Computer Speech and Language (1989) 3, 265–275

Lexical stress and lexical discriminability: Stressed syllables are more informative, but why?

Gerry Altman*

Centre for Speech Technology Research, University of Edinburgh, Edinburgh, U.K.

and David Carter

SRI International Cambridge Research Centre, Cambridge, U.K.

Abstract

Recent studies have suggested that recognition systems should concentrate their efforts on the identification of stressed syllables, as they contain disproportionately more information than do unstressed syllables. The paper investigates whether this increased informativeness may be outweighed by the informational disadvantage associated with transcribing consecutive segments within the same syllable. Phonotactic correlations between such adjacent segments suggest that the most informative transcription of a polysyllabic word may be one where reliable phonemic information is scattered across different syllables. Lexical statistics are presented which support this view. In addition, the paper considers the reasons for the increased informativeness of stressed syllables, and shows that this is because lexical stress preserves vowel distinctions (and hence information) which would otherwise be lost in lexically unstressed syllables.

1. Introduction

Statistical studies of large lexicons have been concerned recently with determining which portions of a word are more informative in terms of distinguishing that word from rival candidates in the lexicon. Lexical statistics typically generate information concerning equivalence classes, where an equivalence class is simply a set of words in the lexicon which are indistinguishable from one another following the application of some selection procedure. For instance, knowing only the stress pattern of a word (e.g. stressed unstressed) would not discriminate between the words "vocal" and "speechless", although partitioning the lexicon by stress pattern does lead to equivalence classes with membership sizes considerably smaller than the size of the lexicon as a whole (15–19%, Waibel, 1982; Aull & Zue, 1984; Carter, 1987). Thus knowing the stress pattern of a word does constrain the candidate search space, although used alone it would not go particularly far in helping to identify the word. Lexical statistics based on a particular

*Current address: Laboratory of Experimental Psychology, Sussex University, Brighton BN1 9QG, UK.

0885–2308/89/030265 + 11 $03.00/0

partitioning of the lexicon (e.g. according to stress pattern alone) establish the magnitude of the constraining power of that partition. By manipulating the specificity of information located word-internally, the relative informativeness of different portions of the same word can be assessed. For instance, preserving phonemic information in just the first syllable, and ignoring the information in subsequent syllables, allows the contribution of word-initial syllables to constraining the candidate search space to be assessed.

On the basis of a series of experiments on different kinds of partitioning, Huttenlocher (1984) claimed that the phonetic information contained within a *lexically* stressed syllable is more constraining, or informative, than information contained within unstressed syllables. Huttenlocher based his claims on the following experiment. The 20 000 words from the Merriam–Webster pocket dictionary were all transcribed to just the "broad class" phonetic level (Shipman & Zue, 1982). At this level there are in effect just six "phonemes"; vowels, nasals, liquids/glides, strong fricatives, weak fricatives, and stops. The word "speechless" would thus be transcribed as /sfric stop vowel sfric liq-gli vowel sfric/. Huttenlocher showed (following Shipman & Zue) that using only broad class distinctions led to equivalence classes which contained at most just 1% of the original size of the lexicon, and that 32% of the lexicon was uniquely discriminable. Huttenlocher then applied two further partitionings: in one, he transcribed to broad class only those segments contained within the lexically stressed syllable, and used a "wild card" to represent the presence of an unstressed syllable (e.g. for "speechless": /sfric stop vowel sfric */, where "*" is the wild card); in the other, he transcribed to broad class only those segments contained within the unstressed syllables, using the wild card to represent the stressed syllable (e.g. /* liq-gli vowel sfric/). He found that both the expected class size (a weighted version of the mean class size) and the maximum class size were considerably larger when information was deleted from stressed segments than when it was deleted from unstressed segments. He concluded that information contained within the stressed syllable of a word is more informative than the information contained within the unstressed syllables, and that this additional informativeness is apparent irrespective of whether broad classes or full phonemes (fine classes) are used in the partitioning.

As Huttenlocher pointed out, stressed segments are likely to be acoustically more reliable than unstressed segments, and hence are more likely to be recognized successfully. Is it fortuitous that the maximally reliable parts of words are also the maximally informative? It is this question that the present paper seeks to investigate.

