

# LEXICAL ACCESS IN CONNECTED SPEECH RECOGNITION

Ted Briscoe

Computer Laboratory  
University of Cambridge  
Cambridge, CB2 3QG, UK.

## ABSTRACT

This paper addresses two issues concerning lexical access in connected speech recognition: 1) the nature of the pre-lexical representation used to initiate lexical look-up 2) the points at which lexical look-up is triggered off this representation. The results of an experiment are reported which was designed to evaluate a number of access strategies proposed in the literature in conjunction with several plausible pre-lexical representations of the speech input. The experiment also extends previous work by utilising a dictionary database containing a realistic rather than illustrative English vocabulary.

## THEORETICAL BACKGROUND

In most recent work on the process of word recognition during comprehension of connected speech (either by human or machine) a distinction is made between lexical access and word recognition (e.g. Marslen-Wilson & Welsh, 1978; Klatt, 1979). Lexical access is the process by which contact is made with the lexicon on the basis of an initial acoustic-phonetic or phonological representation of some portion of the speech input. The result of lexical access is a *cohort* of potential word candidates which are compatible with this initial analysis. (The term *cohort* is used descriptively in this paper and does not represent any commitment to the particular account of lexical access and word recognition provided by any version of the *cohort theory* (e.g. Marslen-Wilson, 1987).) Most theories assume that the candidates in this cohort are successively whittled down both on the basis of further acoustic-phonetic or phonological information, as more of the speech input becomes available, and on the basis of the candidates' compatibility with the linguistic and extralinguistic context of utterance. When only one candidate remains, word recognition is said to have taken place.

Most psycholinguistic work in this area has focussed on the process of word recognition *after* a cohort of candidates has been selected, emphasising the role of further lexical or 'higher-level' linguistic constraints such as word frequency, lexical semantic relations, or syntactic and semantic congruity of candidates with the linguistic context (e.g. Bradley & Forster, 1987; Marslen-Wilson & Welsh, 1978). The few explicit and well-developed models of lexical access and word recognition in continuous speech (e.g. TRACE, McClelland & Elman, 1986) have small and unrealistic lexicons of, at most, a few hundred words and ignore phonological processes which occur in fluent speech. Therefore, they tend to overestimate the amount and reliability of acoustic information which can be directly extracted from the speech signal (either by human or machine) and

make unrealistic and overly-optimistic assumptions concerning the size and diversity of candidates in a typical cohort. This, in turn, casts doubt on the real efficacy of the putative mechanisms which are intended to select the correct word from the cohort.

The bulk of engineering systems for speech recognition have finessed the issues of lexical access and word recognition by attempting to map directly from the acoustic signal to candidate words by pairing words with acoustic representations of the canonical pronunciation of the word in the lexicon and employing pattern-matching, best-fit techniques to select the most likely candidate (e.g. Sakoe & Chiba, 1971). However, these techniques have only proved effective for isolated word recognition of small vocabularies with the system trained to an individual speaker, as, for example, Zue & Huttenlocher (1983) argue. Furthermore, any direct access model of this type which does not incorporate a pre-lexical symbolic representation of the input will have difficulty capturing many rule-governed phonological processes which affect the pronunciation of words in fluent speech, since these processes can only be characterised adequately in terms of operations on a symbolic, phonological representation of the speech input (e.g. Church, 1987; Frazier, 1987; Wiese, 1986).

The research reported here forms part of an ongoing programme to develop a computationally explicit account of lexical access and word recognition in connected speech, which is at least informed by experimental results concerning the psychological processes and mechanisms which underlie this task. To guide research, we make use of a substantial lexical database of English derived from machine-readable versions of the *Longman Dictionary of Contemporary English* (see Boguraev et al., 1987; Boguraev & Briscoe, 1989) and of the Medical Research Council's psycholinguistic database (Wilson, 1988), which incorporates word frequency information. This specialised database system provides flexible and powerful querying facilities into a database of approximately 30,000 English word forms (with 60,000 separate entries). The querying facilities can be used to explore the lexical structure of English and simulate different approaches to lexical access and word recognition. Previous work in this area has often relied on small illustrative lexicons which tends to lead to overestimation of the effectiveness of various approaches.

