LEXICAL SEGMENTATION AND AMBIGUITY: INVESTIGATING THE RECOGNITION OF EMBEDDED WORDS

Matt Davis and William Marslen-Wilson
MRC Cognition and Brain Sciences Unit, 15 Chaucer Road, Cambridge, UK

Gareth Gaskell
Department of Psychology, University of York, York, UK.

Abstract:

The lack of acoustic markers of word boundaries may create ambiguity between words like *cap* and the start of longer words like *captain*. This potential ambiguity has motivated models of spoken word recognition in which lexical competition allows information after the end of an embedded word to assist identification. We review the results of cross-modal priming experiments demonstrating that additional acoustic cues assist listeners in distinguishing embedded words from longer competitors. Recurrent network simulations in which bottom-up cues and following context are combined in the identification of onset-embedded words show an activation profile consistent with the priming data. We suggest that recurrent networks provide an appropriate model of the recognition of embedded words in connected speech.

Embedded Words and Ambiguity

Two accounts have been proposed of how word recognition can contribute to lexical segmentation:

1. sequential recognition (Marslen-Wilson & Welsh, 1978)
2. lexical competition (TRACE McClelland & Elman, 1986) (Shortlist Norris, 1994)

The temporary ambiguity of onset-embedded words (*cap* in *captain*) is used as an argument against sequential accounts (Luce, 1986; McQueen, Norris, Briscoe & Cutler, 1995)

Competition between short and long words delays identification such that information after word offset can be used to recognise embedded words (Grosjean, 1985; Bard, Shillcock & Altmann, 1988; McQueen, Norris & Cutler, 1994)

However, acoustic differences between syllables in short and long words may provide an additional cue for the identification of embedded words (Lehiste, 1972; Klatt, 1976)

We tested whether the perceptual system uses these cues in identifying embedded words or whether ambiguity with longer words require delayed identification (Davis, Marslen-Wilson & Gaskell, 1997)

Expts:

Four cross-modal repetition-priming experiments investigated the recognition of onset-embedded words in *lexical garden paths* (sequences matching a long word - e.g. *cap tucked*) and longer competitors (*captain*) in non-biasing sentences. Primes and targets following:

"The soldier saluted the flag with his..."

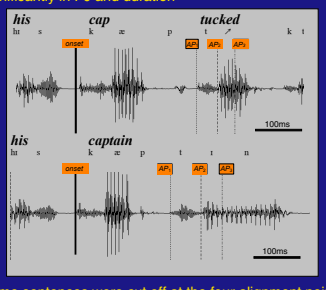
Prime Type	*Prime Word* continuation	Short Target	Long Target
Short Test	***cap*** tucked under his arm	CAP	CAPTAIN
Long Test	***captain*** looking on	CAP	CAPTAIN
Control	***rifle*** by his side	CAP	CAPTAIN

Alignment Points:

Comparisons between short and long word stimuli were made at aligned positions in the paired sequences

*AP*1 - offset of embedded word [kæp]
*AP*2 - onset of following syllable [kæpt]
*AP*3 - vowel of second syllable [kæpt ʌ or [kæptɪ]
*AP*4 - 100ms after *AP*3

Initial syllables in short and long word sequences differed significantly in F0 and duration



Prime sentences were cut off at the four alignment points to probe lexical activation at specific points in the stimuli.

Magnitude and significance of priming at these probe positions is shown in the figures alongside lexical activations predicted by a recurrent network model.

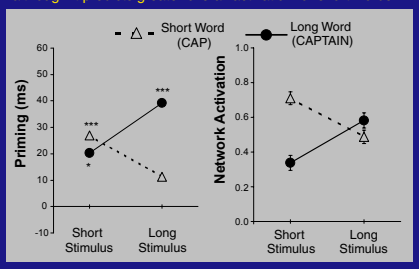
Refs:

Bard, E. G., Shillcock, R. C., & Altmann, G. T. M. (1988). The recognition of words after their acoustic offsets in spontaneous speech: Effects of subsequent context. Perception and Psychophysics, 44, 395-408.
Davis, M. H., Marslen-Wilson, W. D., & Gaskell, M. G. (1997). Ambiguity and competition in lexical segmentation. In Shafto, M. (ed) Proceedings 19th Cognitive Science Society Conference. LEA.
Davis, M. H., Gaskell, M. G., & Marslen-Wilson, W. D. (1997). Recognising embedded words in connected speech: Context and competition. In Bullinaria, et al. (Eds.). Proceedings 4th Neural Computation and Psychology Workshop. London: Springer-Verlag.
Grosjean, F. (1985). The recognition of words after their acoustic offset: Evidence and implications. Perception & Psychophysics, 38(4), 299-310.
Klatt, D. (1976). Linguistic uses of segmental duration in English: acoustic and perceptual evidence. Journal of the Acoustical Society of America, 59(5), 1208-1221.
Lehiste, I. (1972). The timing of utterances and linguistic boundaries.
Luce, P. A. (1986). A computational analysis of uniqueness points in auditory word recognition. Perception and Psychophysics, 39, 155-158.
Marslen-Wilson, W. D., & Welsh, A. (1978). Processing interactions and lexical access during word recognition in continuous speech. Cognitive Psychology, 10, 29-63.
McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. Cognitive Psychology, 18, 1-86.
McQueen, J. M., Norris, D., & Cutler, A. (1994). Competition in spoken word recognition: spotting words in other words. Journal of Experimental Psychology: Learning, Memory and Cognition, 20(3), 621-638.
McQueen, J. M., Cutler, A., Briscoe, T., & Norris, D. (1995). Models of continuous speech recognition and the contents of the vocabulary. Language and Cognitive Processes, 10(3-4), 309-331.
Norris, D. (1994). Shortlist: a connectionist model of continuous speech recognition. Cognition, 52, 189-234.