2. Lexical partitioning and information theory

Carter (1987) points out that the evaluation of lexical partitions on the basis of equivalence class size is misleading, and offers an information-theoretic approach to the problem of evaluating lexical partitionings. The new measure, the *percentage of information extracted* (PIE), is based partly on the following premise (ignoring, for the moment, the effects of word frequency): if the original size of the lexicon is 10 000, and a given partitioning leads to an expected class size of 1000 (i.e. 10% of the lexicon), then the contribution of that partitioning to the identification of any one word is 25%, not the 90% that is implied by quoting the 10% figure—implied because as long as we can identify just which equivalence class the word falls in, we can eliminate 90% of the lexicon from our search space. The figure of 25% arises because in reducing the lexicon from 10 000 to 1000, we have effected just *one* of the four 10-fold reductions that would be needed to achieve a unique identification—we have extracted only 25% of the information that would be required for complete identification. Information theory assumes that the reduction from 1000 to 100 requires the *same* amount of information as is required by the reduction from 100 to 10, or the reduction from 10 to 1, or even that from 10 000 to 1000.

Carter shows that the PIE measure is also superior to measures of equivalence class size in another respect: it is sensitive to the *relative* frequencies of words *within* an equivalence class, whereas class-size measures are, at best, only sensitive to *total* class frequencies (that is, the sum of the frequencies of the member words). Once a partial transcription has determined an equivalence class, the remaining uncertainty (information still to be extracted) is a function of the frequency *distribution* of all the words in the class, not of the total class frequency (for details of the mathematics involved in calculating frequency weighted PIEs, see Carter, 1987).

Using both his own information-theoretic approach and Huttenlocher's class sizes, Carter repeated Huttenlocher's original experiments. Although Huttenlocher's findings were replicated using expected class size, the PIE values indicated that the informational advantage of stressed syllables only exists if the full phonemic content of the stressed syllable is extracted. There was no informational advantage when only broad class information was extracted.

The following section describes some new studies which investigate whether the informational advantage of stressed syllables persists when *two* levels of phonemic specification are introduced, and applied differently to stressed and to unstressed segments. The two levels are fine class, and an intermediate "mid" class which lies between the broad and fine classes (Dalby, Laver, & Hiller, 1986).

3. Mid classes, fine classes, and lexical stress

In Huttenlocher's original paradigm, one syllable would be transcribed to broad class, and the rest would be left untranscribed, with a wild card to signify the presence of each untranscribed phoneme. Carter's (1987) figures show that the informational advantage of stressed syllables exists only when fine class, and not broad class, information is extracted. But what would happen if the critical syllable (whether stressed or unstressed) was transcribed to fine class, and the remaining syllables were transcribed to an intermediary class that is less specific than fine class, but more specific than broad class? Would the informational advantage of fine class stressed syllables still be apparent?

In order to address these questions, we have adopted the intermediary "mid class" level of transcription proposed by Dalby *et al.* (1986). Unlike the broad classes, distinctions are made between voiced and voiceless stops; voiced and voiceless weak or strong fricatives; liquids and glides; front, back, and central vowels; diphthongs, and nasals. Whereas there are only six broad class phonemes, there are 13 of these mid class phonemes.

Table I compares the following transcriptions:

(1) in which all the segments were transcribed to fine class (e.g. /s p ii ch l @ s/).[1]

[1] For the purposes of this paper we adopt the machine readable phonetic alphabet developed at Edinburgh. Its interpretation for the word /s p ii ch l @ s/ is transparent except for the vowel transcribed as /ii/ (the vowel common to "speech" and "peach") and the vowel /@/ which is schwa.

(2) in which all the segments were transcribed to mid class (e.g. /unvoiced_sfric unvoiced_stop front_vowel unvoiced_sfric liquid central_vowel unvoiced_sfric/).
(3) in which all the stressed segments (that is, segments falling within stressed syllables) were transcribed to fine class and all the unstressed segments to mid class (e.g. /s p ii ch liquid central_vowel unvoiced_sfric/).
(4) in which unstressed segments were transcribed to fine class with all the stressed segments transcribed to mid class (e.g. /unvoiced_sfric unvoiced_stop front-_vowel unvoiced_sfric 1 @ s/).

