The strong/weak syllable distinction in English

Beverley D. Fear
Department of Engineering, University of Cambridge, Cambridge, United Kingdom

Anne Cutler
MRC Applied Psychology Unit, Cambridge, United Kingdom and Max-Planck-Institute for Psycholinguistics, Nijmegen, The Netherlands

Sally Butterfield
MRC Applied Psychology Unit, Cambridge, United Kingdom

(Received 30 June 1993; revised 18 October 1994; accepted 24 October 1994)

Strong and weak syllables in English can be distinguished on the basis of vowel quality, of stress, or of both factors. Critical for deciding between these factors are syllables containing unstressed unreduced vowels, such as the first syllable of *automata*. In this study 12 speakers produced sentences containing matched sets of words with initial vowels ranging from stressed to reduced, at normal and at fast speech rates. Measurements of the duration, intensity, $F0$, and spectral characteristics of the word-initial vowels showed that unstressed unreduced vowels differed significantly from both stressed and reduced vowels. This result held true across speaker sex and dialect. The vowels produced by one speaker were then cross-spliced across the words within each set, and the resulting words' acceptability was rated by listeners. In general, cross-spliced words were only rated significantly less acceptable than unspliced words when reduced vowels interchanged with any other vowel. Correlations between rated acceptability and acoustic characteristics of the cross-spliced words demonstrated that listeners were attending to duration, intensity, and spectral characteristics. Together these results suggest that unstressed unreduced vowels in English pattern differently from both stressed and reduced vowels, so that no acoustic support for a binary categorical distinction exists; nevertheless, listeners make such a distinction, grouping unstressed unreduced vowels by preference with stressed vowels.

PACS numbers: 43.70.Fq, 43.71.Es

INTRODUCTION

The speech rhythm of English is principally determined by the opposition between strong and weak syllables. English is a stress language, and its rhythm is stress based. All stressed syllables are strong, and all weak syllables are unstressed. Whether a syllable is strong or weak could therefore be seen as wholly a function of whether or not it is stressed. According to this distinction, strong syllables are defined as stressed syllables, while weak syllables are defined as unstressed syllables (this is the definition used in verse metrics, for example; Halle and Keyser, 1971). An alternative definition, however, equates strong syllables with those containing full vowel quality, while weak syllables are defined as those with central, or reduced, vowels, usually schwa (see, e.g., Bolinger, 1981).

The two definitions are not equivalent, though, because some unstressed syllables have unreduced vowels; the first vowel of *automata*, for instance, is noncentral, but carries neither primary nor secondary stress. According to a stress-based definition it would be weak, but according to a vowel-based definition it would be strong. It could be argued therefore that the strong–weak distinction is in fact not an exhaustive categorical division at all, but maps rather onto a continuum in which both stress and vowel quality play a role, and on which unstressed unreduced syllables occupy an intermediate position between stressed syllables and reduced syllables.

The question is important because although the distinction between strong and weak syllables is a phonological one, recent evidence has demonstrated that it plays a role in perception. Studies of speech segmentation show that in English, listeners use the strong–weak distinction as a guide to locating boundary points in a speech signal. For instance, strong–weak sequences such as [lɛtəs] tend to be perceived as one word (*lettuce*) rather than two (*let us*; Taft, 1984); monosyllabic words embedded in nonsense bisyllables are easy to detect if they span a boundary between a strong and a weak syllable, but hard to detect if they span a boundary between two strong syllables (e.g., *mint* is easier to detect in [mɪntəf] than in [mɪntef]; Cutler and Norris, 1988; McQueen *et al.*, 1994; Norris *et al.*, 1995), and segmentation errors in speech more often consist of postulating erroneous boundaries before strong syllables and overlooking boundaries before weak syllables than of the converse (Cutler and Butterfield, 1992).

Production studies lend further support to the importance of the strong/weak distinction; when speakers are deliberately trying to articulate clearly, they pause at word boundaries preceding weak syllables but not at word boundaries preceding strong syllables, i.e., they mark precisely those boundaries which the observed listener behaviors would *not* detect (Cutler and Butterfield, 1990).

Cutler and Norris (1988) and Cutler and Butterfield (1992) proposed that listeners treat strong syllables as if they

 0001-4966/95/97(3)/1893/12/$6.00

are highly likely to be lexical word onsets, and that segmenting speech at strong syllable onsets is a way of solving the problem posed to the listener by the absence of robust and reliable cues to word boundaries in speech. Indeed, corpus studies showed that such a segmentation strategy would be highly efficient at locating actual lexical word boundaries (Cutler and Carter, 1987). Note that grammatical (or function) word boundaries would not be detected by such a strategy, since by far the majority of grammatical words in English speech are weak monosyllables; this distinction between word classes may, however, be a useful further byproduct of listeners' exploitation of the strong/weak syllable distinction (Cutler, 1993).

In Taft (1984) a stress-based definition of strong versus weak syllables was assumed, whereas Cutler and Norris (1988) and Cutler and Butterfield (1992) adopted a vowel-based definition. However, none of these experiments actually addressed the issue of how syllables containing unstressed unreduced vowels are perceived, and in fact either definition is compatible with any of the results. Not only do the results not enable us to decide whether the distinction which listeners were drawing is based on stress or on vowel quality, they also do not enable us to decide whether or not it is an exhaustively categorical distinction, or a more continuous one.

The present study was designed to shed further light on this issue. Since the crucial case is provided by unstressed unreduced vowels (hereafter, U vowels) such as that in the first syllable of *automata*, these vowels were explicitly included in the study. A production experiment first assessed the acoustic characteristics of U vowels in comparison to stressed vowels and reduced vowels produced by the same range of speakers in the same range of contexts at the same two speech rates. A subsequent perceptual experiment asked whether listeners would treat U vowels as more like reduced vowels (a stress-based distinction), as more like stressed vowels (a vowel-based distinction), or as a true intermediate category, by assessing with which other vowels U vowels were perceived to be freely interchangeable.

I. PRODUCTION STUDY

A. Method

1. Materials

Five sets of four words were constructed, each set having one word each with an initial syllable bearing (1) primary stress, (2) secondary stress, (3) no stress, but with an unreduced vowel, and (4) no stress with a reduced vowel. The four vowel types of these syllables will be referred to below as P, S, U, and R, respectively. The phonetic context following the vowel was the same in each word in a set. Finding sets of vowel-initial words in English with the same vowel in P, S, and U realizations, and in addition with the same following consonant, proved far from easy. We succeeded in constructing five usable sets, although one of these began with a glide-plus-vowel: [ju]. The sets were (i) *autumn, automation, automata, atomic*, (ii) *authorize, authorization, authentic, authority*, (iii) *audiences, auditoria, audition, addition*, (iv) *idle, ideology, idolatry, adoption*, and (v) *unity, unification, united, y'know*. The sets were compiled on the basis of the vowels and stress patterns listed in Jones (1958) and Wells (1990); all the U vowels were listed in those sources with full vowel quality.

A meaningful context was constructed for each word. The contexts were designed as natural occurrences of each word in a focused position, with syntactic and phonetic context controlled: to this end, the critical word occurred in each context after the word *but*, in the beginning of a second clause. The full set of contexts is listed in the Appendix.

Three sets of four distractor sentences were also constructed. These were also two-clause sentences linked with *but*, and each set was constructed around four related words, e.g., *shift, shifting, shifty, shiftless*; the distractor sentences are also listed in the Appendix. Each of the 20 experimental sentences and 12 distractor sentences was typed onto an index card.

2. Subjects

Twelve subjects participated voluntarily in the production experiment. All were students at Cambridge University. Four were speakers of standard southern British English,[1] four were speakers of American English (with no marked regional accent), and four were speakers of Scottish English. Of each group of four, two speakers were female and two male.

3. Procedure

Recordings were made in a sound-damped booth, using an Ampex microphone and a Revox B77 tape recorder. The microphone was adjusted to be 6 in. from the speaker's mouth, and subjects were instructed to maintain this speaking distance. In the interests of maintaining a natural speaking style, no physical constraints were applied; however, the subjects' position was monitored.[2] The cards with the sentences were presented to the subjects in pseudo-random order (constrained such that no two instances from the same set occurred in succession, and that the first three sentences were always distractors). A different order was used for each subject. Each subject first read each sentence three times at a normal speaking rate. Once all 32 sentences had been read, the subjects were asked to read the 32 again, again three times each, speaking at as fast a rate as they could comfortably manage.

The second of the three recordings of each sentence at each rate by each subject was selected for analysis (except in a few instances in which the subject had stumbled over a word, or rustled the cards in the background, during the second version; in those cases the more fluent of the other two versions was selected). Twelve speakers, 20 experimental sentences, and two rates of speech produced 480 utterances for analysis. We analyzed the vowels' duration, $F0$ height, $F0$ movement, intensity, and spectral characteristics, all of which should vary as a function of stress.

The utterances were digitized at 12 bits using a sampling rate of 10 kHz, and stored on disk. The CAMSED speech editing system was used to determine the onset and offset of each of the critical vowels on the digitized waveform. For

word set 5 the prevocalic glide was included in the vowel measurement. The duration of each vowel was recorded in milliseconds.