There are two broad questions to ask concerning the process of lexical access. Firstly, what is the nature of the initial representation which makes contact with the lexicon? Secondly, at what points during the (continuous) analysis of the speech signal is lexical look-up triggered?

We can illustrate the import of these questions by considering an example like (1) (modified from Klatt via Church, 1987).

(1)

- a) Did you hit it to Tom?
- b) [dIjEhIdl?tEtam]

(Where 'I' represents a high, front vowel, 'E' schwa, 'd' a flapped or neutralised stop, and '?' a glottal stop.) The phonetic transcription of one possible utterance of (1a) in (1b) demonstrates some of the problems involved in any 'direct' mapping from the speech input to lexical entries not mediated by the application of phonological rules. For example, the palatalisation of final /d/ before /y/ in /did/ means that any attempt to relate that portion of the speech input to the lexical entry for *did* is likely to fail. Similar points can be made about the flapping and glottalisation of the /t/ phonemes in /hit/ and /it/, and the vowel reductions to schwa. In addition, (1) illustrates the well-known point that there are no 100% reliable phonetic or phonological cues to word boundaries in connected speech. Without further phonological and lexical analysis there is no indication in a transcription like (1b) of where words begin or end; for example, how does the lexical access system distinguish word-initial /I/ in /I?/ from word-internal /I/ in /hIdl/?

In this paper, I shall argue for a model which splits the lexical access process into a pre-lexical phonological parsing stage and then a lexical entry retrieval stage. The model is similar to that of Church (1987), however I argue, firstly, that the initial phonological representation recovered from the speech input is more variable and often less detailed than that assumed by Church and, secondly, that the lexical entry retrieval stage is more directed and discriminatory, in order to reduce the number of spurious lexical entries accessed and to compensate for likely indeterminacies in the initial representation.

## THE PRE-LEXICAL PHONOLOGICAL REPRESENTATION

Several researchers have argued that phonological processes, such as the palatalisation of /d/ in (1), create problems for the word recognition system because they 'distort' the phonological form of the word. Church (1987) and Frazier (1987) argue persuasively that, far from creating problems, such phonological processes provide important clues to the correct syllabic segmentation of the input and thus, to the location of word boundaries. However, this argument only goes through on the assumption that quite detailed 'narrow' phonetic information is recovered from the signal, such as aspiration of /t/ in /tE/ and /tam/ in (1) in order to recognise the preceding syllable boundaries. It is only in terms of this representation that phonological processes can be recognised and their effects 'undone' in order to allow correct matching of the input against the canonical phonological representations contained in lexical entries.

Other researchers (e.g. Shipman & Zue, 1982) have argued (in the context of isolated word recognition) that

the initial representation which contacts the lexicon should be a broad manner-class transcription of the stressed syllables in the speech signal. The evidence in favour of this approach is, firstly, that extraction of more detailed information is notoriously difficult and, secondly, that a broad transcription of this type appears to be very effective in partitioning the English lexicon into small cohorts. For example, Huttonlocher (1985) reports an average cohort size of 21 words for a 20,000 word lexicon using a six-category manner of articulation transcription scheme (employing the categories: Stop, Strong-Fricative, Weak-Fricative, Nasal, Glide-Liquid, and Vowel).