The proportion of unstressed syllables transcribed to fine class (67·4%, chosen at random) was chosen to make the total (frequency-weighted) proportion of fine class (unstressed) segments in the lexicon equal to the proportion of stressed segments in the lexicon (40·3%). This therefore ensured that the same numbers of segments were transcribed to fine class in conditions (3) and (4).

The present studies used the 12 850 words in the *Longman Dictionary of Contemporary English* for which word frequency information is supplied by the Brown corpus (Kucera & Francis, 1967) as represented in the MRC Psycholinguistic Database (Coltheart, 1981). This set of words was the same as that used in the experiments reported in Carter (1987). However, the treatment of stress in monosyllabic words was altered in the work described here in order to reflect phonetic reality more closely. Rather than treating *all* monosyllabic function words as unstressed, stress values (and transcriptions) were assigned according to whether or not LDOCE indicated alternative pronunciations for the monosyllable: if it did (as usually was the case for function words), then a pronunciation containing a schwa (or "reduced i" as found in the word "it" in a sentence like "NO, give it to ME"—where capitals indicate the main sentence accents) was always selected in preference to one without. This preference reflects the fact that reduced pronunciations are more common for such words. So for instance, the function words "by" and "no" were treated as stressed, whilst any word with a schwa or reduced i was treated as unstressed. The net effect of this decision was that slightly fewer unstressed syllables were assumed than in Carter (1987). All lexical, or content, monosyllabic words were treated as stressed.

Since even an all mid class transcription, the least detailed of those presented here, is capable of extracting the greater part of the information in a typical word, the results are presented not as PIE but as PINE values (*Percentage of Information Not Extracted*, i.e. 100% − PIE).

TABLE I. Frequency-weighted PINE values

(1) All segments transcribed to fine class	2·02%
(2) All segments transcribed to mid class	5·94%
(3) Stressed segments transcribed to fine class	3·27%
(4) 40·3% unstressed segments transcribed to fine class	5·08% ± 0·04%

The difference between conditions (2) and (1) shows that approximately 3·9% of the total information required to achieve unique lexical discrimination is lost by replacing fine class transcriptions with mid class transcriptions; this compares with approximately

18·9% that would be lost replacing fine classes with broad classes (Carter, 1987).[2] The difference between conditions (2) and (3) shows that adding fine class information around stressed syllables (with mid classes elsewhere) adds approximately 2·7% of the total information required. Adding the extra fine class information for the same overall proportion of segments (40·3%), but in unstressed position [condition (4)] gains only 0·9% extra information.[3] Although small, there does appear to be some informational advantage to stressed syllables when a mixed transcription (fine and mid class) is adopted. Does this mean, therefore, that there is some advantage to be gained by designing speech recognition systems to concentrate on stressed syllables? The following section considers whether such a system would be any better off than one which worked, say, at random.

4. Phonotactic constraints and the transcription of consecutive segments

In condition (3) above, but not in condition (4), the segments that were transcribed to fine class were all *consecutive*. However, consecutive segments are subject to short-range phonotactic dependencies; it is well known that the probability of one phoneme given its neighbours is in general higher than its probability given some other non-adjacent segment. This is the essence of phonotactic constraints (or *phoneme sequence constraints*—Harrington, Watson, & Cooper, 1988). In terms of partitioning the lexicon into word classes, the higher the probability of co-occurrence of two (or more) phonemes, the larger the word class defined by that co-occurrence; the less probable a co-occurrence, the more informative is the transcription which captures that co-occurrence. Altmann (1986) points out that it should follow that all other things being equal, a reliable transcription of segments subject to strong phonotactic constraints (e.g. within syllables, as opposed to across syllable boundaries) is *less* informative than a reliable transcription of segments which are *not* subject to phonotactic constraints (e.g. a reliable transcription of non-consecutive segments). So transcriptions with reliable information located in stressed syllables are not, therefore, necessarily *maximally* informative.

