1. Introduction

One of the most fundamental of human skills is the perception of connected speech. The communication abilities provided by spoken language are the most obvious division between humans and other animals. This thesis investigates one small aspect of the cognitive skills that contribute to our ability to understand speech – the segmentation and recognition of words in connected speech.

1.1. Spoken word recognition

Word recognition plays a central role in the processes by which acoustic waveforms are converted into a representation of the meaning of utterances. Accounts of spoken language comprehension typically postulate initial processing stages which extract relevant perceptual information from the acoustic signal. The term *word recognition* applies to the processes by which these perceptually-derived input representations make contact with stored representations of the words being identified. Processing stages following word recognition are then concerned with integrating the individual syntactic and semantic properties of the recognised words into a representation of an utterance's meaning.

In this very abstract description, the processes associated with spoken word recognition appear little different from those involved in the comprehension of printed text. Indeed, early models of spoken word recognition were derived from existing knowledge of visual word recognition (Forster, 1976; Morton, 1969). This debt remains apparent in some more recent accounts of spoken language processing (see for instance, Bradley & Forster, 1987; Luce, Pisoni, & Goldinger, 1990). However, with increasing competence in the experimental manipulation of speech stimuli (mostly through the use of digital computers for the recording and playback of speech) many of the unique properties of spoken language have become available to psycholinguistic investigation.

Perhaps the most obvious difference between spoken and written language is that while written language can be perceived in parallel for as long as is required for processing (at least at the level of individual words) spoken language is sequentially ordered and transient. Since only a small amount of speech can be retained in the auditory system in an unanalysed or echoic form (see for instance Crowder & Morton, 1969; Huggins, 1975) it seems that the processing infrastructure for speech perception must operate rapidly if connected speech is to be processed efficiently. The question of how closely the processing of connected speech tracks the auditory input is an important issue in research on speech perception (Mattys, 1997). However, the immediacy with which we perceive and are able to respond to incoming speech (see for instance Marslen-Wilson & Tyler, 1980) suggests that fast and effective processing of spoken language is a normal property of adult comprehension.

Another important problem for spoken language comprehension that is absent in reading printed words is created by inherent variability in the speech signal. The development of written language and use of the printing press required explicit agreement about the exact form of written letters (though consider the difference between **a**, ***a*** and **A**). The evolution of language and the development of the speech articulators in humans allowed no equivalent standardisation (though Stevens & Blumstein (1981) have described potentially invariant cues to the identification of some classes of phonemic segment). For these reasons, variation in the acoustic form of the speech signal is pervasive at many levels of analysis.

One source of variation in the acoustic form of the speech signal is caused by differences between speakers. These may be caused by differences in accent or dialect, or through differences between individual speakers in the rate and pitch of their speech. Experimental evidence suggests that familiarity with the particular characteristics of individuals' speech can facilitate identification of words even with several days delay between successive experiences of a given speaker uttering a particular word (Goldinger, 1996a).

Another form of variation is caused by pressure to fit the discrete phonological gestures associated with segments into a connected stream of speech. The resulting coarticulation and deletion of segments, where speech sounds are altered through the influence of preceding and following phonemes, may result in dramatic differences in the production of individual words. The word *stand* for instance will rarely be pronounced in its canonical form /stænd/ but may surface as /stæn/ in *stand down*, as /stæŋ/ in *stand close* or as /stæm/ in *stand back* (for experimental and computational investigations into the processing of phonological variation, see Gaskell, Hare, & Marslen-Wilson, 1995; Gaskell & Marslen-Wilson, 1996).

A third class of problem specific to spoken as distinct from written language is that of segmentation. Alphabetic writing systems are composed of discrete units (letters) formed into larger chunks (words) which are then organised into longer sequences (sentences and paragraphs). There is therefore a physical separation of orthographic units at multiple levels which is likely to provide at least an initial structure for the mental representation of written language. However, this physical organisation is not found in the structure of spoken language, which is to a great extent continuous and formed out of connected and co-articulated units not bounded by spaces or other breaks in the speech signal. Consequently the processes by which speech is divided into low level units, as well as the nature of these units, remains open to widespread debate. Different authors have proposed that perceptual processing of speech proceeds from spectral representations (Klatt, 1979), phonetic features (Warren & Marslen-Wilson, 1987; 1988), phonemes (Pisoni & Luce, 1987) or syllables (Mehler, Dommergues, Frauenfelder, & Segui, 1981).