This claim suggests that the English lexicon is functionally organised to favour a system which initiates lexical access from a broad manner class pre-lexical representation, because most of the discriminatory information between different words is concentrated in the manner articulation of stressed syllables. Elsewhere, we have argued that these ideas are misleadingly presented and that there is, in fact, no significant advantage for manner information in stressed syllables (e.g. Carter et al., 1987; Carter, 1987, 1989). We found that there is no advantage *per se* to a manner class analysis of stressed syllables, since a similar analysis of unstressed syllables is as discriminatory and yields as good a partitioning of the English lexicon. However, concentrating on a full phonemic analysis of stressed syllables provides about 10% more information than a similar analysis of unstressed syllables. This research suggests, then, that the pre-lexical representation used to initiate lexical access can only afford to concentrate exclusively on stressed syllables if these are analysed (at least) phonemically. None of these studies consider the extractability of the classifications from speech input; however, whilst there is a general belief that it is easier to extract information from stressed portions of the signal, there is little reason to believe that manner class information is, in general, more or less accessible than other phonologically relevant features.

A second argument which can be made against the use of broad representations to contact the lexicon (in the context of connected speech) is that such representations will not support the phonological parsing necessary to 'undo' such processes as palatalisation. For example, in (1) the final /d/ of *did* will be realised as /j/ and categorised as a strong-fricative followed by liquid-glide using the proposed broad manner transcription. Therefore, palatalisation will need to be recognised before the required stop-vowel-stop representation can be recovered and used to initiate lexical access. However, applying such phonological rules in a constrained and useful manner requires a more detailed input transcription. Palatalisation illustrates this point very clearly; not all sequences which will be transcribed as strong-fricative followed by liquid-glide can undergo this process by any means (e.g. /gl/), but there will be no way of preventing the rule overapplying in many inappropriate contexts and thus presumably leading to the generation of many spurious word candidates.

A third argument against the use of exclusively broad representations is that these representations will not support the effective recognition of syllable-boundaries and some word-boundaries on the basis of phonotactic and other phonological sequencing constraints. For example, Church (1987) proposes an initial syllabification of the input as a prerequisite to lexical access, but his syllabification of the speech input exploits phonotactic constraints and relies on the extraction of allophonic features, such as aspiration, to guide this process. Similarly, Harrington et al. (1988) argue that approximately 45% of word boundaries are, in principle, recognisable because they occur in phoneme sequences which are rare or forbidden word-internally. However, exploitation of these English phonological constraints would be considerably impaired if the pre-lexical representation of the input is restricted to a broad classification.

It might seem self-evident that people are able to recognise phonemes in speech, but in fact the psychological evidence suggests that this ability is mediated by the output of the word recognition process rather than being an essential prerequisite to its success. Phoneme-monitoring experiments, in which subjects listen for specified phonemes in speech, are sensitive to lexical effects such as word frequency, semantic association, and so forth (see Cutler et al., 1987 for a summary of the experimental literature and putative explanation of the effect), suggesting that information concerning at least some of the phonetic content of a word is not available until after the word is recognised. Thus, people's ability to recognise phonemes tells us very little about the nature of the representation used to initiate lexical access. Better (but still indirect) evidence comes from mispronunciation monitoring and phoneme confusion experiments (Cole, 1973; Miller & Nicely, 1955; Shepard, 1972) which suggest that listeners are likely to confuse or conflate phonemes along the dimensions predicted by distinctive feature theory. Most errors result in reporting phonemes which differ in only one feature from the target. This result suggests that listeners are actively considering detailed phonetic information along a number of dimensions (rather than simply, say, manner of articulation).

Theoretical and experimental considerations suggest then that, regardless of the current capabilities of automated acoustic-phonetic front-ends, systems must be developed to extract as phonetically detailed a pre-lexical phonological representation as possible. Without such a representation, phonological processes cannot be effectively recognised and compensated for in the word recognition process and the 'extra' information conveyed in stressed syllables cannot be exploited. Nevertheless in fluent connected speech, unstressed syllables often undergo phonological processes which render them highly indeterminate; for example, the vowel reductions in (1). Therefore, it is implausible to assume that any (human or machine) front-end will always output an accurate narrow phonetic, phonemic or perhaps even broad (say, manner class) transcription of the speech input. For this reason, further processes involved in lexical access will need to function effectively despite

the very variable quality of information extracted from the speech signal.