The $F0$ of each vowel was measured, using the Schäfer-Vincent (1982, 1983) algorithm; this algorithm determines, in several stages, the most likely location of quasiperiodic portions of the waveform. The stages include analysis of the periodic structure of an amplitude-against-time representation of the waveform; the reciprocal of the distance between successive pitch periods is used as an instantaneous measure of the frequency of $F0$. Voicing is determined by assessing adjacent measures for similarity and combining those which pass the similarity tests (Schäfer-Vincent, 1983, pp. 180–183) into a period chain. The output of the algorithm therefore is a sequence of $F0$ values for voiced segments, one value corresponding to each pitch period. Thus the number of $F0$ values calculated for any vowel varies according to the length of the vowel and its $F0$. Within each vowel, the mean and standard deviation of the $F0$ values were calculated. The mean $F0$ value across each vowel gives an estimate of relative $F0$ *height*; the standard deviation of this mean for each vowel gives an estimate of relative $F0$ *movement* on the vowel. We analyzed each $F0$ measure separately.

The algorithm failed to calculate values for a small proportion of the vowels. In these cases the missing data point was replaced by the average across subjects for that condition.

We also measured the intensity of the vowels. For the intensity measurements, we used an algorithm which calculated the average power in the signal across the duration of each analyzed vowel, and expressed the values on an arbitrary decibel scale (on which 0 dB corresponded to a sine wave with intensity equal to one minimum quantum of input voltage).

Finally, we assessed spectral characteristics of each vowel segment. In choosing which spectral characteristics to analyze we sought a simple measure on which stressed and reduced vowels might be expected to differ, which would then allow us to determine whether U vowels resembled stressed vowels, reduced vowels, or neither; we also sought a measure on which we could meaningfully conduct statistical analyses in the same manner as on the duration, $F0$, and intensity measures. We were further aware that in vowels extracted from continuous speech, as opposed to vowels spoken in isolation, $F3$ is often hard to track (Koopmans–van Beinum, 1980). We therefore decided to analyze the *difference* between $F1$ and $F2$ (expressed on a log scale) for each analyzed vowel. Effectively this produces larger values for those vowels with high $F2$ and smaller values for those vowels with low $F2$; reduction, or centralization, would be expressed as a decrease in large values but an increase in small values. In practice, for the vowels used in the present study increasing centralization would consistently show up as an increase in the $F1/F2$ difference. Speech formant trajectories and related information were computed using the Entropic signal processing system (ESPS 4) formant program and a sampling rate of 10 kHz. The program performed a 12th-order linear predictive analysis using the autocorrelation method, with a window duration of 0.049 s. The program default value for frame duration was 0.01 s. For each frame the $F1$ and $F2$ values were extracted and converted to log values. The difference between each pair of log values was computed, and a mean across these differences was then determined for each vowel.

Separate analyses of variance were conducted on the duration, mean $F0$, $F0$ standard deviation, intensity, and spectral measures. Since the five word sets contained different vowels which would exhibit intrinsic differences on at least some of our measures, we did not feel it appropriate to average across word sets. Accordingly, each analysis was conducted with only subjects as random factor. The independent variables were dialect, speaker sex, speech rate, vowel type, and word set. Where appropriate, *post hoc* analyses of significant differences between means were conducted using the modified studentized range statistic q (the Newman–Keuls method; Winer, 1972).

B. Results

1. Duration

The average duration of vowel segments spoken at normal rate was 126 ms, at fast rate 103 ms. This difference was statistically significant ($F[1,6]=16.64$, $p<0.01$). This confirms that speakers responded appropriately to the instruction to speak faster. There were no differences in mean duration as a function of dialect or of speaker sex, nor were the interactions of these factors with one another or with speech rate statistically significant. The mean values agree well with those reported for American English by Crystal and House (1988).

The average duration of P vowels was 148, of S vowels 129, of U vowels 104, and of R vowels 78 ms, again a significant difference ($F[3,18]=177.93$, $p<0.001$). *Post hoc* comparisons revealed that all four vowel types were significantly different from each other in duration. The vowel type effect interacted with speech rate ($F[3,18]=4.8$, $p<0.02$), although *post hoc* comparisons showed all four vowel types to be significantly different from each other at both rates. However, while the vowel durations were sequentially ordered P-S-U-R for both speech rates, the difference between S and U was the smallest intertype difference at normal rate, but the largest intertype difference at fast rate. This interaction is depicted in Fig. 1.

Vowel type also interacted with dialect ($F[6,18]=4.73$, $p<0.01$); although in fact all differences between vowel type were significant for all three dialect groups, the largest intertype difference for English and Scottish speakers was that between S and U vowels, whereas the largest intertype difference for American speakers was between U and R.

The five word sets differed significantly in vowel duration ($F[4,24]=10.01$, $p<0.001$). *Post hoc* analyses revealed (predictably) that the three sets which had the same stressed vowel ([ɔ]) did not differ among themselves at 114, 110, and 110 ms for sets 1, 2, and 3, respectively. The average durations for set 5 (with [ju]) were somewhat shorter at 105 ms, and the average durations for set 4 (with the diphthong [ai]) were significantly longer at 133 ms. The word set variable interacted significantly with speech rate ($F[4,24]$

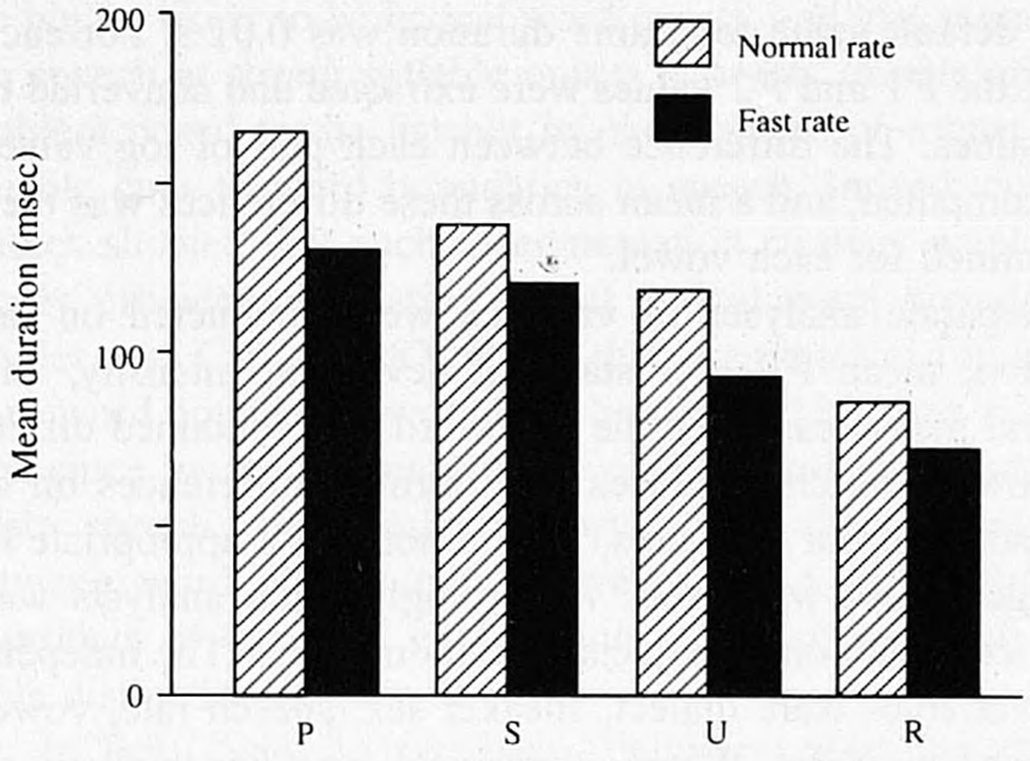

FIG. 1. Mean durations (across subjects and word sets) of the four vowel types in milliseconds, at normal and at fast speech rates.

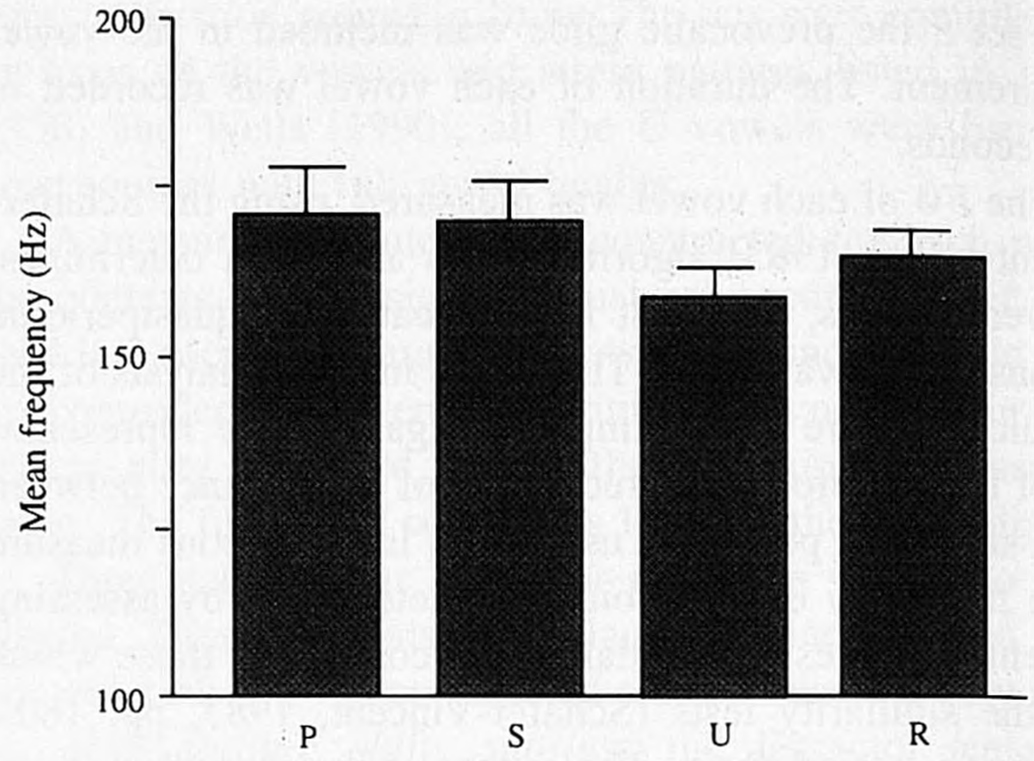

FIG. 2. Mean (across subjects and word sets) of the two $F0$ measures for the four vowel types in hertz: mean $F0$ across the vowel (main bars) and standard deviation of that mean (subsidiary bars).