Results:

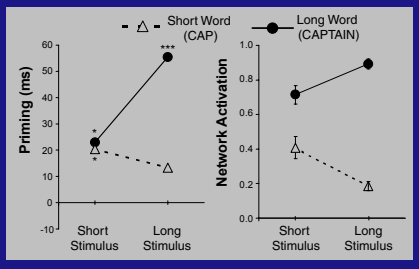
*AP*1 [kæp]

Priming data at *AP*1 shows a significant cross-over interaction, indicating that listeners favour the correct interpretation of these stimuli at the offset of the embedded syllable. The model supplied with additional input cues shows a similar profile - although it predicts greater overall activation for short words.



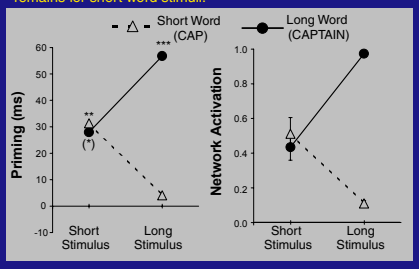
*AP*2 [kæpt]

The onset of the following syllable increases activation for long interpretations - as indicated by greater priming for long word targets and increased activation in the network.



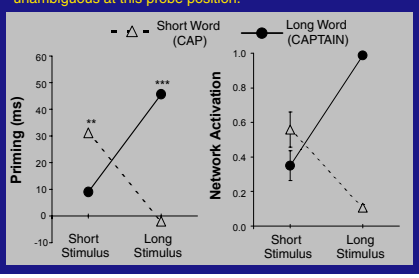
*AP*3 [kæpt ʌ or [kæptɪ]

Despite phonemic mismatch between short stimuli and long targets in the vowel of the second syllable, there is still a bias towards long word interpretations at *AP*3. Some ambiguity remains for short word stimuli.



*AP*4

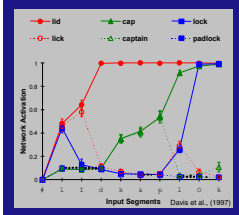
At *AP*4 ambiguity between embedded words and longer competitors has been resolved - shown by the cross-over interaction in the priming data. The similar pattern shown by the model indicates that both short and long word stimuli are unambiguous at this probe position.

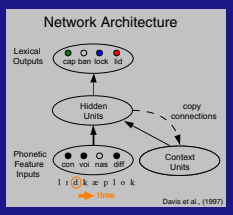


Modelling lexical competition:

TRACE and Shortlist simulate the delayed recognition of embedded words using mutual inhibition between candidates that span word boundaries. Lexical competition allows information 'ruling-out' longer words to 'rule-in' embedded words.

Simple recurrent networks show effects of lexical competition without mutual inhibition. When trained to identify sequences of words (rather than single words) SRNs also show delayed recognition of onset-embedded words. (Davis, Gaskell & Marslen-Wilson, 1997)





Adding acoustic cues:

A simple-recurrent network was trained on a set of 20 lexical items, including onset-embedded words with longer competitors (e.g. *cap* in *captain*)

To simulate the effect of acoustic cues that distinguish short and long words, additional input units were provided representing differences in duration (or F0) between syllables in short and long words.

Three units coded this acoustic cue in a contextually dependent fashion. Duration codes for each sequence will depend on an overall 'rate'. The intermediate code could therefore come from either a short or a long word, but can be disambiguated by preceding context. The ambiguous middle code replaced the short and long codes in 20% of words chosen at random.

Syllable Duration	Code	Network Input	Speech Rate Fast	Speech Rate Slow
short	1	1 1 1	bisyllable	-
medium	2	0 1 1	monosyllable	bisyllable
long	3	0 0 1	-	monosyllable

Ten networks were trained for 500 000 sequences and tested on lexical garden path sequences analogous to those used in the experimental stimuli (e.g. *cap tap*) and sequences containing longer words (*captain*). Test words were in the middle of the sequence, allowing the network to use prior context to discriminate input from short and long words.

In simulating the experimental data, probe positions corresponded to the phonemes at each alignment point. Activations for short and long lexical units are averaged and analysed over the 10 networks and compared to the priming data.

Discussion:

Results from cross-modal priming experiments show that additional acoustic cues distinguish syllables from short and long words; reducing the ambiguity of onset-embedded words. However, longer words are activated by embedded words in lexical garden path sequences, suggesting that following context affects identification.

A simple recurrent network simulates this experimental data where the network is provided with appropriate bottom-up input cues - without requiring direct inhibtion between lexical units.

Acknowledgements:

This work was supported by EPSRC research studentship 94700590 and by an MRC Program Grant to William Marslen-Wilson and Lorraine K. Tyler. Thanks to members of the Centre for Speech and Language at Birkbeck College and in Cambridge for advice and assistance.