This last point creates a serious difficulty for the design of effective phonological parsers. Church (1987), for example, allows himself the idealisation of an accurate 'narrow' phonetic transcription. It remains to be demonstrated that any parsing techniques developed for determinate symbolic input will transfer effectively to real speech input (and such a test may have to await considerably better automated front-ends). For the purposes of the next section, I assume that some such account of phonological parsing can be developed and that the pre-lexical representation used to initiate lexical access is one in which phonological processes have been 'undone' in order to construct a representation close to the canonical (phonemic) representation of a word's pronunciation. However, I do not assume that this representation will necessarily be accurate to the same degree of detail throughout the input.

## LEXICAL ACCESS STRATEGIES

Any theory of word recognition must provide a mechanism for the segmentation of connected speech into words. In effect, the theory must explain how the process of lexical access is triggered at appropriate points in the speech signal in the absence of completely reliable phonetic/phonological cues to word boundaries. The various theories of lexical access and word recognition in connected speech propose mechanisms which appear to cover the full spectrum of logical possibilities. Klatt (1979) suggests that lexical access is triggered off each successive spectral frame derived from the signal (i.e. approximately every 5 msec.), McClelland & Elman (1986) suggest each successive phoneme, Church (1987) suggests each syllable onset, Grosjean & Gee (1987) suggest each stressed syllable onset, and Cutler & Norris (1985) suggest each prosodically strong syllable onset. Finally, Marslen-Wilson & Welsh (1978) suggest that segmentation of the speech input and recognition of word boundaries is an indivisible process in which the endpoint of the previous word defines the point at which lexical access is triggered again.

Some of these access strategies have been evaluated with respect to three input transcriptions (which are plausible candidates for the pre-lexical representation on the basis of the work discussed in the previous section) in the context of a realistic sized lexicon. The experiment involved one sentence taken from a reading of the 'Rainbow passage' which had been analysed by several phoneticians for independent purposes. This sentence is reproduced in (2a) with the syllables which were judged to be strong by the phoneticians underlined.

(2)

- a) The rainbow is a division of white light into many beautiful colours
- b) WF-V reIn bEu V-SF V S-V vI SF-V-N V-SF walt laIt V-N S-V men V bju: S-V WF-V-G K^I V-SF

This utterance was transcribed: 1) fine class, using phonemic transcription throughout; 2) mid class, using phonemic transcription of strong syllables and a six-category manner of articulation transcription of weak syllables; 3) broad class, as mid class but suppressing voicing distinctions in the strong syllable transcriptions. (2b) gives the mid class transcription of the utterance. In this transcription, phonemes are represented in a manner compatible with the scheme employed in the *Longman Dictionary of Contemporary English* and the manner class categories in capitals are Stop, Strong-Fricative, Weak-Fricative, Nasal, Glide-liquid, and Vowel, as in Huttenlocher (1982) and elsewhere. The terms, fine, mid and broad, for each transcription scheme are intended purely descriptively and are not necessarily related to other uses of these terms in the literature. Each of the schemes is intended to represent a possible behaviour of an acoustic-phonetic front-end. The less determinate transcriptions can be viewed either as the result of transcription errors and indeterminacies or as the output of a less ambitious front-end design. The definition of syllable boundary employed is, of necessity, that built into the syllable parser which acts as the interface to the dictionary database (e.g. Carter, 1989). The parser syllabifies phonemic transcriptions according to the phonotactic constraints given in Gimson (1980) and utilises the *maximal onset principle* (Selkirk, 1978) where this leads to ambiguity.

Each of the three transcriptions was used as a putative pre-lexical representation to test some of the different access strategies, which were used to initiate lexical look-up into the dictionary database. The four access strategies which were tested were: 1) phoneme, using each successive phoneme to trigger an access attempt; 2) word, using the offset of the previous (correct) word in the input to control access attempts; 3) syllable, attempting look-up at each syllable boundary; 4) strong syllable, attempting look-up at each strong syllable boundary. That is, the first strategy assumes a word may begin at any phoneme boundary, the second that a word may only begin at the end of the previous one, the third that a word may begin at any syllable boundary, and the fourth that a word may begin at a strong syllable boundary.