$=6.41$, $p<0.01$); however, at both speech rates the relations between sets were as just described (set 5 vowels shortest, sets 1, 2, and 3 intermediate, set 4 longest). The source of the interaction was a smaller percentage reduction in duration from normal to fast rate in set 5 than in the other sets. A three-way interaction between dialect, word sets, and speech rate showed that in fact this small percentage reduction for set 5 was true only of English and Scottish speakers; Americans speeded up as much in set 5 as in the other four sets.

Finally, vowel type interacted with the word set variable ($F[12,72]=5.4$, $p<0.001$). Pairwise comparison of vowels for each word set separately produced statistically significant differences *except* in three instances, namely the comparisons between P and S vowels on set 3 only, between S and U vowels on set 1 only, and between U and R vowels on set 2 only; in these three cases the differences were nonsignificant. However, the P-S-U-R ordering of the durational values was maintained for all five sets. A three-way interaction between set, vowel type, and speech rate showed that the similarity in duration of U and R vowels on set 2 was in fact true of fast speech only.

In summary, with respect to our main question, namely the status of U vowels, the durational measures suggest that they form in general a true intermediate case, significantly different from stressed vowels and reduced vowels.

2. F0

The mean $F0$ of male vowels was significantly lower at 124 Hz than the mean $F0$ of female vowels at 209 Hz ($F[1,6]=97.87$, $p<0.001$). Female vowels also exhibited significantly more $F0$ movement (average s.d. of mean $F0$ 6.76 Hz) than male vowels (4.18 Hz; $F[1,6]=11.24$, $p<0.02$). The average $F0$ for each vowel accorded well with the values for the same vowels given in Lehiste (1970).

There was a main effect of dialect on the mean $F0$ measure: our American speakers' voices (male mean 103, female mean 189 Hz) were lower than the voices of our speakers from England (male mean 126, female mean 202 Hz) and Scotland (male mean 141, female mean 235 Hz). Given the very small sample of speakers, we hesitate to draw general conclusions from this finding.

The vowel-type effect was significant on both $F0$ measures ($F[3,18]=3.84$, $p<0.03$ for mean $F0$, $F[3,18]=6.37$, $p<0.01$ for $F0$ s.d.). *Post hoc* analyses showed that P and S vowels did not differ significantly from one another on either measure; U and R vowels differed significantly from one another and from P and S on mean $F0$, but did not differ significantly from one another or from S on $F0$ s.d. The most noticeable feature of the vowel-type effect, however, was the ordering of the four types, which on both measures was P-S-R-U; Fig. 2 depicts this result.

There were no effects of speech rate on either measure and no relevant interactions. The word set variable had a significant effect on the s.d. analysis only ($F[4,24]=4.6$, $p<0.01$); *post hoc* analyses revealed that set 5 vowels ([ju]) displayed more $F0$ movement than vowels from the other four sets.

In summary, although U vowels cannot in this instance accurately be called an intermediate case, the $F0$ analyses did again reveal that they cannot be grouped either with S or with R vowels, since they differed significantly from both on mean $F0$, and were indistinguishable from both simultaneously on $F0$ s.d.

3. Intensity

There were no main effects of dialect, speech rate, or speaker sex in this analysis. The only significant interaction of any kind was one between dialect and speaker sex ($F[2,6]=13.36$, $p<0.01$): female speakers' vowels were more intense for the English and American groups, male speakers' more intense for the Scottish group. The average intensity for each vowel agreed well with previous measures on another large corpus in our own laboratory using the same algorithm (Cutler and Butterfield, 1991).

There were significant differences between the vowel types ($F[3,18]=51.5$, $p<0.001$); *post hoc* analyses showed that P and S vowels were not significantly different,

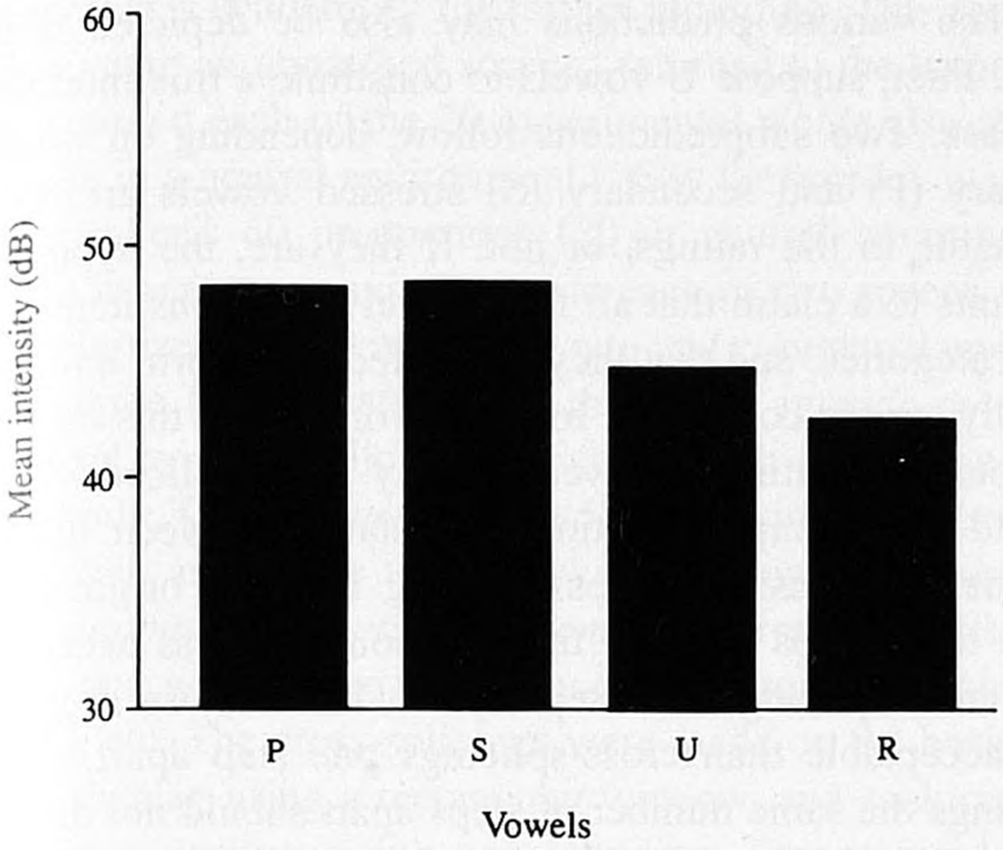

FIG. 3. Mean intensity (across subjects and word sets) of the four vowel types in decibels.

