ambiguity (by comparing whether listeners are able to distinguish appropriate from inappropriate segmentations) found rather less ambiguity than other results might have predicted. However, this study focussed on words embedded at the offset of a longer word rather than the onset-embedded words that are more critical for the theoretical accounts under discussion here.

Consequently, there remains a conflict between work describing acoustic differences that might provide a means by which to discriminate short and long words and models that assume that onset-embedded words are ambiguous at their offset. This conflict remains unresolved since experiments have not compared the activation of short and long words that share the same onset during connected speech. For the same reason, it is not possible to draw strong conclusions regarding the differential predictions of recurrent network and lexical competition accounts regarding short word biases during the identification of onset-embedded words. Without comparison of the activation of short and long words during the processing of stimuli containing either short or long words it is unclear how a bias towards short word interpretations could be established.

4.2. Experimental design

In the various experiments reviewed here, a variety of techniques were used to investigate the processing of monosyllables embedded in longer words and of longer words that contain onset-embeddings. However, none of these studies have adequately investigated the time course with which embedded words are recognised in the speech stream. Consequently, although there is evidence showing the activation of longer competitors during the recognition of embedded words (and vice versa), there are few firm conclusions about how these competing lexical hypotheses are resolved on-line and

whether acoustic cues that distinguish short from long words play a role in the identification of embedded words.

In order to compare the time course of identification of onset-embedded words and longer competitors in connected speech it is necessary to use methods in which sentences can be presented to participants. This rules out tasks such as word-spotting in which only bisyllabic stimuli can be used. Furthermore, to track the activation of alternative lexical hypotheses across precisely measured sections of speech, the standard form of the cross-modal priming task, in which stimuli continue after the presentation of the visual target, is also unsuitable. For this reason all the experiments reported in this thesis used sentence fragments, with stimuli being cut off at positions of interest in the speech stream. By comparing interpretations of short and long stimuli at specific points in the stimuli, it is possible to evaluate the extent to which different sources of information in the speech stream contribute to the identification of onset-embedded words and longer competitors. Concerns may be raised that by cutting off speech a cue to the location of a word boundary is provided (in the silence that follows the offset of the gated speech). However, since stimuli will be cut-off during a word as well as after its offset, participants who attempted to use such a cue would find it unhelpful.

In the standard form of the gating task, participants are presented with progressively longer fragments of speech and have to write down the words that they can identify at each gate. Comparing responses to stimuli containing short and long words allows investigation of the extent to which onset-embedded words create ambiguity between short and long lexical candidates. Investigating how listeners' interpretations change across gates enables measurement of how the recognition of these words is affected by different sources of information that are available in the speech stream.

Concerns have been raised about the effect of successive presentations of the same stimuli and the off-line nature of responses in gating (Cotton & Grosjean, 1984; Tyler & Wessels, 1985; Walley, Michela & Wood, 1995). Consequently results obtained in a gating study will be compared with experiments in which cross-modal priming is used to provide an on-line measure of lexical activation at each gate (Gaskell & Marslen-Wilson, 1997; Zwitserlood, 1989; Zwitserlood & Schriefers, 1995). Since the competing interpretations of embedded words are distinct lexical items, repetition priming of lexical decision can be used to provide a measure of the activation of the prime that avoids

possible confounds that could be produced by differences in semantic or associative relatedness.

In order to conclude that listeners are able to detect subtle acoustic differences (such as syllable duration) between short and long word stimuli, any confounding factors must first be ruled out. The initial series of experiments reported in this thesis will therefore use stimuli that maximised the potential ambiguity between short and long words. This was achieved by using *lexical garden-paths*, stimuli in which speech coming after the offset of an embedded word continues to match a longer competitor – for example, the sequence *cap tucked* in which the onset of the following word matches the onset of the second syllable of the competitor *captain*. In this way, even allowing for co-articulation, syllables which can either be a monosyllabic word or the start of a longer word will be as acoustically similar as possible (except for the acoustic differences that were described in Chapter 2).

The activation of short and long words for these lexical garden-path sequences was compared with matched sentences including longer lexical items that contained these embedded words at their onset. It should therefore be possible to rule out strategic accounts of the priming effects observed and allow investigation of whether the on-line activation of words in connected speech is biased towards short words, as predicted by lexical competition accounts of spoken word recognition.