The strong syllable strategy uses a separate look-up process for typically unstressed grammatical, closed-class vocabulary and allows the possibility of extending look-up 'backwards' over one preceding weak syllable. It was assumed, for the purposes of the experiment, that look-up off weak syllables would be restricted to closed-class vocabulary, would not extend into a strong syllable, and that this process would precede attempts to incorporate a weak syllable 'backwards' into an open-class word.

The direct access approach was not considered because of its implausibility in the light of the discussion in the previous section. The stressed syllable account is very similar to the strong syllable approach, but given the problem of stress shift in fluent speech, a formulation in terms of strong syllables, which are defined in terms of the absence of vowel reduction, is preferable.

Work by Marslen-Wilson and his colleagues (e.g. Marslen-Wilson & Warren, 1987) suggests that, whatever access strategy is used, there is no delay in the availability of information derived from the speech signal to further select from the cohort of word candidates. This suggests that a model in which units (say syllables) of the pre-lexical representation are 'pre-packaged' and then used to trigger a look-up attempt are implausible. Rather the look-up process must involve the continuous integration of information from the pre-lexical representation immediately it becomes available. Thus the question of access strategy concerns only the *points* at which this look-up process is initiated.

In order to simulate the continuous aspect of lexical access using the dictionary database, database look-up queries for each strategy were initiated using the two phonemes/segments from the trigger point and then again with three phonemes/segments and so on until no further English words in the database were compatible with the look-up query (except for closed-class access with the strong syllable strategy where a strong syllable terminated the sequence of accesses). The size of the resulting cohorts was measured for each successively larger query; for example, using a fine class transcription and triggering access from the /h/ of *rainbow* yields an initial cohort of 89 candidates compatible with /reɪ/. This cohort drops to 12 words when /n/ is added and to 1 word when /b/ is also included and finally goes to 0 when the vowel of is is added. Each sequence of queries of this type which all begin at the same point in the signal will be referred to as an access path. The difference between the access strategies is mostly in the number of distinct access paths they generate.

Simulating access attempts using the dictionary database involves generating database queries consisting of partial phonological representations which return sets of words and entries which satisfy the query. For example, Figure 1 represents the query corresponding to the complete broad-class transcription of *appoint*. This query matches 37 word forms in the database.

```
[[pron
  [nsylls 2]
  [s1
    [peak ?]
  [s2
    [stress 2]
    [onset (OR b d g k p t)]
    [peak ?]
    [coda (OR m n N)
      (OR b d g k p t)]]]]
```

Figure 1 - Database query for 'appoint'.

The experiment involved generating sequences of queries of this type and recording the number of words found in the database which matched each query. Figure 2 shows the partial word lattice for the mid class transcription of the *rainbow is*, using the strong syllable access strategy. In this lattice access paths involving successively larger portions of the signal are illustrated. The number under each access attempt represents the size of the set of words whose phonology is compatible

with the query. Lines preceded by an arrow indicate a query which forms part of an access path, adding a further segment to the query above it.

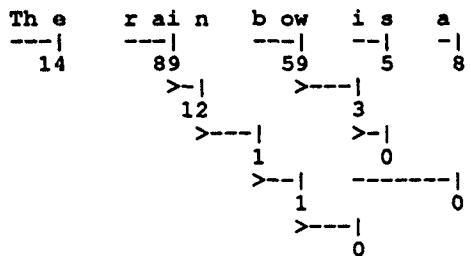


Figure 2 – Partial Word Lattice

The corresponding complete word lattice for the same portion of input using a mid-class transcription and the strong syllable strategy is shown in Figure 3. In this lattice, only words whose complete phonology is compatible with the input are shown.