A similar debate exists regarding the representation of higher-level lexical units in spoken word recognition. Most authors have assumed that the units of lexical storage correspond to an orthographic word as written on the page. However, research on the representation and processing of derivational morphology has been used to argue for a decomposed mental lexicon in which representations of stems and affixes combine to represent morphologically complex words (Marslen-Wilson, 1999; Marslen-Wilson, Tyler, Waksler, & Older, 1994). Conversely, experimental evidence demonstrating a ‘word superiority effect’ for familiar word combinations (e.g. *greasy spoon*) has been used as evidence for stored representations of larger units consisting of more than one orthographic word (Harris, 1994; Harris, 1996). This conflict suggests that the lexicon is unlikely to contain one single size of lexical unit. Consequently questions about the computational mechanisms that divide the speech stream into lexical units are likely to be essential in describing the operation of the language processor.

1.2. Lexical segmentation

The segmentation of connected speech into lexical units is a major topic for psycholinguistic research and is the primary focus of the research reported in this thesis.

The use of the term ‘lexical segmentation’ throughout this thesis is not intended to imply that only processes at a lexical level are involved in segmentation; rather this term refers to the segmentation of meaningful units, as distinct from the segmentation of lower-level phonetic or phonemic units in the speech stream. The term ‘segmentation’, used in this thesis without a modifier, refers to the segmentation of connected speech into lexical units. These units may or may not correspond to orthographic or dictionary words depending on the details of the particular theory being discussed at the time.

Various theories have been proposed describing the nature of the information used to divide the speech stream into lexical units. There have been three main classes of proposal. Firstly specific acoustic markers in the speech stream are used in segmentation (Lehiste, 1972; Nakatani & Dukes, 1977; Nakatani & Schaffer, 1978). Secondly, knowledge of the statistical or distributional structure of lexical items in the language can provide a cue to segmentation; this statistical approach may be applied in different domains, including phonology (Brent & Cartwright, 1996; Cairns, Shillcock, Chater, & Levy, 1997); metrical stress (Cutler & Norris, 1988; Grosjean & Gee, 1987) or prosody (Christophe, Guasti, Nespor, Dupoux and Ooyen, 1997). Thirdly, segmentation is achieved through the identification of lexical items in connected speech (Marslen-Wilson & Welsh, 1978; McClelland & Elman, 1986; Norris, 1994).

While there is considerable experimental evidence showing that each of these strategies is effective in lexical segmentation, very little research has attempted to provide a unified account of the effect that different combinations of these strategies have on the segmentation of words in connected speech (for exceptions see Christiansen, Allen & Seidenberg, 1998; Norris, McQueen, Cutler & Butterfield, 1997; Norris, McQueen & Cutler, 1995). One aim of this thesis is to attempt to evaluate these different accounts of segmentation by investigating which mechanisms operate in the case of words whose boundaries are predicted to be ambiguous by one or more of these theories. The words concerned are embedded at the onset of longer words (such as *cap* in *captain*). This thesis uses sentences containing these embedded words to investigate how processes using different sources of information may interact during the segmentation and identification of words in connected speech.

1.3. Models of spoken word recognition

Although the work reported in this thesis is intended to address issues of lexical segmentation and word recognition relevant to all current theories, this research relies on a general theoretical framework. This section describes the main assumptions behind this framework and reviews the experimental results that have been used to support these assumptions.