4.2.1. Stimuli

Short and long word pairs

Starting from the CELEX lexical database (Baayen, Pipenbrook & Guilikers, 1995) bisyllabic words were selected which had a morphologically unrelated monosyllable embedded at their onset (e.g. *captain*, containing the embedded word *cap*). Only bisyllables with a metrically stressed first syllable were chosen and in all cases the monosyllabic word exactly matched the syllabification of the longer word. Pairs such as *cat* and *cattle* were excluded, since by the maximal-onset principle (Selkirk, 1984) *cattle* would be syllabified as [kæ][t̪l̪] with a boundary within the embedded word. The monosyllables chosen all had at least three letters and consisted of three or more phonological segments.

Items were rejected if they were not of the same syntactic class, or if either word was orthographically unusual (such as the pair *pizza* and *peat*). However, items were not required to be fully orthographically embedded (pairs such as *track* and *tractor* were included as well as *captain* and *cap*[2]). A further criterion was that at the offset of the monosyllable there should be a limited number of longer items that contain the embedded word (this excluded items like *con* embedded in *concrete* where *con-* as a prefix is found in over 200 words). The mean number of words in which the monosyllables were embedded was 12 items (maximum, 43; minimum, 1). In all cases the long word was the most frequent word in this group. To avoid biases towards either short or long words in our test stimuli, pairs were rejected if they did not occur with approximately equal frequency in the language. Across the set of 40 pairs of words that were used in the experiments a paired t-test showed that there were no significant differences in the frequency of the pairs of short and long words (mean frequency short words = 35/million, long words = 25/million, $t(39)=1.07$, $p>.1$).

Test sentences

Given that one goal of these experiments was to test for effects of acoustic cues (such as greater syllable duration in monosyllabic words) that may only be contrastive by comparison with prior context, items were placed approximately in the middle of test sentences. Each sentence contained an average of 6 syllables of neutral preceding context (range 3 to 11 syllables) so that subjects would be able to use ongoing prosodic cues that were present in these stimuli. Cloze tests were carried out on these sentence contexts to ensure neither of the target words were predictable. Each test sentence had several words after the test item, so that listeners would not be able to use prosodic cues to the end of a sentence as potential evidence of a short rather than a long word being present. No major clause boundaries followed the short test words to avoid possible intonation differences

---

[2] Auditory priming experiments by Jakimik, Cole and Rudnicky (1985) report differences in priming from a bisyllables to an onset-embedded monosyllable depending on whether the monosyllable was spelt in the same or different way. However, these effects were observed at an SOA of two seconds, much longer than the typical SOA used in cross-modal experiments. This suggests that strategic effects may have contributed to these results.

that might distinguish them from their longer competitors (Christophe, Guasti, Nespor, Dupoux & Ooyen, 1997).

In order to create as much ambiguity in these stimuli as possible and provide the most stringent test of the claim that there are acoustic cues that distinguish between short and long words in connected speech, it is necessary to exclude acoustic differences in the embedded syllables caused by co-articulation from following segments. Continuations for the short word stimuli were therefore chosen that started with the same onset segment or segments as the second syllable of the longer word. An example pair of sentences from the set of 40 used in the experiments are shown below (with target words emphasised):

(Short word) The soldier saluted the flag with his **cap** tucked under his arm.

(Long word) The soldier saluted the flag with his **captain** looking on.

The set of 40 sentence pairs shown in Appendix A were recorded by the author onto digital audio tape (DAT) in a sound-proof booth. Each pair of sentences was recorded successively to help ensure that the sentences were produced with near-identical intonation patterns and without prosodic breaks after the monosyllabic words. These recordings were then passed through an anti-aliasing filter and digitised at a sampling rate of 22khz using a DT2821 sound-card attached to a Dell PC.

4.2.2. Acoustic analysis and alignment points

In order to determine whether listeners are sensitive to duration differences between syllables in monosyllabic and bisyllabic words, it is important to ascertain whether these and other possible acoustic differences are present in these stimuli. Furthermore, given the intention to compare interpretations of stimuli that contain embedded words and longer competitors, it is necessary to make these contrasts between stimuli containing equivalent acoustic-phonetic information. Consequently, *alignment points* (hereafter *AP*) were set up at phonetically equivalent positions in each sentence. Measuring acoustic differences and differences in participants' interpretations with respect to these alignment points helps ensure that results reflect cues to the location of word boundaries and are not artefacts caused by information from subsequent segments or syllables in either set of stimuli.