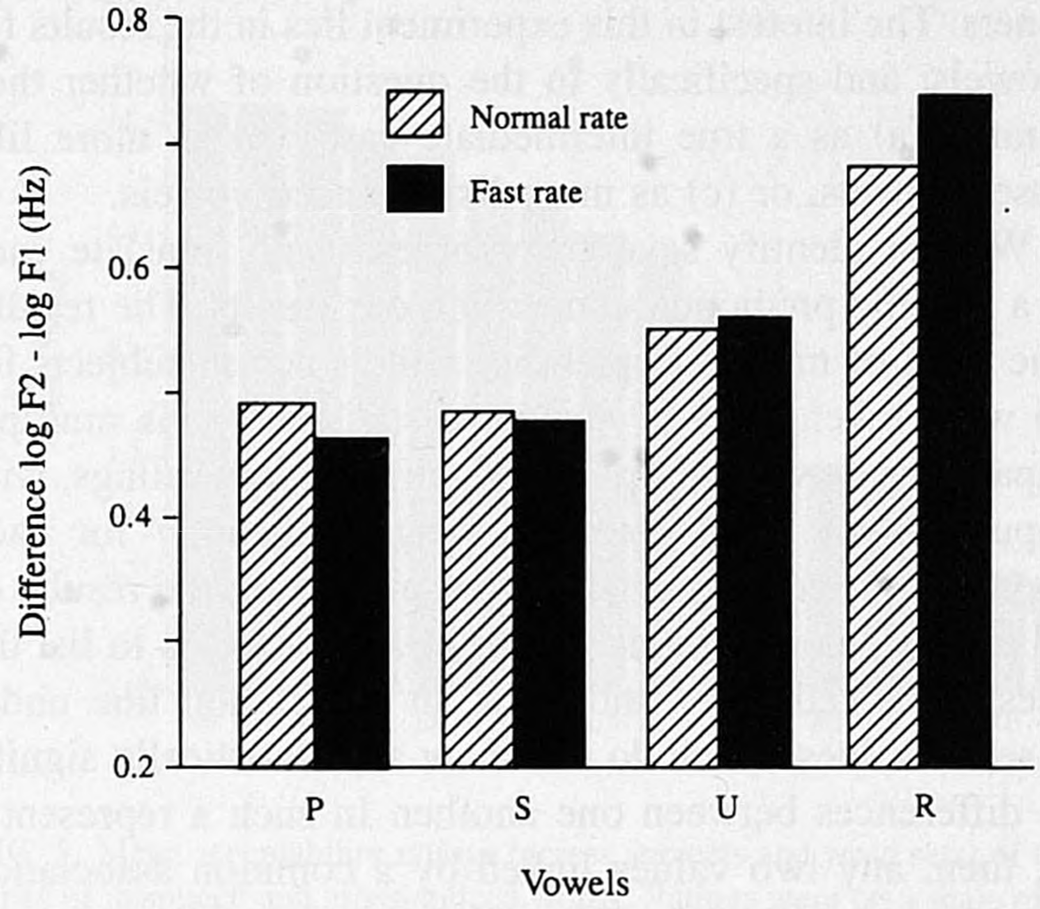

FIG. 4. Mean log $F2$-log $F1$ difference (across subjects and word sets) of the four vowel types in hertz, at normal and at fast speech rates.

but both were significantly more intense than U vowels which in turn were significantly more intense than R vowels. This result is depicted in Fig. 3.

There was also a significant difference between word sets ($F[4,24]=8.39$, $p<0.001$): *post hoc* analyses showed that the two most intense sets of vowels (sets 3 and 4) were not significantly different, but the three other sets differed significantly from these two and from each other. Since this factor did not interact with any other factor (e.g., any speaker-related factor such as sex or dialect), it is difficult to assign it a meaningful interpretation.

On the intensity measure, to summarize, U vowels pattern as an intermediate case once again, significantly different from both stressed vowels and reduced vowels.

4. Spectral characteristics

There were no effects of speaker sex, dialect, or speech rate, and no interaction between these factors. The values obtained for the stressed and reduced vowels were in agreement with those reported elsewhere for such tokens, in British English (Howell and Williams, 1992), American English (Fourakis, 1991), and Dutch (van Son and Pols, 1990).

Vowel types differed significantly on this measure ($F[3,18]=62.92$, $p<0.001$); *post hoc* analyses revealed that P and S vowels did not differ significantly from one another, but U and R vowels differed significantly from one another and from P and S vowels. The ordering, from least to greatest difference, was P-S-U-R.

Vowel type also interacted significantly with speech rate ($F[3,18]=3.75$, $p<0.03$). Figure 4 shows this interaction. The change from normal to fast speech rate reduces the $F1/F2$ difference for P and S vowels, but increases it (indicating increasing centralization) for U and R vowels. *Post hoc* analyses revealed that all four vowels differed significantly from each other at fast rate, but P and S vowels did not differ significantly at normal rate.

There was a main effect of word sets ($F[4,24]=183.74$, $p<0.001$); this was hardly surprising given that the sets involved tokens of different vowels. In a separate analysis on the three sets in which the stressed vowel was [ɔ], the above ordering of vowel types was maintained and the main effect of vowel type was again significant, with all four vowel types differing significantly from one another on *post hoc* analyses. However, the interaction between speech rate and vowel type did not reach significance in this subanalysis.

In summary, on spectral measures U vowels again did not pattern like stressed vowels or like reduced vowels, but were significantly different from both.

5. Conclusion

The four measures we have taken are in striking agreement with regard to the status of U vowels. These are neither like P and S vowels, nor like R vowels. On three of the measures—duration, intensity, spectral characteristics—the U vowels occupy an intermediate position between the stressed and the reduced vowels, significantly different from both. This pattern holds true irrespective of speaker sex and across the three dialects from which we sampled. Also, importantly, it is equally true at normal and at fast rates of speech. U vowels are therefore an intermediate case between stressed and reduced vowels, at least as far as speakers' productions are concerned. In our next investigation we turned to their status in perception.

II. PERCEPTION STUDY

The perceptual aspect of our investigation addressed the issue of how listeners treat the four vowel types, and in particular whether U vowels are perceived as more like stressed vowels or as more like reduced vowels. The method we chose was to interchange the vowels within each set by cross-splicing, and to collect listeners' ratings of the interchanged versions.

Some results of this procedure are of course entirely predictable. It is obvious that cross-splicing the initial vowels of, say, *autumn* and *atomic* is going to produce deviant and unacceptable results. Likewise, it is entirely obvious that the original unspliced words will prove the most acceptable to

listeners. The interest in this experiment lies in the results for U vowels, and specifically in the question of whether they are rated (a) as a true intermediate case, (b) as more like stressed vowels, or (c) as more like reduced vowels.

We can identify several hypotheses and translate each into a specific prediction concerning our results. The results, in the form of mean acceptability ratings across subjects for each word token, may be evaluated statistically via multiple comparisons between all possible pairs of mean ratings, with computation of the studentized range statistic q for each comparison. A conventional way of presenting the results of such multiple comparisons (Winer, 1972, p. 84) is to list the values in ranked order and draw an association line under any set of values which do not show any statistically significant differences between one another. In such a representation, then, any two values linked by a common association line are not significantly different; any two values which do not share a common association line are significantly different.

The various predictions may also be depicted in this form. First, suppose U vowels to constitute a true intermediate case. Two subpredictions follow, depending on whether primary (P) and secondary (S) stressed vowels are distinguishable in the ratings, or not. If they are, the hypothesis amounts to a claim that all four vowel types constitute separate categories, and that they thus effectively form a roughly equally spaced continuum for each word set. In this case, the acceptability rating received by any cross-spliced version should be a simple function of distance between the two original versions: any cross-splicing between original versions three steps apart in the set should be less acceptable than cross-splicings two steps apart which in turn should be less acceptable than cross-splicings one step apart. Cross-splicings the same number of steps apart should not differ in acceptability. This "continuity hypothesis" thus predicts an order defined by number of steps between vowel and body in a P-S-U-R continuum, and four statistically defined groupings among the values:

P-P S-S U-U R-R P-S S-P S-U U-S U-R R-U P-U U-P R-S S-R R-P P-R

If, on the other hand, P and S vowels do *not* differ, there are effectively three separate categories of vowel: stressed vowels (P and S), unreduced unstressed vowels (U), and reduced vowels (R). This, the most plausible "intermediacy hypothesis," predicts that cross-splicings between P and S should be perfectly acceptable, while cross-splicings across category boundaries should be less acceptable if they cross two such boundaries (P and R, S and R) than if they cross just one (P and U, S and U, U and R). Thus we might depict that prediction with much the same ordering as above, but with only three statistically defined groupings:

P-P S-S U-U R-R P-S S-P S-U U-S U-R R-U P-U U-P R-S S-R R-P P-R

Two further hypotheses assert that there will be a single category boundary within the set. In such a case, any cross-splicing which crosses the category boundary should be less acceptable than any cross-splicing which does not. A stress-based definition of the strong–weak difference predicts a category boundary between the two stressed vowels and the two unstressed vowels. In other words, this "stress-based categoricality" hypothesis predicts that P and S vowels should be freely interchangeable and that U and R vowels should be freely interchangeable, but that cross-splicings between these groups should be unacceptable. Hence the category boundary should fall as follows:

P-P S-S U-U R-R P-S S-P S-U U-S P-U U-P U-R R-U R-S S-R R-P P-R

Finally, a vowel-based definition of the strong/weak distinction predicts a category boundary between the three unreduced vowels and the one reduced vowel. Thus P, S, and U vowels should be freely interchangeable but cross-splicings between any of these and R vowels should be unacceptable. Hence a single category boundary should be observed between R vowels and all others:

P-P S-S U-U R-R P-S S-P S-U U-S P-U U-P U-R R-U R-S S-R R-P P-R

All four identifiable hypotheses, as can be seen, predict (reasonably) that the original versions of the words will receive the highest acceptability ratings and the cross-splicings between stressed and reduced vowels the lowest. The difference between the hypotheses lies solely in the relative ordering of cross-spliced versions involving U vowels, and in the statistical pattern of association between means.

A. Method

1. Materials

Since in the production experiment the words had been recorded in meaningful contexts, the acceptability rating received by any cross-spliced version of these tokens amounts to a rating of how closely it resembled the original word. However, it is also worth asking whether a cross-spliced word is acceptable as a possible word. To this end, we added to our existing meaningful contexts, spoken at normal and at fast rates, a third environment: a (normal rate) neutral context.