4.1. Random versus consecutive transcriptions

At issue is whether the extra informativeness of stressed syllables can outweigh the phonotactic disadvantage of transcribing adjacent segments. Given local context effects due to coarticulation, it is in fact more likely that a front-end will reliably transcribe

[2] Statistical comparisons were not performed because the calculation of PINE values does not involve averaging over *populations* of equivalence classes—equivalence classes are not actually constructed during the calculation of PINE values. The PINE value is a single descriptive statistic, and it would to our knowledge be impossible to compare two PINE values using any inferential statistic which relied on assessing population variance (or some equivalent). Our use of them here is confined, therefore, to providing a descriptive statistic which defines a property of the particular lexicon and associated frequency values which were used for this study.

[3] Because of the random element involved, experiment (4) was run 1000 times for monosyllables and 200 times for polysyllables. This gave a standard error (to 0·1% likelihood) of around 0·04% in the averaged PINE figure shown here.

A random element in a transcription means, as argued in Carter (1989) that the classes it invokes are not *equivalence* classes; that is, the transcription does not partition the lexicon into disjoint sets. All the PINE values reported here are therefore derived from the *consistency* classes that Carter shows result from such transcriptions. However, the distinction is not central to the arguments in this paper.

adjacent segments; if a segment's context is correctly identified, then the likelihood of that segment being itself correctly identified increases. The question is whether this property counts against the processor in terms of any associated informational disadvantage. In order to assess this, PINE values were computed for the following conditions:

(5) in which the segments to be transcribed to fine class are chosen at random within each word (the rest being transcribed to mid class)—and in which the number of fine class segments in the word is the same as the number of stressed segments *in that word* (e.g. /s unvoiced_stop front_vowel ch 1 central_vowel s/). The results are thus directly comparable to those reported under condition (3) above.

(6) in which the same number of segments is transcribed to fine class (on a word-for-word basis), but this time all the segments within the word are contiguous—the place in the word where the contiguous block starts being chosen at random (e.g. /unvoiced_sfric unvoiced_stop ii ch 1 @ unvoiced_sfric/).

(7) in which, as in (5), the same number of segments is transcribed to fine class, and the segments so transcribed are chosen at random, but this time with the constraint that as few stressed segments as possible are transcribed fine (e.g. /unvoiced_sfric unvoiced_stop front_vowel ch 1 @ s/).

On the assumption that stressed segments are not sufficiently informative to overcome phonotactic effects, condition (5) should produce a lower PINE value (that is, should extract more information) than condition (3) from Table I, in which all the (consecutive) segments falling within stressed syllables were transcribed to fine class. Moreover, it should also be the case that the PINE value for condition (6) should also be lower than that for condition (3). This is because in condition (6) there will, in general, be segments which are transcribed to fine class and which fall in *different* syllables, and which will therefore not be subject to phonotactic constraints. Consequently more information should be extracted in condition (6) than in condition (3) (again, this assumes that stressed segments are not sufficiently informative to overcome such phonotactic effects). Given the earlier finding that stressed syllables do have some increased information value, then it also follows that condition (7), which minimizes the number of stressed segments transcribed to fine class, should produce a higher PINE value than condition (5).

It is unclear what the expected relationship between conditions (5) and (6) should be; phonotactic effects should reduce the informativeness of transcriptions based on consecutive segments which lie within the *same* syllable. In both conditions (5) and (6), segments lying in *different* syllables are transcribed to fine class (although there may well be individual words in condition (6) for which this is not true, and for which the onset of the contiguous block coincided with the onset of a syllable). Both conditions, therefore, benefit from the phonotactic advantages that accrue by extracting fine class information from different syllables.

As described above, in conditions (5) to (7) the number of fine class segments are kept the same as the number of stressed segments in that word. This is in contrast to condition (4), in which the proportions were calculated over the lexicon as a whole. Given that the majority of words in the frequency-weighted lexicon are monosyllables (approximately 75%—calculated by taking into account the relative frequency in the language of each word in the lexicon), which never mix stressed and unstressed segments, the transcriptions used in conditions (5) to (7) will not in most cases (i.e. in monosyllabic words) do anything different from the transcriptions used in condition (3). Consequently, there should be little difference overall in the PINE figures for conditions (3), (5), (6), and (7). PINE values will therefore be reported separately for (a) all words, (b) monosyllabic words, and (c) polysyllabic words. The PINE values for monosyllabic words should not differ across the three conditions, whilst the PINE values for the polysyllabic words should reflect the differences predicted earlier. Of course, looking only at the figures for polysyllabic words is an artificial test of the predicted informational disadvantage of transcribing consecutive segments; the frequency-weighted lexicon, with its predominance of monosyllables, most closely reflects the target language over which a speech recogniser will have to operate, and thus the PINE values for the lexicon as a whole constitute a more realistic model of the "real-life" effect of phonotactic constraints.