The start and the end points of each sentence were marked using the BLISS speech editing system (Mertus, 1989). Additional markers were placed at the onset of the target

word – a point at which each pair of sentences should be as identical as possible. This similarity was confirmed by listening to the sentence onsets and by visual inspection of the speech wave and fundamental frequency (F0) contours for a selection of the test items. Acoustic analysis of the duration of the word immediately preceding this marker (usually an article such as *the*) showed no reliable differences in duration (short word duration = 97ms, long word duration = 98ms; t(39)=0.070; p>.1).

Cursor positions	Measure	Short word stimulus	Long word stimulus	Difference
onset – AP_1	Duration (ms)	291	243	**
	Voicing time (ms)	184	165	**
	F0 (hz[a])	112	113	ns
	mean RMS[a]	2476	2657	(*)
$AP_1 – AP_2$	Duration (ms)	79	77	ns
$AP_2 – AP_3$	Duration (ms)	42	44	ns

Table 4.1: Alignment points and acoustic measurements for stimuli in experiment 1. [a]F0 and RMS energy for voiced section of syllable only. Statistical significance: ns p>.1, (*) p<.1, ** p<.01

The second alignment point (AP_2) was placed following the onset segment (or segments) of the second syllable. A paired t-test showed that there were no significant differences in the duration of onsets which were word initial in the short word stimuli and word medial in the long word stimuli (t(39)=0.42, p>.1). This contrasts with the stimuli used by Gow and Gordon (1995), as well as with other findings in acoustic phonetics showing the significantly greater duration of word-initial segments (Klatt, 1976). This difference in the acoustic properties of our stimuli may be attributable either to the absence of prosodic boundaries before the onset-segments in our stimuli or to acoustic differences being obscured by subsequent phonetic differences.