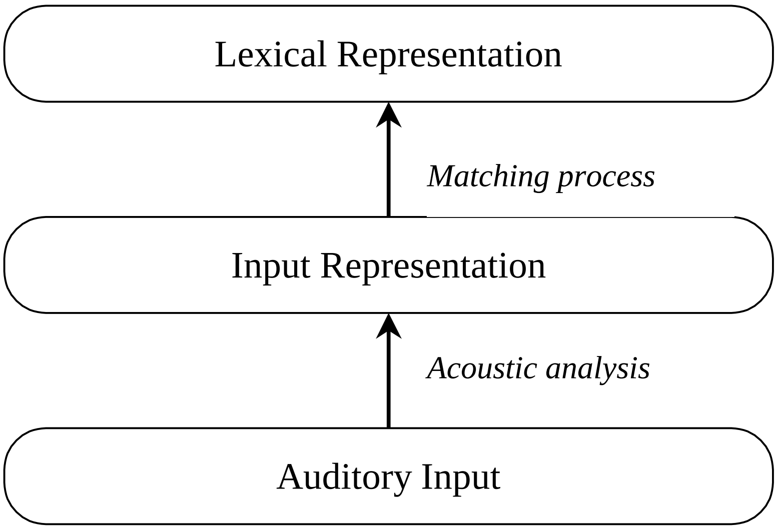

Figure 1.1: Processing stages involved in auditory lexical access

The general approach to speech recognition that is taken in the literature involves progressive stages of abstraction, going from a representation of the acoustic signal, to low-level linguistic or perceptual units, to lexically based representations and ultimately to syntactic and semantic properties of the spoken input. In descriptions of this framework, a commonly made distinction is between processes that are involved in deriving an initial representation of the speech input, and later stages involved in accessing lexical representations that match the input representation (see for instance Tyler & Frauenfelder, 1987). In the initial stages, *acoustic analysis* is carried out to construct a form-based representation of the speech signal. Later stages of analysis then involve a *matching process* whereby this input representation is matched to a representation of specific lexical candidates. The goal of this lexical access process is to identify or recognise specific lexical items contained in the speech signal. These processing stages are illustrated in Figure 1.1.

Despite general agreement regarding this processing framework for speech perception, there is still considerable debate and disagreement regarding the nature of the representations that are involved in each of these stages, as well as discussion of how different levels of representation interact during processing. One important issue regards the nature of the pre-lexical representation of the speech input that contacts the mental lexicon.

1.3.1. The input representation

Psycholinguistics has inherited from the linguistic tradition an assumption that a string of undifferentiated phonemes (represented as bundles of phonetic features) is an appropriate representation of the speech stream. For a historical perspective on this assumption see Chomsky and Halle (1968). For a more recent discussion of the limitations of this approach from a phonetic perspective see Manaster-Ramer (1996) and Port (1996). Some accounts of spoken word recognition, such as the Neighbourhood Activation Model of Luce and colleagues (Luce et al., 1990; Pisoni & Luce, 1987) assume a phonemically-labelled level of representation as an input to the lexical identification system. Similarly, computational models such as TRACE (McClelland & Elman, 1986) and Shortlist (Norris, 1994) incorporate a phonemically coded representation at a pre-lexical level - although these may be preceded by other lower-level representations of the speech input.

Alternatively, some authors have proposed that the primary input representation of the speech signal is in terms of larger, syllabic units. Evidence supporting this position has come from experiments in French showing that the detection of word fragments is facilitated in words that contain these fragments as syllables compared to words in which these fragments cross a syllable boundary (Mehler et al., 1981). Although this finding has been replicated in other languages such as Spanish, it appears that languages such as Japanese and English are an exception to this pattern. Input representations of Japanese speakers appear to be based around a smaller, moraic unit (Otake, Hatano, Cutler, & Mehler, 1993) while research in English has failed to provide unequivocal evidence of pre-lexical representations organised at any single level (Dupoux & Hammond (submitted), see Pallier, Christophe, & Mehler (1998) for a recent review of this research).

Other accounts of lexical access and speech perception propose a sub-phonemic representation as input to the lexical access process. For instance, Warren and Marslen-

Wilson (1987, 1988) postulate a featural representation of the speech input, with fine-grained sub-phonemic information playing an important role in the recognition process. Experiments showing that mismatching information at a sub-phonemic level can disrupt the recognition process have been cited as evidence to support these accounts (Andruski, Blumstein, & Burton, 1994; Marslen-Wilson, Moss, & van Halen, 1996; Marslen-Wilson & Warren, 1994).

This conflict between accounts proposing different sizes of units as being of primary importance in spoken word processing may reflect a distinction between experimental tasks that tap into perceptual awareness of the form of speech and tasks that infer the structure of pre-lexical representations from the properties of the recognition process. Some theoretical accounts suggest a functional separation between representations involved in the pre-lexical processing of the speech signal and representations that are involved in producing a 'percept' of the speech input (Marslen-Wilson & Warren, 1994). Recent computational models (Gaskell & Marslen-Wilson, 1997; Norris, McQueen & Cutler, in press) consequently provide separate levels of representation for pre-lexical acoustic processing and for the perceptual representations used in phoneme detection and other similar tasks. This separation between acoustic processing and perceptual representation suggests that the units involved in forming a speech percept may not be the same units by which the acoustic signal is processed pre-lexically.

