

# Lexical Segmentation in Spoken Word Recognition

**Matthew Harold Davis**

Birkbeck College,

University of London

Thesis submitted for the degree of Doctor of Philosophy

March 2000

## Abstract

This thesis examines an important issue in spoken word recognition; how the perceptual system segments connected speech into lexical units or words. Research on this topic has investigated the role of different sources of information in dividing up the speech stream: acoustic cues in the speech signal, statistical regularities in the structure of the language or through the identification of individual lexical items.

This research focuses on cases in which the location of word boundaries may be ambiguous by one or more of these segmentation mechanisms using words embedded at the onset of longer words (such as *cap* in *captain*). The ambiguities proposed for onset-embedded words have motivated accounts of segmentation based on competition between alternative parses of speech into words. In these accounts, the recognition of embedded words is delayed until after their offset when subsequent input rules out longer competitors. In this thesis it is demonstrated that training a simple recurrent network to activate a representation of all the words in a sequence allows a connectionist network to learn the appropriate delay to allow the identification of onset-embedded words without requiring directly implemented competition between words.

Both lexical competition and recurrent network models assume ambiguity between onset-embedded words and equivalent syllables in longer competitors. Acoustic analysis carried out in this thesis confirms the presence of reliable acoustic differences between syllables in short and long words. A series of experiments using gating and cross-modal priming suggest that the perceptual system uses these acoustic differences to discriminate embedded words from the onset of longer competitors and that match or mismatch with longer competitors may be less important for the identification of onset-embedded words. These results are interpreted within a revised version of the recurrent network model, incorporating input representing the acoustic differences between syllables in short and long words.

## Acknowledgements

I am extremely grateful to my supervisors, William Marslen-Wilson and Gareth Gaskell, for their guidance and support throughout the course of this work. I have also benefited from discussions with John Bullinaria, Morten Christiansen, Gary Cottrell, Tom Loucas and Billi Randall on many of the topics contained in this thesis.

At a physical rather than intellectual level, this thesis could not have been written without the assistance of Dr. Whelan, Mr. Cobb and many others at the Middlesex hospital. The care and support of friends and family during my illness helped make this time a frustration rather than a torment.

A further debt of thanks is also owed to my colleagues past and present in the Centre for Speech and Language both at Birkbeck and in Cambridge who provided such a friendly atmosphere in which to work. In particular, thanks must go to Billi Randall and Tom Loucas who helped make my time at CSL so enjoyable. Thanks are also due to Lolly Tyler who faced my divided loyalties in the past year with great stoicism and to Helen Moss and Kathy Rastle for help with proof-reading.

I would like to thank Joe Devlin, Mike Ford and Richard Russell for hospitality and barbecues in Cambridge and Liz Gresham for assisting with matters of style. I also wish to thank my parents for their encouragement throughout my studies. I have kept them from retirement for more years than I should have.

Most important of all though, I would like to thank Maggie Kemmner for advice and encouragement throughout and for the love and support that helped me through the most difficult times.

This research was supported by a studentship from the Engineering and Physical Sciences Research Council.

# Contents

<b>List of tables.....</b>	<b>7</b>
<b>List of figures.....</b>	<b>8</b>
<b>Frontispiece.....</b>	<b>9</b>
<b>1. Introduction.....</b>	<b>10</b>
1.1. Spoken word recognition.....	10
1.2. Lexical segmentation.....	12
1.3. Models of spoken word recognition.....	14
1.3.1. The input representation.....	15
1.3.2. The matching process.....	16
1.3.3. Lexical representations.....	21
1.4. Computational modelling.....	21
1.5. Overview of thesis.....	23
<b>2. Segmentation in lexical access.....</b>	<b>27</b>
2.1. Acoustic cues to segmentation.....	28
2.1.1. Segmental cues to word boundaries.....	28
2.1.2. Duration cues to word boundaries.....	31
2.2. Psycholinguistic accounts of lexical segmentation.....	34
2.2.1. Statistical accounts of segmentation.....	35
2.2.2. Lexical accounts of segmentation.....	45
2.3. Onset-embedded words and segmentation.....	50
2.3.1. Counting embedded words.....	52
2.3.2. Counting morphological embeddings.....	54
2.3.3. Semantic relatedness.....	58
2.3.4. Discussion.....	59
2.4. General discussion.....	60
<b>3. Connectionist models of spoken word recognition.....</b>	<b>62</b>
3.1. The TRACE model.....	63
3.2. Recurrent network accounts.....	68
3.2.1. Prediction tasks and lexical identification.....	70
3.2.2. Modelling word recognition in connected speech.....	72
3.3. Simulation 1 – A distributed account of lexical acquisition .....	76
3.3.1. Method.....	81
3.3.2. Results.....	85
3.3.3. Discussion.....	92
3.4. Simulation 2 – Combining distributional and lexical accounts.....	93
3.4.1. Method.....	94
3.4.2. Results.....	94
3.4.3. Discussion.....	97
3.5. General discussion.....	98