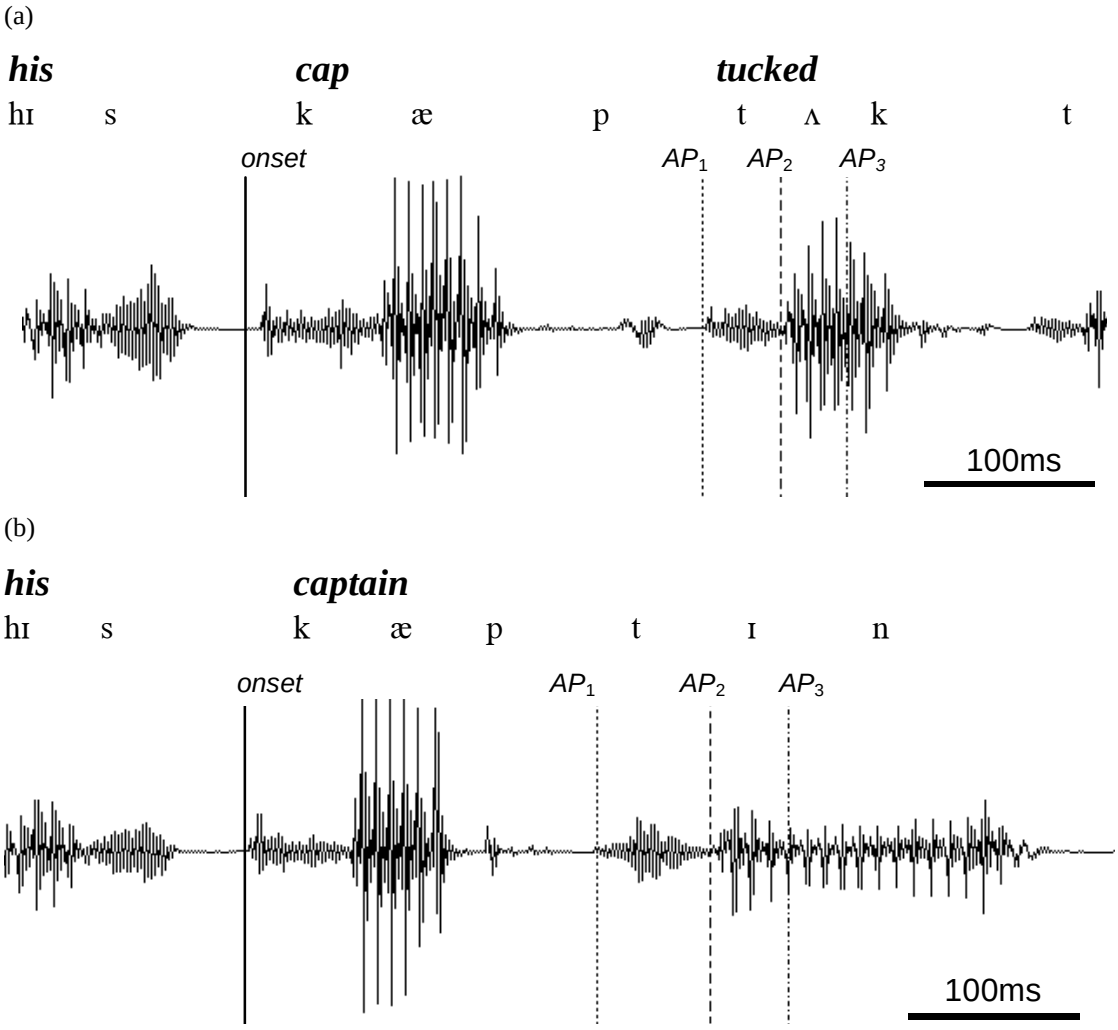

Figure 4.1: Speech waveforms and alignment points for the stimuli in experiment 1. *onset* – onset of target word, AP_1 – offset of target word, AP_2 – onset of second syllable, AP_3 – vowel of second syllable. Stimulus items are:

(a) *"The soldier saluted the flag with **his cap tucked** under his arm."*
(b) *"The soldier saluted the flag with **his captain** looking on."*

The first alignment point (AP_1) was placed at the offset of the first syllable of the prime word (at the end of the closure of the final segment). Measurements of the acoustic duration of the first syllable (from the onset of the target word to AP_1) shows the expected difference in acoustic duration in monosyllabic and bisyllabic words, as shown in Table 4.1. This 48 ms difference in duration was highly significant across the 40 test items ($t(39)=9.35$, $p<.01$). Further analysis confirmed that a large proportion of this difference could be accounted for by differences in the duration of the vowel. Analysis of the duration of the voiced portion of the target syllable showed reliable differences in duration ($t(39)=3.11$, $p<.01$). However, this difference apart, by the criteria described by

Fear, Cutler & Butterfield (1995) there was no major difference in the amount of stress applied to these syllables. There were no significant differences in the fundamental frequency of the two syllables ($t(39)=0.53$, $p>.1$) and measurements of the acoustic energy in the vowel of these syllables showed a minor difference in mean RMS energy. Syllables at the onset of long words had greater energy, suggesting they were more strongly stressed than the equivalent syllable as a monosyllable though this effect was marginal ($t(39)=2.01$, $p<.1$).

The third alignment point (AP_3) marks the earliest location where the stimuli are expected to contain different phonemes. This marker was placed 4 pitch periods into the vowel of the second syllable, for instance 52ms into the vowel [ʌ] of *cap tucked*. Again, there were no overall differences in the duration of this section of vowel ($t(39)=1.35$, $p>.1$). An example of the position of these alignment points is illustrated in Figure 4.1.

4.3. Experiment 1 – Gating

The first experiment carried out to test these hypotheses used the gating task reviewed previously. This method was used to assess the role of the acoustic differences described in

Table 4.1 in identifying stimuli that are potentially ambiguous between a bisyllabic word and an onset-embedded monosyllable. For instance, if listeners are sensitive to duration differences in the syllable /kæp/ for words like *cap* and *captain* it would be expected that responses to the pairs of stimuli will diverge at or before AP_1 – the offset of the first syllable. However, if sub-phonemic cues in the production of segments that are word onsets in short word stimuli and word medial in long word stimuli are of greater importance (as suggested by Gow and Gordon (1995) and Nakatani and Dukes (1977)) then responses will diverge at AP_2. Finally, if recognition requires phonemic mismatch between short and long stimuli then listeners would be unable to distinguish these paired stimuli prior to AP_3; this being the earliest point at which our experimental stimuli differ phonemically.

4.3.1. Method

Participants

Twenty-four English speakers from the Birkbeck Speech and Language subject pool were tested. Most were University of London students, all were aged between 18 and 45 and were paid for their participation. All were native speakers of British English and had normal hearing and no history of language impairment.