TABLE II. Frequency-weighted PINE values

Condition	(a) All words	(b) Monosyllables	(c) Polysyllables
(3) Stressed	3·268%	5·332%	0·321%
(5) Random	3·234% ± 0·005%	5·332%	0·213% ± 0·01%
(6) Contiguous	3·235% ± 0·005%	5·332%	0·215% ± 0·01%
(7) Avoiding stressed	3·268% ± 0·005%	5·332%	0·320% ± 0·01%

4.2. Discussion

Table II suggests that in keeping with Altmann's (1986) predictions, there is a phonotactic disadvantage to transcribing consecutive segments which fall within the same syllable. The two random conditions (5) and (6) yield reliably lower PINE values than the stress based transcription (3),[4] and thus the increased informativeness of stressed syllables [which is confirmed by the difference between conditions (7) and (5)] is not sufficient to overcome the informational disadvantage associated with consecutive transcriptions within the same syllable. For polysyllabic words the random transcription procedure (5) did yield a slightly lower PINE value than the contiguous transcription procedure (6). However, this difference is smaller than the error values found for either condition, and is thus unreliable.[5]

In summary, phonotactic correlations between adjacent segments do reduce the informativeness of contiguous transcriptions, but given both a speech recognizer which can output reliable transcriptions of stressed segments, and the high probability with

[4] The error margins shown in Table II represent three standard deviations in the mean of the distribution of values obtained over 1000 runs for the monosyllabic sub-lexicon and 200 runs for the polysyllabic one, for all cases where the result was sensitive to random variation. Thus, the reported difference was significant at $p<0{\cdot}001$—in other words, the difference was consistent given that a number of different (random) transcriptions were tested. The statistical variations catered for here all arise from the randomness in some of the transcription conditions, and not from any other source, such as, for example, the original choice of texts for the Brown corpus—that is, given the corpus and our initial analysis of it, our results are not significantly affected by the random variations introduced by subsequent processing.

[5] Errors for the polysyllabic cases of conditions (5) to (7) are again for 0·1% likelihood, derived from 200 runs. There was no element of randomness in monosyllabic (5) to (7) because in each stressed monosyllable, all the segments were transcribed fine, and in each unstressed one, none were.

which that recognizer will encounter monosyllabic words, it appears to be the case that this reduced informativeness will hardly impair the recognizer's ability to partition the lexicon (the differences in PINE values for all words, as opposed to just polysyllabic words, are very slight indeed).

Phonotactic effects aside, it is still the case that stressed syllables *are* more informative than unstressed syllables. Moreover, the information gained by having a fine (as opposed to mid) class transcription in stressed position (2·7%, i.e. condition (2) minus condition (3)) is about three times the gain in unstressed position (0·9%, i.e. condition (2) minus condition (4)). In the following section it will be shown that this difference is also reflected in the *entropy* values for stressed and unstressed syllables with both fine and mid class transcriptions.

5. Information content and entropy

The amount of entropy in a system is a measure of the degree of uncertainty about just which of several possible states that system is in at any one time. For instance, if each of the 46 or so phonemes in the English language were equiprobable, then the entropy value, or information content, of each phoneme would be $\log_2 46$ (5·52)—roughly, the minimum number of bits of information required to encode the character set consisting of the 46 phonemes. The greater the character set—that is, the greater the choice at any one time—the more bits required, and the greater the information content of each character. Of course, the 46 phonemes in English are not equiprobable, but by taking into account their relative frequencies, one can calculate the weighted average information content of the character set. The weighted average will always be less than the unweighted average based on equal probabilities, because even distributions give rise to higher entropy values than uneven distributions. The difference between the two arises because more frequent symbols, if there are any, can be assigned shorter bit sequences in the encoding, and these sequences will occur relatively often.