The utterances produced by one of the standard southern British English speakers were selected from the production experiment corpus. The selection was made prior to performance of the acoustic analyses and was determined solely by

the speaker's availability for further recording. This speaker, a male, with an unmarked accent,[1] returned to the laboratory and recorded each of the 20 experimental words also at normal rate in a neutral environment ("Say the word ... again"). The speaker's 60 productions (20 in neutral environment, and 20 in sentence environment at each of two speech rates) were digitized, and, within each rate and contextual environment, three further versions of the words in each set were produced by cross-splicing each vowel with each decapitated word body. Thus *autumn* occurred in a version with its original vowel, as well as versions with the vowels of *automation, automata*, and *atomic; atomic* occurred with its own vowel and with the vowels of *autumn, automation*, and *automata*, etc. The cross-splicings were made on the basis of a visual display using a rectangular window, and each cut was made at a positive-going zero crossing. The manipulations resulted in a total of 240 tokens: 5 sets ×4 words ×3 environments ×4 versions.

An experiment containing all 240 stimuli, and hence lasting about an hour and a half, would be very fatiguing for listeners; therefore three separate tapes were made, each tape containing one-third of the experimental stimuli. Every tape contained all 16 original and cross-spliced words, in context, in one environment for each set. All five sets occurred on each tape, and all three environments were represented at least once on each tape. (One tape contained, for example, the *autumn* and *unity* sets in the normal rate sentence environment, the *audiences* and *idle* sets in the fast rate sentence environment, and the *authorize* set in the neutral environment.) A practice set of words (*upper, upset, appeal*) was recorded in similar sentence and neutral environments, at normal and at fast rate, and a few cross-spliced versions were made to provide a small set of practice examples which was recorded at the beginning of each tape.

2. Subjects

Twenty-four listeners took part in the experiment for a small payment. They were students and staff of Magdalene College, Cambridge, or students of other Cambridge colleges; all were native speakers of British English. None had participated in the production experiment.

3. Procedure

The subjects were tested at Magdalene College in groups of four, two such groups hearing each tape. The sentences were presented over Sennheiser headphones from a Revox B77 tape recorder. The subjects were given written instructions to rate the naturalness of each critical word on a scale from 1 to 5, with a rating of 1 signifying that the pronunciation of the word on the tape could not be recognized as a contextually appropriate word, and a rating of 5 signifying that the pronunciation on the tape was appropriate and exactly as would be expected for the perceived word. The experiment lasted approximately 25 min.

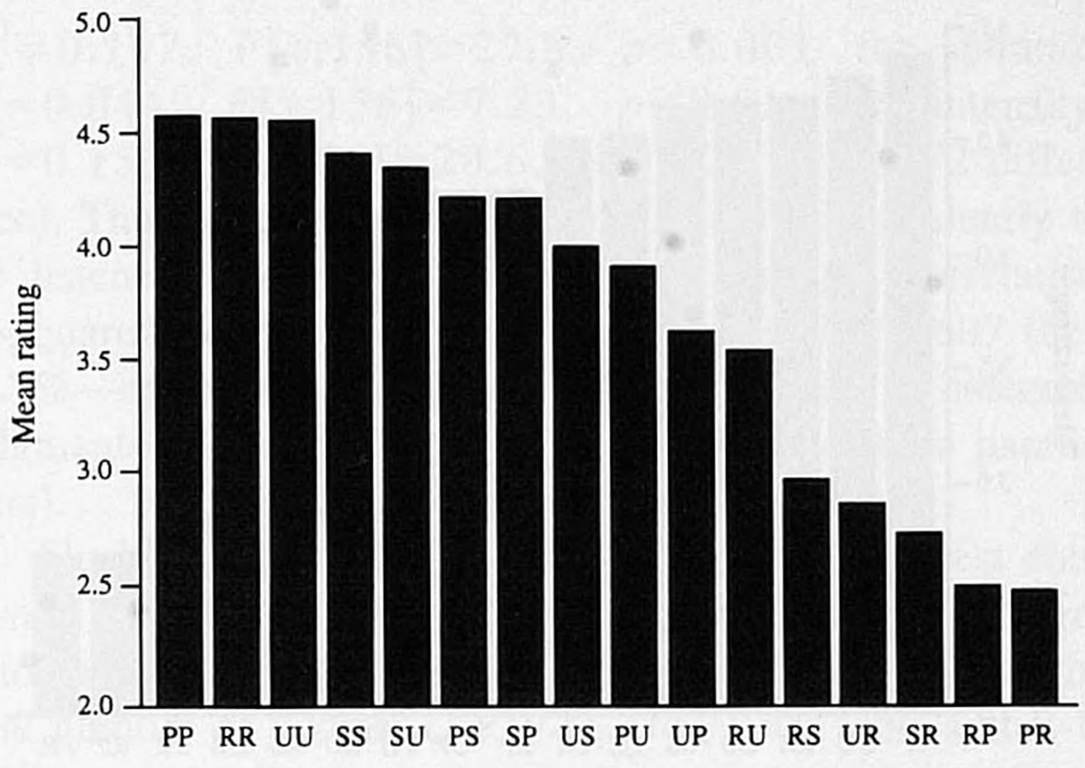

FIG. 5. Mean acceptability ratings (across subjects and word sets) of the 16 types of unspliced and cross-spliced words. Ratings were on a scale of 1–5, with 5 signifying maximum acceptability. The two-letter code for each word signifies first the presented vowel and second the original vowel in that word body; thus UP is an unstressed unreduced vowel spliced into a word which originally had primary stress on the vowel.

B. Results

1. Acceptability ratings

Mean acceptability ratings were computed for each version of each word in each environment. The mean ratings for the 16 word types (four words with four vowels), averaged across the five word sets and the three environments, are shown in Fig. 5. They are ranked in order of rated acceptability (recall that 5 signifies most acceptable, 1 least acceptable).

The multiple (in fact, 120) comparisons between the means averaged over all continua (i.e., the data in Fig. 5) showed a pattern of statistical significance which (as described above) can be represented as follows:

P-P R-R U-U S-S S-U P-S S-P U-S P-U U-P R-U R-S U-R S-R R-P P-R

This is certainly not the statistical pattern predicted by the continuity hypothesis. In fact, Fig. 5 shows that the different combinations do not rank in the order predicted by the continuity hypothesis. Specifically, combinations of primary stress and unreduced zero stress (P-U and U-P, two steps apart) are ranked higher than would be expected, while combinations of reduced and unreduced zero stress (U-R and R-U, one step apart) are ranked lower than expected. Principally, however, the continuity hypothesis cannot predict the observed statistical grouping between original versions and

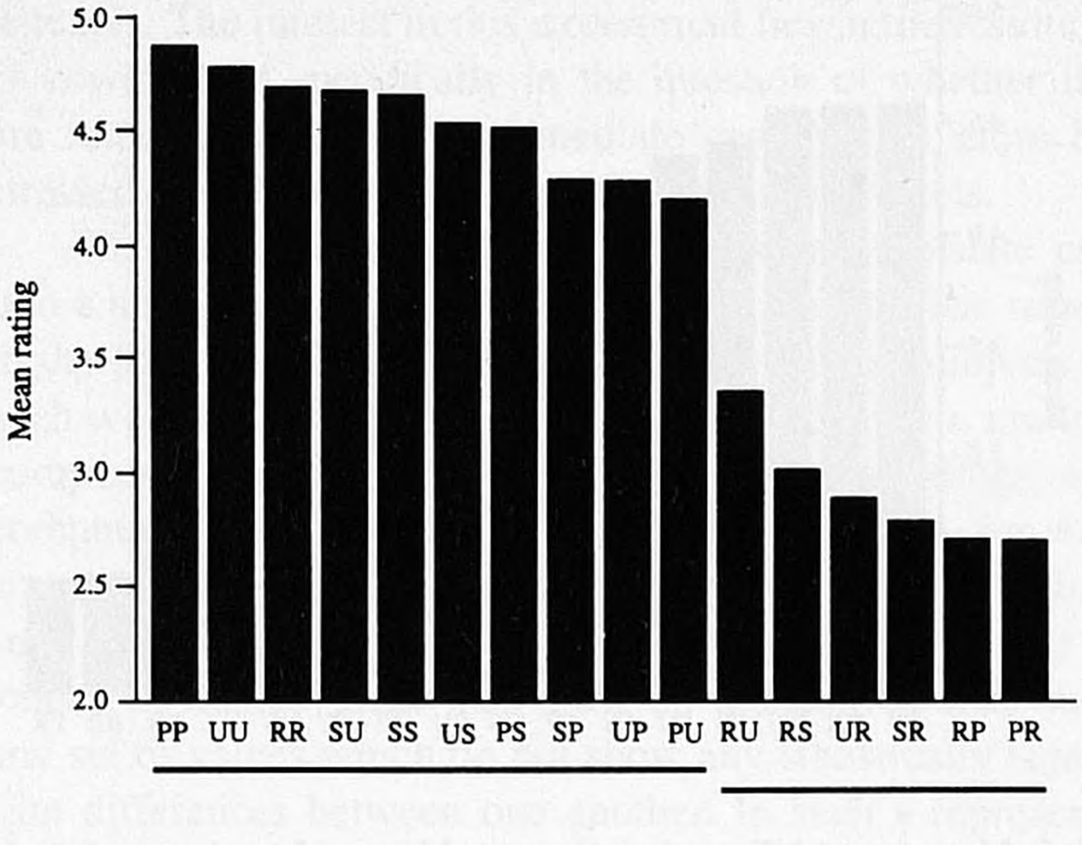

FIG. 6. Mean acceptability ratings (across subjects and word sets) of the 16 types of unspliced and cross-spliced words, presented in the neutral context "Say the word ... again." Ratings were on a scale of 1–5, with 5 signifying maximum acceptability. Ratings for word types linked by underlining did not differ statistically.