1.3.2. The matching process

Many different models have been proposed as accounts of the process by which input representations are mapped onto representations of lexical form. Rather than listing the models themselves, this section will illustrate the theoretical distinctions made by these models. Describing these distinctions is the most productive means of categorising the various models described in the literature.

Parallel vs. serial search

One important distinction made in the literature is between models based on a serial comparison of the input representation with successive lexical items (as in the search model of Forster, 1976; Bradley & Forster, 1987; Forster, 1989) and those based on the parallel comparison of multiple candidates (as originally proposed for visual word recognition in the logogen model of Morton (1969) and extended to spoken word recognition by Marslen-Wilson & Welsh (1978)).

In the serial search model, lexical access occurs via a search through a frequency-ordered list of word candidates. Each candidate is compared in turn to the currently perceived input, with lexical access occurring when a match is found between a lexical candidate and the current input. This ordered search provides a very natural account of effects of word frequency, whereby highly frequent words are accessed more quickly than low frequency words (Rubenstein, Garfield, & Millikan, 1970). However, since this effect is less reliable with spoken than written words (Bradley & Forster, 1987; Marslen-Wilson, 1984; though see Marslen-Wilson, 1990) it is unclear whether frequency effects can be used as evidence for serial search models of spoken word recognition.

In contrast, parallel access models propose that multiple lexical candidates are compared simultaneously, with the activation of any given item indicating the current degree of fit of a lexical hypothesis (cf. Selfridge, 1959). By these accounts, identification involves accumulating sufficient evidence to activate a single lexical item past some threshold level. As a consequence, lexical access is no longer an all-or-nothing process, but may initially involve the simultaneous, partial access of multiple candidates before a single lexical item is recognised. Parallel activation models can also account for frequency effects, either through postulating differences in the resting activation of the representational units for words with different frequencies (McClelland & Rumelhart, 1981; Morton, 1969) or through stronger connections to units involved in the representation of higher frequency words (Plunkett & Marchman, 1991; Seidenberg & McClelland, 1989).

One advantage of parallel access over serial search models is that they offer a simple account of the activation of multiple lexical candidates during spoken word recognition. For instance, Zwitserlood (1989) showed that during the auditory presentation of a fragment of the Dutch word *kapitein* (captain), significant facilitation could be observed for words that are semantically and associatively related to a competing word *kapitaal* (capital) that also matched the fragment. This transient activation of multiple lexical items (and associated meanings) is readily accommodated in models proposing that the speech input activates all the candidates that match the initial spoken input, with selection processes operating to narrow down the set of activated candidates to those that continue to match the speech input (Cohort model – Marslen-Wilson & Welsh, 1978; TRACE – McClelland & Elman, 1986; Shortlist – Norris, 1994).

Although these results may be simulated in a serial system by initiating searches at regular intervals during the speech stream and accessing the meaning of matching candidates after each search, this simulation of parallel effects is only achieved through reducing the role of serial search to a minimum, and making the recognition process functionally equivalent to a parallel access account. The research carried out in this thesis will be based around a parallel-access, activation-based account of the word recognition process.

A further implication of the results presented by Zwitserlood (1989), as well as other results used to motivate the Cohort model (Marslen-Wilson & Welsh, 1978) is that the spoken word processing system tracks the speech input, continuously updating the activation of different representations in the light of new input. Consequently these data are beyond the scope of spoken word recognition models which do not make explicit predictions for processes that occur during the time-course of the speech signal, such as the Neighbourhood Activation Model of Luce and colleagues (Luce et al., 1990). This, and other models that do not adequately capture the temporal processing of the speech input will not be considered in detail in this thesis.

Autonomy, interaction and competition

Models using an activation metaphor to represent the goodness of fit between currently available evidence and lexical items provide a coherent account of the matching process. However, it remains unclear what sources of evidence are evaluated in this way during lexical access. One commonly made distinction in the literature is between accounts whereby only information from lower levels affects the activation of lexical candidates – *autonomous* models, such as those proposed by Forster (1976) and Norris (1994) – and interactive models in which information from higher levels (such as from syntactic or semantic constraint) can also influence the recognition system (such as in TRACE - McClelland & Elman, 1986).