In the remainder of this section, entropy values will first be used to confirm that the extra information gained from knowing fine class rather than mid class information in stressed syllables is approximately three times that for knowing fine class information in unstressed syllables. After calculating entropy values for syllables, average entropy values for vowels and consonants will be reported, in order to identify just which part of the stressed syllable conveys the extra information.

5.1. Entropy values for syllables

By calculating the entropy values for stressed and unstressed syllables with both fine and mid class transcriptions, the informativeness of each kind of syllable can be assessed, and hence the difference in informativeness between mid and fine class transcriptions in stressed or unstressed position can also be assessed. The values are shown in Table III. The entropy values shown in Table III demonstrate that by transcribing to mid class rather than to fine class, much less information is lost on unstressed syllables than on stressed syllables—in fact, about a third, confirming the figure derived on the basis of the PINE values reported in Section 3. Once again, the evidence suggests that stressed syllables are more informative than unstressed syllables. But what is it about stressed syllables that makes them more informative?

TABLE III. Entropy values for stressed and unstressed syllables

Transcription	Position: Stressed	Unstressed
Fine class	9·87	6·70
Mid class	8·12	6·11
Fine–Mid	1·74	0·59

5.2. Entropy values for consonants and vowels

In the preceding sub-section, the entropy associated with different parts of words, and specifically the different syllables, was calculated. The paradigm can be extended by going one step further and calculating the entropy associated with different parts of syllables, thereby uncovering where within the syllable the "informational load" is located. For instance, the increased informativeness of stressed syllables might simply arise from the increased informativeness of vowels in stressed position. Indeed, the results reported here and elsewhere (e.g. Huttenlocher, 1984; Carter, 1987) may be due to an uneven distribution of vowel categories across stressed and unstressed position. The latter will be dominated by central vowels in general, and *schwa* in particular. Thus, there may be more vowel categories to choose between in stressed position than in unstressed position, and hence to know the identity of a vowel in stressed position will be more informative than to know the identity of a vowel in unstressed position.

Table IV shows the entropy values for fine and mid class transcriptions of arbitrary vowels and consonants in both stressed and unstressed syllables. These results demonstrate quite conclusively that for consonants, there is virtually no difference in their information content between stressed and unstressed syllables, whether transcribing to mid class or fine class. For vowels there is comparatively little difference when they are transcribed to mid class—that is, mid class vowels convey almost as much information in unstressed position as they do in stressed position.

This is not true of fine class vowels, however. Fine class vowels convey only just over

TABLE IV. Entropy values for arbitrary consonants and vowels in stressed (S) or unstressed (U) position

	Transcription					
	Fine			Mid		
Position	S	U	U/S	S	U	U/S
Consonant	4·07	3·93	0·97	2·87	2·99	1·04
Vowel	3·83	2·19	0·57	1·84	1·44	0·78

half the information when in unstressed position than when in stressed position.[6] Perhaps surprisingly, the entropy value for mid class consonants is actually higher in the unstressed case than in the stressed case. This is because many of the unstressed syllables are words beginning with a voiced /th/ (e.g. the, this, they, etc.) or with a /w/ (e.g. why, which, where, etc.). Voiced weak fricatives and glides are quite rare in stressed syllables but not in unstressed syllables—there is thus a more even distribution of consonants in unstressed syllables (and as described earlier, the more even a distribution, the higher the entropy).

6. Summary and conclusions

This paper has considered the degree to which stressed syllables are more informative than unstressed syllables. The information-theoretic approach to evaluating lexical partitions developed in Carter (1987) provides a quantitative measure of such differences. Although the extra intrinsic informativeness of stressed syllables does not outweigh the phonotactic disadvantage of interpreting adjacent segments within the same syllable, it was shown that the preponderance of monosyllables in the language (75%—as represented by the frequency values taken from the corpus used) decreases the impact of such a disadvantage; it is not so much a disadvantage to transcribe adjacent segments within the same syllable as it is an advantage to transcribe segments in different syllables: monosyllabic words do not therefore permit such an advantage.