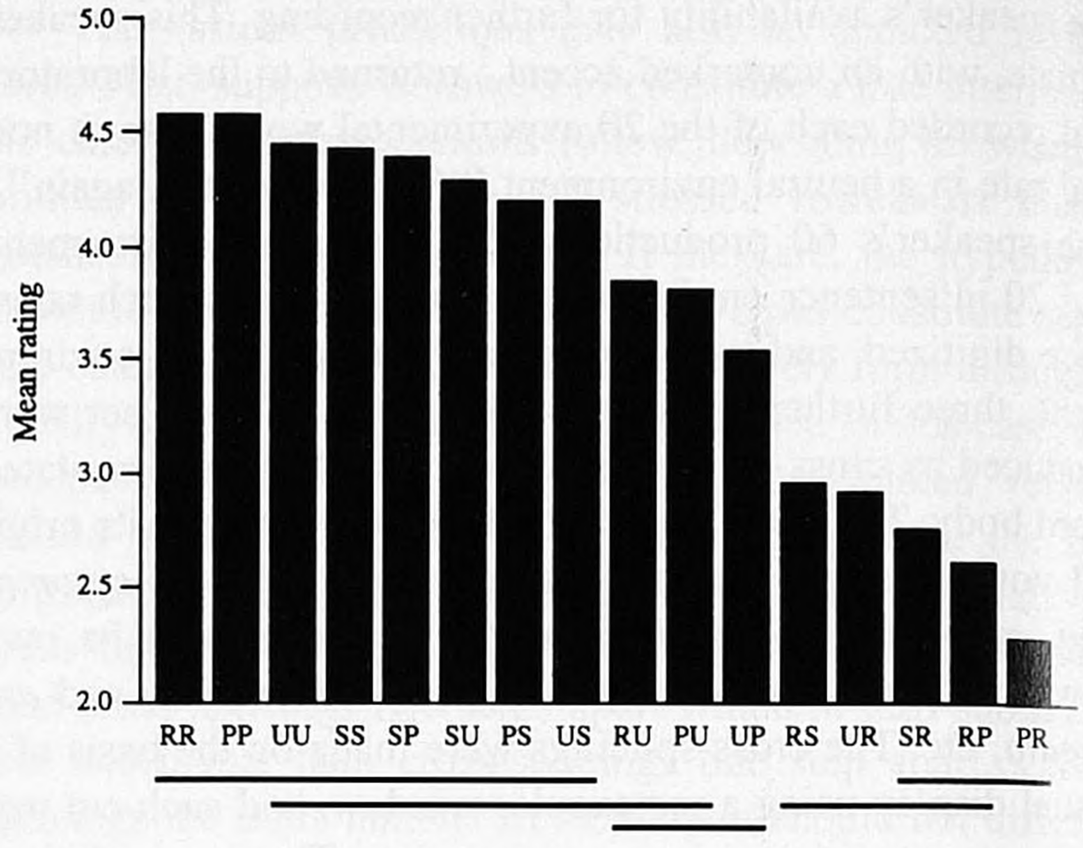

FIG. 7. Mean acceptability ratings (across subjects and word sets) of the 16 types of unspliced and cross-spliced words, presented in the meaningful contexts listed in the Appendix, at a normal rate of speech. Ratings were on a scale of 1–5, with 5 signifying maximum acceptability. Ratings for word types linked by underlining did not differ statistically.

cross-splicings both one and two steps apart. The pattern is also not that predicted by the intermediacy hypothesis; again, the observed grouping of original versions and cross-spliced versions involving stressed and U vowels disconfirms this hypothesis. Finally, the pattern is also not that predicted by the stress-based categoricality hypothesis, for the same reason.

The observed pattern is closest to—although again, not identical with—the pattern predicted by the vowel-based categoricality hypothesis. The ordering is as predicted by that hypothesis in that all cross-splicings involving R are ranked lower than all cross-splicings not involving R. The principal difference from the vowel-based categoricality hypothesis' predictions lies in the statistical indistinguishability of the R-U mean from any other mean.

A clearer pattern emerged when we analyzed each environment separately. The results of these analyses, including their statistical association patterns, are presented in Figs. 6–8.

The neutral context is of particular interest because it offers an index of the simple acceptability of given combinations of vowel and word body, irrespective of whether a particular meaning is intended. The order is similar to the ordering for all contexts, with P-U and U-P being ranked higher than U-R and R-U, and there is a clear break in the rankings, with the gap between P-U and R-U being much larger than all other gaps between adjacent versions. Most noticeably, the acceptability ratings for the cross-spliced words *not* involving schwa do not differ significantly either from each other or from the ratings for unaltered words. This pattern is in fact exactly as predicted by the vowel-based categoricality hypothesis.

In the meaningful context, both at normal and at fast rate, there is just one deviation from the ordering predicted by a vowel-based categoricality hypothesis: R-U words are rated higher than would be predicted. A reduced vowel substituting for an unreduced unstressed vowel is thus reasonably acceptable in a meaningful context (though it is not acceptable in a neutral, i.e., citation-form context). A hypothesis which distinguishes between two categories of vowel on the basis of vowel quality would therefore seem to achieve closer approximation to the acceptability results than any other hypothesis.

2. Correlations of acceptability and acoustic factors

To shed light on the basis for the observed pattern of acceptability judgments, we carried out correlation analyses between the mean acceptability ratings for each token and the acoustic properties of that token. The additional tokens (neutral environment) produced for the perceptual study by the speaker used here were subjected to the same analyses described in the account of the production study above. Only the cross-spliced tokens were included in the correlation

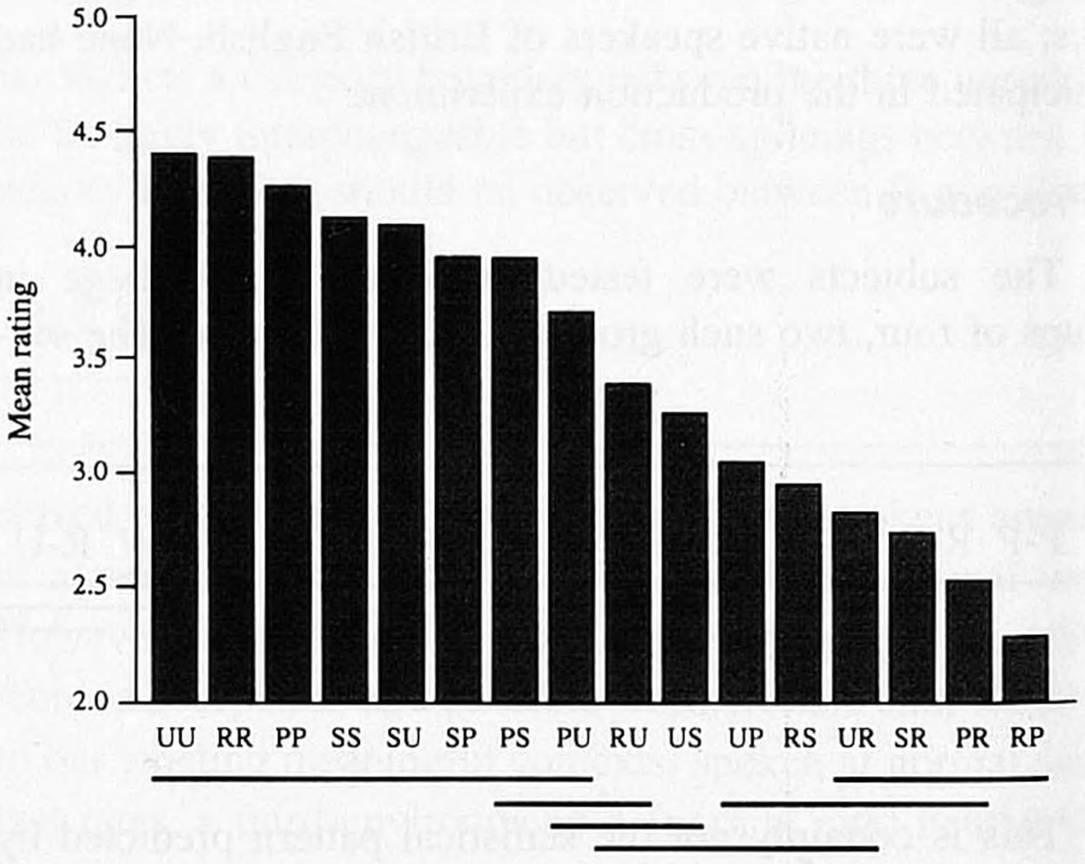

FIG. 8. Mean acceptability ratings (across subjects and word sets) of the 16 types of unspliced and cross-spliced words, presented in the meaningful contexts listed in the Appendix, at a fast rate of speech. Ratings were on a scale of 1–5, with 5 signifying maximum acceptability. Ratings for word types linked by underlining did not differ statistically.

analysis. Six utterances on which one of the measures returned an outlying data value were dropped from these analyses. For each cross-spliced token, a mismatch score was computed on each of the five acoustic dimensions we measured (duration, mean $F0$, $F0$ standard deviation, intensity, $F1/F2$ difference). The mismatch score was the unsigned difference on any measure between the original vowel and its replacement in the cross-spliced word.