In a parallel activation account, the terms autonomous and interactive can be interpreted as constraints on the direction of connectivity and types of connections that can be made to and from lexical units. More specifically, an autonomous model is one in which lexical units only receive input from units at lower levels. An interactive model is one in which these lexical units may also receive feedback from higher levels of representation such as syntax or semantics.

However, with the advent of parallel distributed processing models in which connection strengths are acquired through the application of a gradient descent learning algorithm (Rumelhart, Hinton & Williams, 1986), this categorical distinction between autonomous and interactive models may not provide the best characterisation of different styles of computation. For instance, in many computational simulations the notion of a distinctly lexical level of representation becomes blurred. For example in the triangle framework of Plaut and colleagues (Plaut, McClelland, Seidenberg, & Patterson, 1996) distributed representations of orthography, phonology and semantics form a lexical system through the connectivity that exists between these levels. The nature of the processing interactions that develop between different representations depends on the regularities that exist between different domains and only indirectly on the assumptions made by the modeller (see for instance Harm & Seidenberg, 1999).

These terms are similarly problematic when used in interpreting the results of behavioural experiments. For instance, the Ganong effect whereby ambiguous phonemes are categorised according to the lexical status of the string in which they are contained has been interpreted as evidence for interactive effects on phonetic categorisation (Ganong, 1980). Although this influence of lexical information on phonetic categorisation may be simulated through top-down interactions between lexical and pre-lexical information (in TRACE, for example), these results need not imply top-down connectivity but only that participants’ responses are made from a level of representation that includes lexical influences. Elman and McClelland (1988) showed that phonemes disambiguated by lexical context influence categorisation of subsequent phonemes in the same way as would be expected for unambiguous phonemes. Thus the Ganong effect can also affect mechanisms involved in compensation for co-articulation. Although their results were initially interpreted as evidence for top-down, lexical influences on phonemic processing, subsequent experiments (Pitt & McQueen, 1998) and simulations (Cairns, Shillcock, Chater, & Levy, 1995; Norris, 1993) show that transitional probabilities between phonemes provide a non-lexical account for this effect. Consequently, in order to describe a particular pattern of experimental results as indicating top-down influences or interactive processes it must be demonstrated that no alternative, non-interactive account of the results must be possible.

Another term commonly used to describe models of speech perception is *competition*. This term typically describes models in which the activation of any given lexical candidate is affected by the presence or absence of other activated candidates. For example, if two or more lexical candidates match the current input, a model incorporating competition will activate each candidate to a lesser degree than if only a single candidate matches the input. In localist connectionist models, the presence of competition is commonly modelled as inhibitory connections between jointly activated units (e.g. *captain* and *captive* which would both be activated by the speech input /kæptɪ/)

However, just as caution is advised in the use of the words ‘autonomous’ and ‘interactive’, similar caution should be exercised in using the word ‘competition’. Behavioural data described as indicating competition need not be simulated using direct inhibitory connections and consequently may not only support computational models that include direct competition between lexical units. For instance, recurrent neural networks trained by back-propagation produce output activations that are dependent on the number of simultaneously-activated candidates by representing the probability of all outputs given the current input. Such behaviour would commonly be assumed to reflect competition between output units, yet these recurrent networks do not incorporate direct inhibitory connections between units within a processing level (e.g. Gaskell & Marslen-Wilson, 1997; in press).

It would therefore appear that terms such as autonomy, interaction and competition that are used to relate behavioural data to theoretical accounts are best used sparingly in the absence of implemented computational systems based on these theories. Processing distinctions made on behavioural grounds may not be as effective in distinguishing between different models as previously considered. Throughout this thesis care will therefore be taken to describe precisely the behavioural evidence from experiments using terms that are independent of their simulation in computational models.

1.3.3. Lexical representations

Traditional assumptions about the role of lexical representations viewed these as a bridge between the form of words and their meaning. Consequently, accounts of lexical access in spoken word recognition considered the goal of the process to be to activate a representation of the meaning of lexical items or words in the speech stream. This focus on access to the meanings of words is apparent in work investigating effects of preceding sentential context on the meanings activated in response to words such as *bank* which have multiple meanings (Simpson, 1984; Swinney, Onifer, Prather & Hirshkowitz, 1979).