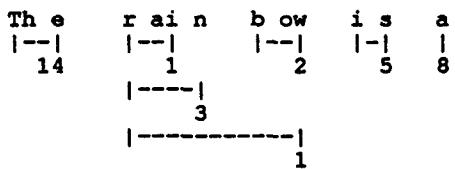


Figure 3 – Complete Word Lattice

The different strategies were evaluated relative to the 3 transcription schemes by summing the total number of partial words matched for the test sentence under each strategy and transcription and also by looking at the total number of complete words matched.

## RESULTS

Table 1 below gives a selection of the more important results for each strategy by transcription scheme for the test sentence in (2). Column 1 shows the total number of access paths initiated for the test sentence under each strategy. Columns 2 to 6 shows the number of words in all the cohorts produced by the particular access strategy for the test sentence after 2 to 6 phonemes/segments of the transcription have been incorporated into each access path. Column 7 shows the total number of words which achieve a complete match during the application of the particular access strategy to the test sentence.

Table 1 provides an index of the efficiency of each access strategy in terms of the overall number of candidate words which appear in cohorts and also the overall number of words which receive a full match for the test sentence. In addition, the relative performance of each strategy as the transcription scheme becomes less determinate is clear.

The test sentence contains 12 words, 20 syllables, and 45 phonemes; for the purposes of this experiment the word *a* in the test sentence does not trigger a look-up attempt with the word strategy because cohort sizes were only recorded for sequences of two or more phonemes/segments. Assuming a fine class transcription serving as pre-lexical input, the phoneme strategy produces 41 full matches as compared to 20 for the strong syllable strategy. This demonstrates that the strong syllable strategy is more effective at ruling out spurious word candidates for the test sentence. Furthermore, the total number of candidates considered using the phoneme strategy is 1544 (after 2 phonemes/segments) but only 720 for the strong syllable strategy, again indicating the greater effectiveness of the latter strategy. When we

Access Strategy	Access Paths	No. of words after x segments:						Complete Matches
		2	3	4	5	6		
<b>Fine Class</b>								
Phoneme	45	1544	251	46	6	2	41	
Word	11	719	193	32	5	2	25	
Syllable	20	1090	210	36	6	2	28	
StrongS	17	720	105	24	5	2	20	
<b>Mid Class</b>								
Word	11	4701	1738	802	54	8	249	
Syllable	20	12995	3221	1530	103	9	380	
StrongS	17	760	232	89	13	4	80	
<b>Broad Class</b>								
Syllable	20	13744	3407	1591	140	23	402	
StrongS	17	1170	228	100	18	9	117	

Table 1

consider the less determinate transcriptions it becomes even clearer that only the strong syllable strategy remains reasonably effective and does not result in a massive increase in the number of spurious candidates accessed and fully matched. (The phoneme strategy results are not reported for mid and broad class transcriptions because the cohort sizes were too large for the database query facilities to cope reliably.)

The word candidates recovered using the phoneme strategy with a fine class transcription include 10 full matches resulting from accesses triggered at non-syllabic boundaries; for example *arraign* is found using the second phoneme of *the* and *rain*. This problem becomes considerably worse when moving to a less determinate transcription, illustrating very clearly the undesirable consequences of ignoring the basic linguistic constraint that word boundaries occur at syllable boundaries. Systems such as TRACE (McClelland & Elman, 1986) which use this strategy appear to compensate by using a global best-fit evaluation metric for the entire utterance which strongly disfavours 'unattached' input. However, these models still make the implausible claim that candidates like *arraign* will be highly-activated by the speech input.

The results concerning the word based strategy presume that it is possible to determinately recognise the endpoint of the preceding word. This assumption is based on the Cohort theory claim (e.g. Marslen-Wilson & Welsh, 1978) that words can be recognised before their acoustic offset, using syntactic and semantic expectations to filter the cohort. This claim has been challenged experimentally by Grosjean (1985) and Bard et al. (1988) who demonstrate that many monosyllabic words in context are not recognised until after their acoustic offset. The experiment reported here supports this experimental result because even with the fine class transcription there are 5 word candidates which extend beyond the correct word boundary and 11 full matches which end before the correct boundary. With the mid class transcription, these numbers rise to 849 and 57, respectively. It seems implausible that expectation-based constraints could be powerful enough to correctly select a unique candidate before its acoustic offset in all contexts. Therefore, the results for the word strategy reported here are overly-optimistic, because in order to guarantee that the correct sequence of words are in the cohorts recovered from the input, a lexical access system based on the word strategy would need to operate nondeterministically; that is, it would need to consider several potential word boundaries in most cases. Therefore, the results for a practical system based on this approach are likely to be significantly worse.