To be able to make use of the extra informativeness of stressed syllables assumes a processor which is capable of reliably transcribing adjacent segments to fine class in stressed syllables. If this assumption were to be violated, and the processor were not so reliable, and perhaps only had a greater probability of correct fine class transcription for any single stressed segment, then the phonotactic advantage of reliably transcribing *non-adjacent* segments in different syllables would work in the processor's favour. This has considerable implications for speech recognition systems; there is an informational advantage to reliably transcribing stressed segments, but there is also an informational advantage to reliably transcribing non-adjacent segments, which do not necessarily fall within the same (stressed or unstressed) syllable. So nothing would be lost, informationally speaking, if the acoustically reliable segments were not all adjacent, as they would be if the entire stressed syllable were to be transcribed. The PIE figures suggest that overall it does not really matter where the reliably transcribed segments are located. Indeed, the results suggest that the emphasis in automated speech recognition should be on the development of acoustic front-ends which are sensitive to acoustic reliability, as opposed to being sensitive to stress *per se*.

In going within the syllable, and deriving entropy values for segments within the syllables, it has been possible to identify just which aspect of stressed syllables is responsible for their increased informativeness. It arises not because of any factors operating at the lexical level itself (e.g. the fortuitous assignment of lexical stress to the most otherwise informative syllable), but because there are, in effect, more vowel *types* in stressed syllables than in unstressed syllables (where central vowels predominate).

[6] A further experiment indicated that the difference observed between mid class vowels was due to the fact, reported in footnote 1, that a pronunciation containing a schwa was always selected in preference to one without. The unreduced forms do occur, however; and in fact, if the selection is made arbitrarily instead of applying any preference, the entropy values for unstressed vowels increase so that the figures for mid class vowels become much closer, while those for fine class vowels remain very different.

Consequently to know the identity of a vowel in stressed position is to have distinguished between more competing vowel hypotheses. Hence the increased value of stressed syllables is not an intrinsic property of the lexicon; it is merely a result of the fact that stress preserves distinctions between vowels which would be lost if they were unstressed. Only if an acoustic front-end is sensitive to these vowel distinctions, can the informational value of stressed segments be used to the advantage of the processor.

The first author is grateful to Richard Shillcock and Ellen Bard for their advice and encouragement. The second author is grateful to Longman Group Ltd for allowing him access to the LDOCE typesetting tape for the purposes of research conducted during his employment at Cambridge University Computer Laboratory, on which the work reported here is based.

References

Altmann, G. (1986). Lexical stress, lexical discriminability, and variable phonetic information in speech recognition. *Proceedings of the Institute of Acoustics, Speech and Hearing*, **8**(7), 471–478.

Aull, A. M. & Zue, V. W. (1984). Lexical stress and its application to large vocabulary speech recognition. Paper presented at the 108th meeting of the Acoustical Society of America.

Carter, D. M. (1987). An information-theoretic analysis of phonetic dictionary access. *Computer Speech and Language*, **2**, 1–11.

Carter, D. M. (1989). LDOCE and Speech Recognition. In *Computational lexicography for natural language processing* (B. Boguraev, & E. Briscoe, eds). London: Longman (in press).

Coltheart, M. (1981). The MRC psycholinguistic database. *Quarterly Journal of Experimental Psychology*, **33A**, 497–505.

Dalby, J., Laver, J. & Hiller, S. M. (1986). Mid-class phonetic analysis for a continuous speech recognition system. *Proceedings of the Institute of Acoustics, Speech and Hearing*, **8**(7), 347–354.

Harrington, J., Watson, G. & Cooper, M. (1988). Word-boundary Identification from Phoneme Sequence Constraints in Automatic, Continuous Speech Recognition. *Proceedings of the 12th international conference on computational linguistics*, Coling, Budapest.

Huttenlocher, D. P. (1984). Acoustic-phonetic and lexical constraints in word recognition: lexical access using partial information. M.S. Thesis, Massachussetts Institute of Technology.

Kucera, H. & Francis, W. N. (1967). *A computational analysis of present-day American English*. Brown University Press, Providence, Rhode Island.

Shipman, D. W. & Zue, V. W. (1982). Properties of large lexicons: implications for advanced isolated word recognition systems. *Proceedings of the international conference on acoustics, speech and signal processing*, pp. 546–549. Paris, France.

Waibel, A. (1982). Towards very large vocabulary word recognition. CMU-CS-82-144, Department of Computer Science, Carnegie-Mellon University.