Recent accounts of sentence processing have been proposed that draw parallels between the resolution of lexical ambiguities (such as those created by homophones) with syntactic ambiguities such as "*the spy saw the cop with the binoculars*" (MacDonald, Perlmutter, & Seidenberg, 1994). The debate between these statistical (constraint-satisfaction) and syntactic (garden-path) accounts is still far from resolved (Frazier & Clifton, 1996). However, both classes of accounts propose that lexically represented information plays an important role in guiding the parsing process. Consequently, lexical access not only activates representations of semantic and conceptual knowledge, but also accesses information about the syntactic constructions in which a word is used.

1.4. Computational modelling

As was illustrated in the preceding discussion, the implementation of computational models is vital for determining whether or not a descriptive model is capable of accounting for a particular set of experimental data. Consequently this thesis combines empirical investigations of spoken word recognition with computational simulations of the time course of identification of words in connected speech. In order to implement a computational model it is necessary to specify every component and assumption that is associated with a theory. By making theories explicit in a working system, insights can be gained into the nature of the tasks undertaken by the language processor. It is also possible to test whether current experimental data can be accounted for by models derived from these theories.

Computational accounts of lexical segmentation and spoken word recognition come in many forms. One important distinction is between models that are implemented using symbolic algorithms (INCDROP - Brent, 1997; PARSER - Perruchet & Vinter, 1998 and MK10 - Wolff, 1977) and connectionist models implemented using networks of simple, neuron-like processing units (Distributed Cohort Model - Gaskell & Marslen-Wilson, 1997; TRACE - McClelland & Elman, 1986; and Shortlist - Norris, 1994). Both symbolic and connectionist models provide numerous advantages over conventional, descriptive theorising, but there are also many differences between them, perhaps the most important of which is in the style of processing they assume.

Conventional computer languages impose strict constraints on the manner and order in which operations can occur on input representations. Computation occurs through a series of discrete, serial steps involving transformations and calculations operating on abstract, symbolic representations. In contrast, connectionist models rely on a parallel process mapping from one representation to another through the operation of simple, distributed processing elements (Rumelhart & McClelland, 1986a; McClelland & Rumelhart, 1986). Although each style of computation could be considered theoretically neutral (after all, both types of system are Turing equivalent and are implemented on serial computers), in practice the two types of model lend themselves most naturally to different types of process and different types of explanation. Connectionist models view cognition as involving similarity-based processes of generalisation, while symbolic models operate through mechanisms of categorisation and rule application.

This distinction between symbolic and connectionist computation is most clearly apparent in the debate on supposed differences in the representation and processing of regular and irregular forms of the English past tense. Symbolic computational systems require separate processes for regular and irregular verbs (Pinker & Prince, 1988; Prasada & Pinker, 1993) whereas connectionist accounts propose that both types of verbs are processed within a single system (Rumelhart & McClelland, 1986b; Plunkett & Marchman, 1991; Plunkett & Marchman, 1993). Although this debate is too lengthy to summarise here, it is becoming increasingly apparent that it will not be resolved either by behavioural data (Ullman et al., 1997) or by computational simulations alone (Joanisse & Seidenberg, 1999). Converging evidence from multiple sources is therefore required to constrain theorising. This multi-disciplinary combination of computational modelling and experimental investigation is utilised in the research reported in this thesis.

The computational simulations carried out here have used connectionist networks exclusively. Although there is nothing intrinsic to these simulations that precludes the use of symbolic algorithms, a list of the desirable computational properties of models of spoken word recognition suggest that connectionist architectures are more parsimonious. For instance, these models are required to account for effects of partial information, to display probabilistic operation, and to be robust in the face of noisy input. Furthermore an important property of the word recognition system is that it be capable of acquiring new vocabulary throughout childhood and to retain this knowledge in the face of progressive damage with ageing. Many of these desirable properties follow naturally from the use of a connectionist model.

One further debate that is mentioned here only in passing, is the distinction between connectionist models built exclusively using localist representations and those that use distributed representations as well. Despite the prevalence of arguments in favour of distributed systems in the psychological literature, it is clear that many useful computational properties can be gained by incorporating localist representations (Page, in press), though possibly at the cost of a certain amount of neural plausibility. Although the models reported in this thesis use a localist representation of phonetic features at the input level and a localist lexical representation at the output, they develop distributed internal representations through training. As will be discussed in Chapter 3, this use of localist input and output representations in a system trained to produce distributed internal representations is chosen for convenience of implementation only, and not through any ideological commitment to either localist or distributed models.