The syllable strategy is effective under the assumption of a determinate and accurate phonemic pre-lexical representation, but once we abandon this idealisation, the effectiveness of this strategy declines sharply. Under the plausible assumption that the pre-lexical input representation is likely to be least accurate/determinate for unstressed/weak syllables, the strong syllable strategy is far more robust. This is a direct consequence of triggering look-up attempts off the

more determinate parts of the pre-lexical representation. Further theoretical evidence in support of the strong syllable strategy is provided by Cutler & Carter (1987) who demonstrate that a listener is six times more likely to encounter a word with a prosodically strong initial syllable than one with a weak initial syllable when listening to English speech. Experimental evidence is provided by Cutler & Norris (1988) who report results which suggest that listeners tend to treat strong, but not weak, syllables as appropriate points at which to undertake pre-lexical segmentation of the speech input.

The architecture of a lexical access system based on the syllable strategy can be quite simple in terms of the organisation of the lexicon and its access routines. It is only necessary to index the lexicon by syllable types (Church, 1987). By contrast, the strong syllable strategy requires a separate closed-class word lexicon and access system, indexing of the open-class vocabulary by strong syllable and a more complex matching procedure capable of incorporating preceding weak syllables for words such as *division*. Nevertheless, the experimental results reported here suggest that the extra complexity is warranted because the resulting system will be considerably more robust in the face of inaccurate or indeterminate input concerning the nature of the weak syllables in the input utterance.

## CONCLUSION

The experiment reported above suggests that the strong syllable access strategy will provide the most effective technique for producing minimal cohorts guaranteed to contain the correct word candidate from a pre-lexical phonological representation which may be partly inaccurate or indeterminate. Further work to be undertaken includes the rerunning of the experiment with further input transcriptions containing pseudo-random typical phoneme perception errors and the inclusion of further test sentences designed to yield a 'phonetically-balanced' corpus. In addition, the relative internal discriminability (in terms of further phonological and 'higher-level' syntactic and semantic constraints) of the word candidates in the varying cohorts generated with the different strategies should be examined.

The importance of making use of a dictionary database with a realistic vocabulary size in order to evaluate proposals concerning lexical access and word recognition systems is highlighted by the results of this experiment, which demonstrate the theoretical implausibility of many of the proposals in the literature when we consider the consequences in a simulation involving more than a few hundred illustrative words.

## ACKNOWLEDGEMENTS

I would like to thank Longman Group Ltd. for making the typesetting tape of the *Longman Dictionary of Contemporary English* available to us for research purposes. Part of the work reported here was supported by SERC grant GR/D/4217. I also thank Anne Cutler, Francis Nolan and Jim Sholicar for useful comments and advice. All errors remain my own.