Design and materials

Experiment 1 used the standard gating method in which participants make written responses to successively longer fragments of recorded speech. For the stimulus sentences used, where more than one word needs to be identified, one set of dependent variables will be the word or words identified by participants following each fragment or gate. In addition, participants were asked to rate the confidence of their responses using a 9 point scale ranging from 1 (guess) to 9 (confident).

The independent variable was whether the sentence fragments played to the participants contained a short or a long word. Each sentence in the experiment was played out as 10 successive fragments with the entire sentence being presented from the start to a cut-off point. Cut-off points were the three alignment points described previously and shown in Figure 4.1, as well as two initial gates 50 and 100ms before AP_1 and five gates (designated gates 6 to 10) placed 50, 100, 200, 300 and 400ms after AP_3. In all cases, it was expected that the gate 400ms after AP_3 would contain sufficient information to enable participants to identify the word following the target item.

The 40 pairs of test sentences (containing a short or long word) were pseudo-randomly divided into two experimental versions such that each version contained only one member of each stimulus pair. An additional 20 sentences were added to each version; four were used as practice items to acquaint participants with the task and the remaining 16 items were fillers to distract from the large number of embedded words in the test sentences.

Procedure

Participants were tested in groups of between 2 and 4, sitting in booths in a quiet room. They were provided with answer books containing the onset of each sentence up to (but not including) the target word and were instructed to identify the word or words that

continued each sentence based on the speech they heard at each gate. Participants were instructed to make a response for every fragment of speech that they heard and to write down as many words as they could hear in each continuation. They were also asked to provide a confidence rating on a 9 point scale with 1 being a guess and 9 representing confidence.

Sentences were played from a PC equipped with a DT2821 soundcard through closed-ear headphones. Fragments were played at 6 second intervals, with an extra 2 second being provided at gates after AP_3 to allow participants more time to write down words coming after the target item. The 56 test and filler sentences were divided into four blocks, each lasting approximately 20 minutes with a five minute break being given between blocks and a 10 minute break at the half way point. The whole testing session, including the practice items, lasted approximately two hours.

4.3.2. Results and discussion

Data from two participants (one from each version) were rejected for failing to comply with the instructions to make a response for each fragment they heard. The remaining 8800 responses (22 subjects, 40 items, 10 responses/item) were coded for the identity of the target word and subsequent words along with the confidence rating and analysed to investigate the proportion of responses (by participants and by items) that matched either of the target words.

All participants produced correct responses for the majority of the test items by the final gate. However, there were three items, (*ban*, *bran* and *win*) for which the short word stimuli were not recognised by 50% of participants at the final gate. Consequently these items (and their corresponding bisyllables, *bandage*, *brandy* and *winter*) were not analysed.

The proportion of responses at different gates that matched either the short or long target words are shown in Figure 4.2. As can be seen in the graph, at early gates (up to and including the offset of the first syllable at AP_1), the majority of participants' responses match the short target word (e.g. CAP). Even at the first gate, 100ms before the offset of the embedded word (AP_1) subjects hear enough of the target word to identify the first syllable.

These results also show that, following the large number of short word responses at early gates the proportion of responses that match the short word target decreases at AP_2. Since the isolation point as conventionally calculated from gating experiments is the average gate at which subjects produce the correct response and then don't change their response at subsequent gates, this U-shaped pattern of short word responses will produce a bi-modal distribution of isolation points for the short word stimuli. For approximately 65% of items and/or participants, isolation points will be at a gate after AP_2, while the remaining 35% of isolation points (where responses don't change subsequently) will be substantially earlier. Consequently, isolation points will be bi-modally distributed with measures of central tendency being unrepresentative of the behaviour of any given participant or item. Statistical analysis will therefore use the proportion of responses that matched the target words at each gate as the dependent measure.