<b>4. Investigating the recognition of embedded words.....</b>	<b>102</b>
4.1. Review of previous experiments.....	104
4.1.1. Gating.....	105
4.1.2. Word-spotting and auditory lexical decision.....	107
4.1.3. Cross-modal priming.....	110
4.2. Experimental design.....	113
4.2.1. Stimuli.....	115
4.2.2. Acoustic analysis and alignment points.....	117
4.3. Experiment 1 – Gating.....	120
4.3.1. Method.....	121
4.3.2. Results and discussion.....	122
<b>5. Cross-modal priming experiments with embedded words.....</b>	<b>127</b>
5.1. Experiment 2a – $AP_1$ .....	128
5.1.1. Method.....	129
5.1.2. Results.....	132
5.1.3. Discussion.....	135
5.2. Experiment 2b-d.....	136
5.2.1. Results of Experiment 2b – $AP_2$ .....	140
5.2.2. Results of Experiment 2c – $AP_3$ .....	142
5.2.3. Results of Experiment 2d – $AP_4$ .....	144
5.2.4. Combined analysis of Experiments 2a-d.....	146
5.3. General discussion.....	148
5.3.1. Acoustic cues to word length.....	149
5.3.2. Lexical competition and short word biases.....	150
<b>6. Acoustic cues to word length in recurrent networks.....</b>	<b>151</b>
6.1. Acoustic cues to word length.....	152
6.1.1. The perception of duration differences.....	152
6.2. Simulation 3 – Processing syllable duration in fast and slow sequences.....	154
6.2.1. Representing duration information in connectionist networks.....	155
6.2.2. Network architecture and training set.....	157
6.2.3. Results.....	159
6.2.4. Discussion.....	163
6.3. Simulation 4 – Representing duration from vowel to vowel.....	164
6.3.1. Network architecture and training set.....	164
6.3.2. Results and discussion.....	166
6.4. Simulating experimental data.....	168
6.4.1 Method.....	169
6.4.2. Experiment 2a – $AP_1$ .....	171
6.4.3. Experiment 2b – $AP_2$ .....	173
6.4.4. Experiment 2c – $AP_3$ .....	174
6.4.5. Experiment 2d – $AP_4$ .....	175
6.4.6. Discussion.....	177
6.5. General discussion.....	179
<b>7. Effects of following context in recognising embedded words.....</b>	<b>181</b>

7.1. Experimental materials.....	182
7.2. Experiment 3 – Gating.....	184
7.2.1. Results and discussion.....	184
7.3. Experiment 4 - Cross-modal priming.....	187
7.3.1. Results.....	188
7.4. Comparison of Experiment 2 and Experiment 4.....	191
7.5. Comparing experimental data and recurrent network simulations.....	194
7.5.1. Recognising embedded words.....	195
7.5.2. Ruling out longer competitors.....	197
7.6. General discussion.....	198
<b>8. Concluding remarks.....</b>	<b>203</b>
8.1 Lexical representation in a distributed system.....	203
8.2 Lexical segmentation and identification.....	205
8.3 Acoustic cues to word boundaries.....	207
8.4 Sequential recognition, lexical competition and embedded words.....	208
8.5 Summary and future directions.....	209
<b>Appendix A – Stimulus sentences.....</b>	<b>211</b>
<b>Appendix B – Supplementary analyses.....</b>	<b>219</b>
B.1. Experiment 2b.....	219
B.2. Experiment 2c.....	219
B.3. Experiment 2d.....	220
B.4. Experiment 4a.....	220
B.5. Experiment 4b.....	221
<b>References.....</b>	<b>222</b>