If listeners' acceptability ratings are affected by a particular dimension of variation, then we can predict that the larger a token's mismatch score (i.e., the larger the deviation on that dimension between the original vowel and its replacement), the lower will be the rated acceptability of that token. Hence a significant negative correlation between rated acceptability and mismatch scores on any acoustic dimension will signify that that dimension is relevant for listeners' judgments of the vowels.

Across all 174 utterances (5 word sets×4 words×3 cross-spliced versions×3 environments=180, minus 6 outliers), only the $F0$ measures showed no relationship with the acceptability ratings. All three of the other measures produced significant negative correlations: $r[173]=-0.42$, $p<0.001$ for duration, $r[173]=-0.47$, $p<0.001$ for intensity, and, the strongest correlation, $r[173]=-0.53$, $p<0.001$ for $F1/F2$ difference. When the three speech environments were considered separately, correlations with $F1/F2$ difference and intensity were significant for all three environments, while correlations with duration were significant for normal rate meaningful-context tokens and for neutral-context tokens but not for fast rate meaningful-context tokens. Tokens from each word set separately also all showed significant correlations with the acoustic measures: on word sets 4 and 5, there were significant correlations with all three measures, on word set 3 there were correlations with spectrum and intensity, on word set 2 with spectrum and duration, and on word set 1 with intensity alone.

Given that the acceptability ratings showed a difference between, say, R-U and U-R tokens, i.e., between cross-splicings in which a more reduced vowel replaced a less reduced vowel in comparison to the reverse, we reanalyzed the tokens incorporating this dimension: whether the spliced-in vowel replaced a vowel higher or lower in the P-S-U-R hierarchy. However, all three acoustic factors showed significant negative correlations for cross-splicings in both directions, i.e., both for tokens in which a less stressed vowel replaced a more stressed (i.e., R-U, R-S, R-P, U-S, U-P, and S-P) and for tokens in which a more stressed vowel replaced a less stressed (i.e., P-S, P-U, P-R, S-U, S-R, and U-R).

Since the acoustic measures showed significant positive correlations among themselves, a further multiple-regression analysis was performed with the acceptability ratings as the dependent variable, to assess the extent to which each of the three acoustic factors could lay claim to a significant explanatory contribution over and above the others. After including terms for the main effects and interactions of speech environments and word sets, the unique variance (squared multiple semipartial correlation; Cohen and Cohen, 1983) of the three measures was calculated. All three were significant ($R^2=0.117$, $F[1,156]=27.8$, $p<0.001$ for duration $R^2=0.044$; $F[1,156]=7.23$, $p<0.01$ for intensity; $R^2=0.151$, $F[1,156]=20.6$, $p<0.001$ for $F1/F2$ difference). Thus all three variables contributed independently to the listeners' judgments. The squared multiple correlation (R-squared) for the three measures combined is 0.607 (i.e., 60.7%—by far the majority—of total variance in listeners' judgments is accounted for by these three acoustic parameters).

Finally, because the principal question of interest concerned the ratings for tokens involving U vowels, the correlation analyses were repeated for these tokens only (half the total number of tokens: U-P, P-U, U-S, S-U, U-R, and R-U only). Again, all three acoustic factors showed significant negative correlations with the acceptability ratings, and again, the strongest correlation was with the spectral measure ($r[87]=-0.3$, $p<0.005$ for duration; $r[87]=-0.38$, $p<0.001$ for intensity; $r[87]=-0.44$, $p<0.001$ for $F1/F2$ difference). Squared multiple semipartial correlations were again significant for all three measures ($R^2=0.182$, $F[1,70]=15.58$; $p<0.001$ for duration $R^2=0.06$, $F[1,70]=4.44$, $p<0.05$ for intensity; $R^2=0.056$, $F[1,70]=4.18$, $p<0.05$ for $F1/F2$ difference).

3. Summary

The perceptual study suggests that U vowels are grouped by listeners rather more consistently with stressed vowels than with reduced vowels; they do not form a clear-cut third, intermediate vowel category. Although listeners use spectral characteristics, duration, and intensity to guide their decisions about vowel tokens, spectral characteristics seem to be given greatest weight.

III. DISCUSSION

This investigation has been aimed at the distinction between syllable types in English, and in particular at the question of whether there are two categories of syllables defined by stress, two categories defined by vowel quality, or a more continuous distribution in which unstressed syllables with unreduced vowels form an intermediate category between stressed syllables and reduced syllables. As described in the Introduction, this question boils down to one about unstressed syllables with unreduced vowels, and it was at these that the present investigation was aimed.

The results of the production study reported here support a continuous distribution of syllable types. On four of the five acoustic dimensions on which we measured the vowel tokens, U vowels were significantly different from *both* reduced vowels and stressed vowels. On the remaining measure they were significantly different from *neither* the reduced nor the stressed vowels. (This one measure on which U vowels failed to differ from the other vowels was the standard deviation of $F0$ values across the vowel, intended as a measure of pitch movement. In fact on this measure, and indeed also on mean $F0$, the U vowels returned lower values than the R vowels. This latter result itself suggests that $F0$ may not be a relevant dimension for the present classification

of syllable types. In any case, it is clear that variation in $F0$ movement across syllable types is unrelated to variation along the remaining dimensions measured.)

Thus there was no indication in the production results that U vowels group systematically either with stressed or with reduced vowels. Their duration, intensity, and spectral quality placed them in an intermediate situation between the stressed P and S vowels (which on most measures did not differ significantly from one another) and the reduced R vowels. The production evidence therefore does not support a binary category distinction between strong and weak syllables based on either stress or vowel quality.

As described in the Introduction, however, there is now abundant empirical evidence for the existence of some dimension along which listeners make a discrimination between (at least) stressed and reduced vowels. Does vowel perception stand in contrast to the acoustic facts, in that listeners make a categorical distinction where speakers produce a more continuous distribution? The present perception study was designed to answer this question by assessing in particular the way in which listeners respond to unstressed unreduced vowels. The results suggested that in citation form at least, these U vowels are grouped firmly with stressed vowels—in other words, listeners draw a binary strong/weak distinction between syllables with full vowels and syllables with reduced vowels.

The statistically indistinguishable interchangeability of P, S, and U vowels in citation form was not fully reproduced when the words were presented in meaningful contexts; here, the results were somewhat more complex. What was noticeable about these results was that reduction was more acceptable to listeners than its converse—i.e., substitution of a less stressed vowel for a more stressed vowel was more acceptable than the reverse substitution. In particular, reduction of U vowels (the R-U case) was relatively acceptable, and certainly more acceptable in context than in citation form. This finding is also compatible with a vowel-based distinction, on which U vowels are properly full in citation form but may be reduced in context. Recall that switching from normal to fast rate of speech in the production study led to a slight increase in centralization for the U and R vowels. In a sense U vowels would therefore be an intermediate case insofar as they can switch categories—but there would still be only two categories in terms of the perception of surface realizations.

The correlations between rated acceptability and the acoustic measures showed that listeners' responses were sensitive to vowel duration, intensity, and spectral characteristics alike. The $F0$ measures showed no relation to acceptability, which is only to be expected given that our vowel tokens patterned less consistently on the $F0$ measures than on the remaining acoustic dimensions. The acceptability of cross-spliced tokens was determined most strongly by spectral characteristics, and, across word sets and rate sets, more strongly by intensity than by duration. (This is consistent with other reports of low sensitivity of English listeners to durational modifications within syllables; Bertinetto and Fowler, 1989.) For U vowels alone, acceptability is again determined most strongly by the appropriateness of spectral characteristics.

In related work, Allerhand *et al.* (1992) reported implementation of a measure of syllable strength based on $F0$ and intensity that was well able to distinguish the strong syllables from the weak syllables in the set of materials used by Cutler and Butterfield (1992). However, relative syllable strength as thus measured did not correlate with relative syllable strength as reflected in the likelihood of segmentation errors involving each syllable in the Cutler and Butterfield data set. Allerhand *et al.* concluded that listeners' differentiation between strong and weak syllables as reflected in their segmentation decisions was not based on the acoustic dimensions incorporated in the $F0$-intensity algorithm for syllable strength measurement. It is clear that these two factors alone would not allow prediction of the present perceptual results either. The perceptual data from the present and from previous studies suggest, indeed, that even though there really is something like continuous variation in syllable types, listeners tend to implement a binary distinction in practice. Since U vowels are somewhere in between the endpoints of this distinction, they will be classified as one or the other depending on how their particular realization is perceived.

Why should listeners make use of a binary distinction between strong and weak syllables? We suggest that listeners will in general prefer to make discriminations which are absolute rather than relational in nature. Absolute judgments can be made immediately; relational judgments require comparison between at least two instances (in this case, two syllables), and hence may involve a delay in making the decision. Studies of spoken word recognition suggest above all that recognition is fast and efficient; recognition decisions are not delayed. If this is indeed the case, then spectral characteristics offer the best basis for an absolute discrimination, on the grounds that category judgments about vowel identity draw upon spectral information. In contrast, duration, intensity, and other prosodic dimensions admit of variation in relational terms only; that is, whether a particular syllable is long or short, loud or soft, and so on can only be judged relative to other syllables. Note that we cannot as yet finally answer the question as to whether listeners prefer absolute to relative information for the strong/weak decision; in the present perceptual study listeners' judgments were significantly correlated with durational and intensity variation as well as with spectral characteristics. All these variables were highly correlated in the present natural materials, and a definitive answer could perhaps only be found via orthogonal manipulation of these dimensions in synthetic materials. However, the results are certainly consistent with a clear role for spectral characteristics.