## REFERENCES

- Bard, E., Shillcock, R. & Altmann, G. (1988). The recognition of words after their acoustic offsets in spontaneous speech: effects of subsequent context. *Perception & Psychophysics*, 44, 395-408.
- Boguraev, B. & Briscoe, E. (1989). *Computational Lexicography for Natural Language Processing*. Longman Limited, London.
- Boguraev, B., Carter, D. & Briscoe, E. (1987). A multi-purpose interface to an on-line dictionary. *3rd Conference of Eur. Assoc. for Computational Linguistics*, Copenhagen.
- Bradley, D. & Forster, K. (1987). A reader's view of listening. *Cognition*, 25, 103-34.
- Carter, D. (1987). An information-theoretic analysis of phonetic dictionary access. *Computer Speech and Language*, 2, 1-11.
- Carter, D., Boguraev, B. & Briscoe, E. (1987). Lexical stress and phonetic information: which segments are most informative. *Proc. of Eur. Conference on Speech Technology*, Edinburgh.
- Carter, D. (1989). LDOCE and speech recognition. In Boguraev & Briscoe (1989) pp. 135-52.
- Church, K. (1987). Phonological parsing and lexical retrieval. *Cognition*, 25, 53-69.
- Cole, R. (1973). Listening for mispronunciations: a measure of what we hear during speech. *Perception & Psychophysics*, 1, 153-6.
- Cutler, A. & Carter, D. (1987). The predominance of strong initial syllables in the English vocabulary. *Computer Speech and Language*, 2, 133-42.
- Cutler, A., Mehler, J., Norris, D. & Segui, J. (1987). Phoneme identification and the lexicon. *Cognitive Psychology*, 19, 141-77.
- Cutler, A. & Norris, D. (1988). The role of strong syllables in segmentation for lexical access. *J. of Experimental Psychology: Human Perception and Performance*, 14, 113-21.
- Frazier, L. (1987). Structure in auditory word recognition. *Cognition*, 25, 157-87.
- Gimson, A. (1980). *An Introduction to the Pronunciation of English*. 3rd Edition, Edward Arnold, London.
- Grosjean, F. & Gee, J. (1987). Prosodic structure and spoken word recognition. *Cognition*, 25, 135-155.
- Harrington, J., Watson, G. & Cooper, M. (1988). Word boundary identification from phoneme sequence constraints in automatic continuous speech recognition. *Proc. of 12th Int. Conf. on Computational Linguistics*, Budapest, pp. 225-30.
- Huttenlocher, D. (1985). Exploiting sequential phonetic constraints in recognizing spoken words. MIT. AI. Lab. Memo 867.
- Klatt, D. (1979). Speech perception: a model of acoustic-phonetic analysis and lexical access. *Journal of Phonetics*, 7, 279-312.
- Marslen-Wilson, M. (1987). Functional parallelism in spoken word recognition. *Cognition*, 25, 71-102.
- Marslen-Wilson, W. & Warren, P. (1987). Continuous uptake of acoustic cues in spoken word recognition. *Perception & Psychophysics*, 41, 262-75.
- Marslen-Wilson, W. & Welsh, A. (1978). Processing interactions and lexical access during word recognition in continuous speech. *Cognitive Psychology*, 10, 29-63.
- McClelland, J. & Elman, J. (1986). The TRACE model of speech perception. *Cognitive Psychology*, 18, 1-86.
- Miller, G. & Nicely, P. (1955). Analysis of some perceptual confusions among some English consonants. *Journal of Acoustical Society of America*, 27, 338-52.
- Sakoe, H. & Chiba, S. (1971). A dynamic programming optimization for spoken word recognition. *IEEE Transactions, Acoustics, Speech and Signal Processing*, ASSP-26, 43-49.
- Selkirk, E. (1978). On prosodic structure and its relation to syntactic structure. Indiana University Linguistics Club, Bloomington, Indiana.
- Shepard, R. (1972). Psychological representation of speech sounds. In David, E. & Denes, P. *Human Communication: A Unified View*, New York: McGraw-Hill.
- Shipman, D. & Zue, V. (1982). Properties of large lexicons: implications for advanced isolated word recognition systems. *IEEE ICASSP*, Paris, 546-549.
- Wiese, R. (1986). The role of phonology in speech processing. *Proc. of 11th Int. Conf. on Computational Linguistics*, Bonn, pp. 608-11.
- Wilson, M. (1988). MRC psycholinguistic database: machine-useable dictionary, version 2.0 *Behaviour Research Methods, Instrumentation & Computers*, 20, 6-10.
- Zue, V. & Huttenlocher, D. (1983). Computer recognition of isolated words from large vocabularies. *IEEE Conference on Trends and Applications*.