1.5. Overview of thesis

This review of the theoretical background provides a framework with which to describe the work carried out as part of this thesis. Since the research approaches segmentation from both an experimental and computational perspective, the review of lexical segmentation in Chapter 2 provides a theoretical introduction without including excessively detailed accounts of either computational or experimental investigations. These will be reserved for more focused reviews at the start of the relevant chapters: Chapter 3 for computational modelling of lexical segmentation and identification and Chapter 4 for experimental investigations of segmentation and lexical access.

The review of lexical segmentation in Chapter 2 outlines an apparent conflict between processes of segmentation that have been hypothesised to operate at different levels of representation of the speech stream. It is argued that onset-embedded words provide a test-case for resolving this conflict since they are predicted to be temporarily ambiguous by several accounts of lexical segmentation. The main body of the thesis will focus on investigating the recognition of onset-embedded words with a view to resolving the conflict between acoustic and lexical accounts of segmentation. Chapter 2 concludes by reporting dictionary searches establishing the extent of the problems created by embedded words within the morphologically decomposed lexicon proposed by Marslen-Wilson, Tyler, Waksler and Older (1994).

Chapter 3 starts with an introductory review of computational models of lexical identification. This review focuses on a particular problem for models of spoken word recognition regarding the identification of onset-embedded words. The apparent need to delay identification of these words until after their acoustic offset has been used to motivate models of spoken word recognition that incorporate inhibitory competition between lexical units. However, simulations reported in this chapter describe how a simple recurrent network can learn to identify embedded words without explicitly implemented competition. The relationship between this model and accounts of vocabulary acquisition is discussed, in particular looking at the relationship between processes involved in learning the statistical structure of speech and processes involved in extracting meaning from sequences of sounds.

Since the recurrent network models reported in Chapter 3 predict a different time course of identification for embedded words than previous, competition-based models, Chapter 4 begins by describing the results of previous experiments that might decide between models with inhibitory competition between lexical units and the recurrent networks described in Chapter 3. Given the lack of appropriately constructed experiments in the literature, Chapter 4 then describes the development of stimuli for experiments with the potential to test these two conflicting accounts of the identification of onset-embedded words. Since computational models of lexical segmentation make strong and possibly unjustified assumptions about the nature of the speech input, the experimental stimuli developed for this experiment are subjected to detailed acoustic analyses to ensure that investigations are based on an accurate description of the acoustic properties of the speech stream. The importance of this analysis is confirmed by the results of a gating study suggesting that acoustic differences between embedded words and longer competitors can be used by the perceptual system. However, caveats regarding the role of response biases in the gating task limit the interpretations of these results with respect to the computational models described previously.

Chapter 5 reports the results of a series of cross-modal priming experiments carried out on the stimuli described in Chapter 4. Using the magnitude of repetition priming as a measure of lexical activation it is shown that acoustic cues allow the perceptual system to differentiate onset-embedded words from longer competitors even before the offset of the embedded word. Despite these acoustic cues, priming experiments tracking the activation of competing interpretations show that longer competitors of onset-embedded words remain active in contexts designed to produce ambiguity for these words.

Since there is now experimental evidence supporting the role of acoustic cues to word length in the identification of onset-embedded words, Chapter 6 describes modifications to the previously developed recurrent network model to incorporate input cues analogous to those hypothesised to be responsible for the discrimination of onset-embedded words. Two sets of simulations are reported, exploring the effect of input cues that require adaptive processing of the preceding spoken context in order to be utilised effectively. The results of these simulations, within the limits created by the small scale of the network, show a similar time course of identification for onset-embedded words and longer competitors as the priming experiments reported in Chapter 5.

In Chapter 7 a further prediction of the model regarding the role of following context in the identification of onset-embedded words is tested. The models reported in Chapter 6 predict that information after the offset of an embedded word plays an important role in ruling out longer competitors, thereby supporting the recognition of an embedded word. One further gating experiment and two cross-modal priming experiments tested this prediction using a set of stimuli derived from those that were used in the initial series of experiments. Results of these experiments suggest that the activation of onset-embedded words appears to be unaffected by the presence or absence of phonological mismatch with the longer words in which they are embedded. Comparisons are made between this behavioural profile and the predictions of the recurrent network account of word recognition presented in the previous chapters. Finally, in Chapter 8, conclusions are drawn from the experimental and computational work presented in this thesis. Future work is also proposed to extend and test the model developed in this thesis.