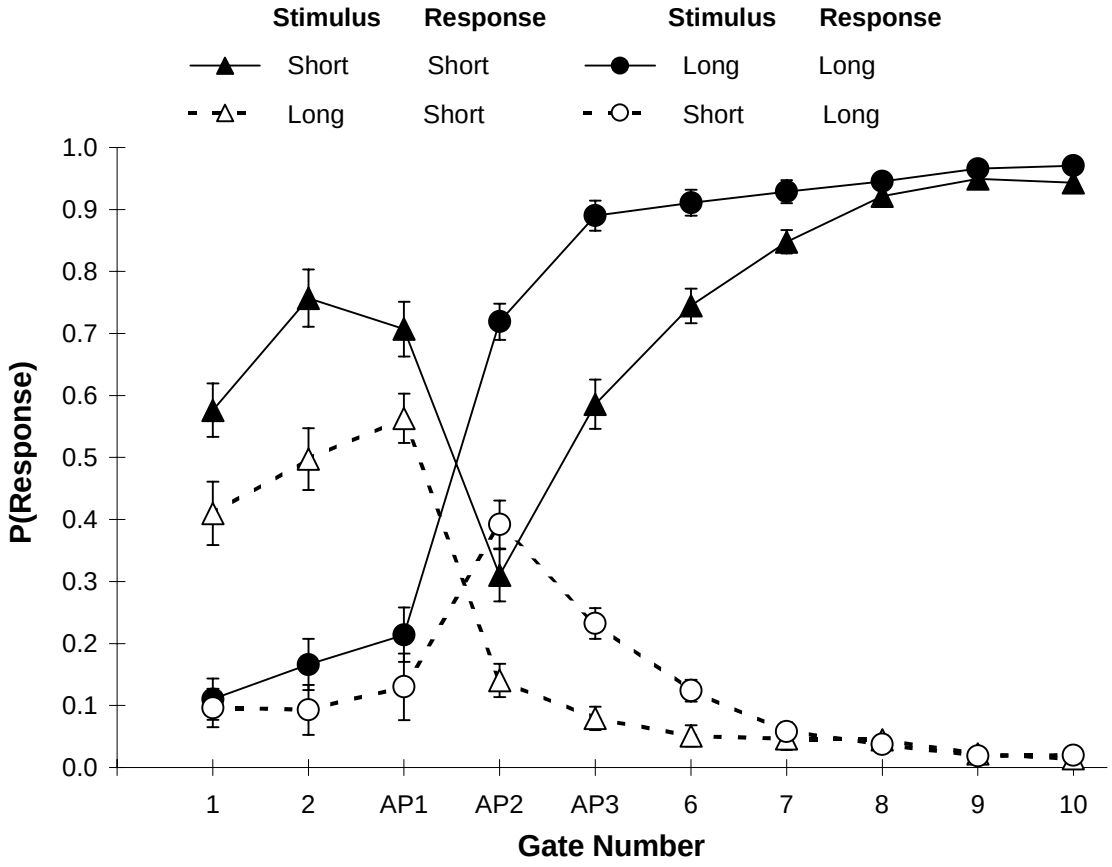

Figure 4.2: Experiment 1 – Gating. Proportion of responses matching short and long target words for stimuli containing short and long words. Error bars are 1 standard error.

Acoustic cues to word length

Although at early gates there is an overall bias towards short word responses, there are differences in the proportion of short word responses made at early gates depending on

which of the pair of stimuli participants were hearing (as shown in Figure 4.2). ANOVA on the proportion of short word responses across the three gates up to AP_1, using the repeated measures factors of stimulus type (short or long word) and gate number (gate 1, 2 or AP_1) shows that significantly more short word responses were made to short word stimuli than to long word stimuli ($F_1[1,20]=60.32$, $p<.001$; $F_2[1,35]=26.86$, $p<.001$). There was also a significant effect of gate ($F_1[2,40]=30.37$, $p<.001$; $F_2[2,70]=9.60$, $p<.001$) reflecting the greater number of short word responses at the later gates and an interaction between stimulus type and gate significant by participants and not items ($F_1[2,40]=5.81$, $p<.01$; $F_2[2,70]=2.35$, $p>.1$).

Parallel effects were found in the analyses of long word responses. Again over the first three gates there were significantly more long word response to long word stimuli than to short word stimuli ($F_1[1,20]=7.34$, $p<.05$; $F_2[1,35]=4.69$, $p<.05$). There was also a significant effect of gate, reflecting the increasing number of responses matching either target word across these three gates ($F_1[2,40]=14.40$; $p<.001$; $F_2[2,70]=8.39$, $p<.001$) and an interaction between stimulus type and gate – though this was of only marginal significance by items ($F_1[2,40]=6.48$, $p<.01$; $F_2[2,70]=2.82$, $p<.1$).