## List of tables

Table 2.1	Proportion of onset-embedded words in LDOCE and CELEX.....	55
Table 2.2	Proportion of onset-embedded words excluding morphology.....	57
Table 3.1	Phonetic feature representation used for computational simulations.....	82
Table 3.2	Vocabulary items used for computational simulations.....	83
Table 4.1	Alignment points and acoustic properties of stimuli in Experiment 1....	118
Table 4.2	Confidence ratings for short words at initial gates in Experiment 1.....	125
Table 5.1	Prime and target stimuli for Experiment 2a.....	130
Table 5.2	Mean response times and error rates for Experiment 2a.....	132
Table 5.3	Prime and target stimuli for Experiments 2b-d.....	138
Table 5.4	Mean response times and error rates for Experiment 2b.....	140
Table 5.5	Mean response times and error rates for Experiment 2c.....	142
Table 5.6	Mean response times and error rates for Experiment 2d.....	145
Table 6.1	Network representation of syllable duration in Simulation 3.....	156
Table 7.1	Mean response times and error rates for Experiments 4a-b.....	189

## List of figures

Figure 1.1 Processing stages in auditory lexical access.....	14
Figure 2.1 Inhibitory connections in Shortlist.....	49
Figure 3.1 A simple recurrent network .....	70
Figure 3.2 A snapshot of the SRN trained in Simulation 1.....	84
Figure 3.3 Activation of lexical units in an example sequence for Simulation 1....	86
Figure 3.4 Cohort competitors and onset-embedded words in Simulation 1.....	88
Figure 3.5 Long word sequences and lexical garden-paths in Simulation 1.....	90
Figure 3.6 Offset-embedded words in Simulation 1.....	91
Figure 3.7 A snapshot of the SRN trained in Simulation 2.....	94
Figure 3.8 Comparison of networks with and without the prediction task.....	96
Figure 4.1 Speech waveforms and alignment points for stimuli in Experiment 1....	119
Figure 4.2 Proportion of responses matching the target words in Experiment 1.....	123
Figure 5.1 Magnitude and significance of priming in Experiment 2a.....	134
Figure 5.2 Magnitude and significance of priming in Experiment 2b.....	141
Figure 5.3 Magnitude and significance of priming in Experiment 2c.....	143
Figure 5.4 Magnitude and significance of priming in Experiment 2d.....	146
Figure 5.5 Normalised magnitude of priming in Experiments 2a-d.....	147
Figure 6.1 Histogram of syllable duration in short and long test items.....	153
Figure 6.2 A snapshot of the SRN trained in Simulation 3.....	158
Figure 6.3 Onset-embedded words with and w-out prior context in Simulation 3...	160
Figure 6.4 Onset-embedded words in ambiguous prior contexts in Simulation 3....	162
Figure 6.5 Onset-embedded words in Simulation 4.....	167
Figure 6.6 Network activation and priming results at $AP_1$ .....	172
Figure 6.7 Network activation and priming results at $AP_2$ .....	173
Figure 6.8 Network activation and priming results at $AP_3$ .....	175
Figure 6.9 Network activation and priming results at $AP_4$ .....	176
Figure 7.1 Proportion of responses matching the target words in Experiment 3.....	185
Figure 7.2 Magnitude and significance of priming in Experiments 4a-b.....	190
Figure 7.3 Normalised magnitude of priming in Experiments 2 and 4.....	192
Figure 7.4 Network activation and priming results for short words in garden-path and non garden-path contexts.....	196
Figure 7.5 Network activation and priming results for long words in garden-path and non garden-path contexts.....	197