These effects of stimulus type suggest that subjects are able to use acoustic differences to discriminate the initial syllables of short and long words. The significant effects of gate show the sensitivity of the gating task to the arrival of new acoustic information, while interactions between stimulus type and gate suggest that acoustic information that allows subjects to discriminate between short and long words becomes more available throughout these gates.

Similar effects can also be observed in the confidence ratings. However, given the small number of long word responses at the first three gates, there were insufficient data points to analyse ratings from these responses. Ratings data for the first three gates were averaged for short word responses and are shown in Table 4.2. Analysis of these confidence ratings confirm the effect of stimulus type and gate shown in the analyses of word responses. Participants produced significantly higher confidence ratings for short word responses to short word stimuli than to long word stimuli ($F_1[1,20]=14.56$, $p<.001$; $F_2[1,27]=9.86$, $p<.01$). There was also a significant effect of gate number ($F_1[2,40]=69.25$, $p<.001$; $F_2[2,54]=106.70$, $p<.001$) and no significant interaction between these variables.

	Confidence Ratings		
Stimulus Type	Gate 1	Gate 2	AP_1
Short Word	3.21	4.73	5.82
Long Word	2.96	3.91	5.07

Table 4.2: Mean confidence ratings for short word responses to short and long word stimuli at initial gates. 1 = guess, 9 = confident

Delayed recognition and response biases

Despite these differences, the recognition of embedded words still appears to be delayed compared to the identification of the longer words in which they are embedded. It is only at gate 8 that there is no significant difference between the proportion of correct responses given to short words and long word stimuli ($t(36)=0.96$, $p>.1$). This delay in recognition appears to result from competition from longer interpretations, since at AP_2 (the onset of the second syllable) participants gave many more long word responses to the short word stimuli than at any previous gate. It is only when there is clear mismatch between the short word stimuli and the long target words (for instance the phonemic differences between *cap tucked* and *captain*) at AP_3 and beyond that subjects are able to revise these hypotheses and identify the short words. This result confirms the role of information coming after the offset of a word in identifying embedded words as suggested by the results of gating (Grosjean, 1985) and word-spotting experiments (Cutler & Norris, 1988).

However, it is also necessary to consider possible effects of response biases. In off-line gating experiments, participants may be biased towards producing the shortest single word that accounts for all the speech segments that they can hear in the current fragment (Tyler, 1984). Such a bias could account for two properties of this data. Firstly, it would explain the predominance of short word responses at the initial three gates. Since participants are hearing some or all of a single syllable such as [kæp], they will be inclined to produce monosyllabic words in response, even in cases where the acoustic duration of the syllable might suggest that it was more likely to come from a bisyllabic

word. This bias may lead to under-estimating the effectiveness of the cue provided by syllable duration, since participants will produce fewer bisyllabic responses at early gates.

The second aspect of the results that could be interpreted in terms of this single word bias is the large increase in long word responses at AP_2. A bias towards responses that account for as much speech as possible would encourage participants to respond with a longer word (e.g. *captain*) for stimuli such as [kæpt]. Response biases might therefore increase the number of long word responses at AP_2 and hence exaggerate the amount of competition between short and long hypotheses – producing the delayed recognition observed for onset embedded words.

In summary, the results of Experiment 1 suggest that listeners are sensitive to acoustic cues that distinguish between syllables of short and long words. The significant differences before the offset of the first syllable confirm that subjects are able to discriminate short and long words before they diverge phonemically. This challenges the assumptions of models in which the onset of embedded words are assumed to be indistinguishable from the initial syllables of longer words.

Despite these acoustic differences, participants display an overall preference towards short word interpretations at early gates. At later gates, where continuations match longer words, long word interpretations are preferred. This appears consistent with models that incorporate lexical-level competition – with short and long words competing during identification even where they do not share word boundaries. Furthermore, the overall bias towards short word responses at early gates is as predicted by lexical competition models.

However, given the questions raised earlier about the possible role of response biases in determining gating responses, it is important to use more on-line methods to measure the activation of embedded words and longer competitors in connected speech. With this gating experiment as background, the next chapter therefore reports a series of experiments using an on-line task (cross-modal priming) to probe the activation of onset-embedded words and longer competitors during connected speech.