Given that the vowel quality discrimination is efficient, a further motivation for using it to separate strong vowels from weak as separate categories then arises from the segmentation problem described in the introduction to this paper. Such discrimination helps to solve the word boundary problem, because in English it is the case that by far the majority of lexical words begin with strong syllables. Note that the categorization is not finer grained: listeners do not appear to calculate likely word initialness for each full vowel individually. Cutler and Butterfield (1992) showed that patterns of segmentation errors did not correlate with actual distribution

of onsets in vocabulary for the six full vowels they tested—there was no greater tendency, for instance, towards erroneous word boundary insertions before syllables with [ʌ] than before syllables with [ɛ], despite the fact that [ʌ] is more than 3× as likely as [ɛ] to be found in a word-initial syllable. Cutler and Norris (1988) proposed that categorizing vowels as strong or weak during speech recognition serves the primary purpose of targeting lexical access attempts at those portions of the speech signal which are relatively more likely to be word initial. The Cutler and Butterfield finding suggests that the categorization is made on the broadest possible grounds; reduced vowels are far less likely to be the initial syllables of lexical words in English; therefore syllables containing reduced vowels can simply be consigned to the bottom of a hierarchy of likely points at which lexical access might be attempted. Thus the effect of the categorization is not so much to favor strong syllables as to disfavor weak. The law of the jungle rules even in speech recognition: strong/weak discrimination is effectively discrimination against the weak.

ACKNOWLEDGMENTS

The larger part of this work was completed by the first author, under supervision of the second author, in partial fulfillment of the requirements of a master's degree in Computer Speech and Language Processing at the University of Cambridge. Financial support for one of the authors (B. D. F.) was provided by the Science and Engineering Research Council, U.K. We are grateful to Tom Cooke, Dennis Norris, Brit van Ooyen, Ken Robinson, and Duncan Young for further assistance, to Ian Nimmo-Smith for extensive assistance with statistical analyses, to Inge Doehring for assistance with the figures, and to Francis Nolan, James McQueen, Ann Syrdal, and an anonymous reviewer for helpful comments on the manuscript. A brief report of part of the perceptual study reported here was presented to the ESCA Workshop on Phonetics and Phonology of Speaking Styles, Barcelona, in October 1991. B. D. F. is now with Cameron Markby Hewitt (London). Correspondence should be addressed to the second author at Max-Planck-Institute for Psycholinguistics, P.O. Box 310, 6500 AH Nijmegen, The Netherlands. E-mail:anne@mpi.nl.

APPENDIX

The five sets of sentences used in both production and perception studies (1–5), and the three sets of distractor sentences used in the production study (6–8) follow:

(1) Summer is the time for berries, but autumn is the time for apples.
The factory once employed 80, but automation reduced this by half.
The workers were treated as if they weren't humans, but automata to be programmed.
Armies used to be a country's main defense, but atomic weapons changed all that.

(2) It may be classified, but authorized copies are available.
It may be tedious, but authorization is required.
It may be inspiring, but authentic artwork is expensive.
Rebellion was brewing, but authority was maintained.

(3) The band wanted more engagements, but audiences were hard to find.
Opera may be an art form, but auditoria have to be filled for it to be economic.
She acts well, but auditions fill her with dread.
Subtraction is easy, but addition is even easier.

(4) That boy is quite bright, but idle and insolent.
Civil rights were violated, but ideology somehow justified it.
Admiration is one thing, but idolatry is quite another.
The couple wanted a child, but adoption was out of the question.

(5) Church leaders may be hopeful, but unity is a long way off.
West Germans may grumble, but unification with East Germany will occur.
Palace played well, but United won.
You can try if you like, but y'know it won't work.

(6) To move the table, first shift the glass, but be careful not to spill the wine.
The shiftless youth daydreamed, but somehow the work was done.
Fagin was shifty, but Oliver was simply easily led.
A whole parish was swallowed up by the shifting sands, but still no action could be taken.

(7) The prince rushed to her rescue, but failed to reach the princess.
The geometric principles determined by the Greeks were used for hundreds of years, but then Newton invented calculus.
I can read this book, but the print is very small.
In school physics the spectrum is produced not by reflection, but by refraction through a prism.

(8) The quality was good, but still the material ripped.
He is a poor scholar, but a matchless athlete.
I tried to open the door, but the mat was in the way.
I don't have a match, but I do have a lighter.

[1]Standard southern British English is the unmarked form of educated English as spoken in England, and is the most common form of English used in the English broadcast media. Standard Scottish English fills the same role in Scotland as standard southern British does in England.

[2]It was important to us to ensure that the speakers' productions were, within the limits of the sentence reading task, as natural as possible, and thus no physical constraints were applied. However, there is always under such circumstances a risk that gradual but systematic movement leading to variations in intensity may occur. We explicitly controlled for the possibility of systematic variation introduced by subjects gradually moving closer to or further from the microphone during the reading of a set of sentences by presenting the sentences in a different order for each subject. There is a further possibility, namely that subjects might gradually move closer to or further from the microphone across the entire recording session. This would introduce a main effect of set, i.e., of speech rate, since normal rate productions were always recorded before fast rate productions. In fact, as the results of the intensity analysis below show, there was no such effect, and no interaction involving speech rate.

Allerhand, M., Butterfield, S., Cutler, A., and Patterson, R. (**1992**). "Assessing syllable strength via an auditory model," Proc. Inst. Acoust. (UK) **14**, 297–304.

Bertinetto, P. M., and Fowler, C. A. (**1989**). "On sensitivity to durational modifications in Italian and English," Riv. Linguist. **1**, 69–94.

Bolinger, D. L. (**1981**). *Two Kinds of Vowels, Two Kinds of Rhythm* (Indiana University Linguistics Club, Bloomington).

Cohen, J., and Cohen P. (**1983**). *Applied Multiple Regression/Correlation Analysis in the Behavioral Sciences* (Erlbaum, Hillsdale, NJ), 2nd ed.

Crystal, T. H., and House, A. S. (**1988**). "Segmental durations in connected-speech signals: Syllabic stress," J. Acoust. Soc. Am. **83**, 1574–1585.

Cutler, A. (**1993**). "Phonological cues to open- and closed-class words in the processing of spoken sentences," J. Psycholinguist. Res. **22**, 109–131.

Cutler, A., and Butterfield, S. (**1990**). "Durational cues to word boundaries in clear speech," Speech Commun. **9**, 485–495.

Cutler, A., and Butterfield, S. (**1991**). "Word boundary cues in clear speech: A supplementary report," Speech Commun. **10**, 335–353.

Cutler, A., and Butterfield, S. (**1992**). "Rhythmic cues to speech segmentation: Evidence from juncture misperception," J. Mem. Lang. **31**, 218–236.

Cutler, A., and Carter, D. M. (**1987**). "The predominance of strong initial syllables in the English vocabulary," Comp. Speech Lang. **2**, 133–142.

Cutler, A., and Norris, D. G. (**1988**). "The role of strong syllables in segmentation for lexical access," J. Exp. Psychol.: Hum. Percept. Performance, **14**, 113–121.

Fourakis, M. (**1991**). "Tempo, stress and vowel reduction in American English," J. Acoust. Soc. Am. **90**, 1816–1827.

Halle, M., and Keyser, S. J. (**1971**). *English Stress: Its Form, Its Growth, and Its Role in Verse* (Harper and Row, New York).

Howell, P., and Williams, M. (**1992**). "Acoustic analysis and perception of vowels in children's and teenagers' stuttered speech," J. Acoust. Soc. Am. **91**, 1697–1706.

Jones, D. (**1958**). *Everyman's English Pronouncing Dictionary* (Dent, London), 11th ed.

Koopmans-van Beinum, F. J. (**1980**). *Vowel Contrast Reduction* (Academische Pers, B. V., Amsterdam).

Lehiste, I. (**1970**). *Suprasegmentals* (MIT, Cambridge, MA).

McQueen, J. M., Norris, D. G., and Cutler, A. (**1994**). "Competition in spoken word recognition: Spotting words in other words," J. Exp. Psychol.: Learn. Mem. Cog. **20**, 621–638.

Norris, D. G., McQueen, J. M., and Cutler, A. (**1995**). "Competition and segmentation in spoken word recognition," J. Exp. Psychol.: Learn. Mem. Cog. **21**.

Schäfer-Vincent, K. (**1982**). "Significant points: Pitch period detection as a problem of segmentation," Phonetica **39**, 241–253.

Schäfer-Vincent, K. (**1983**). "Pitch period detection and chaining: Method and evaluation," Phonetica **40**, 177–202.

Son, R. J. J. H. van, and Pols, L. C. W. (**1990**). "Formant frequencies of Dutch vowels in a text, read at normal and fast rate," J. Acoust. Soc. Am. **88**, 1683–1693.

Taft, L. (**1984**). "Prosodic constraints and lexical parsing strategies," Ph.D. dissertation, University of Massachusetts.

Wells, J. C. (**1990**). *Longman Pronunciation Dictionary* (Longman, Harlow).

Winer, B. J. (**1972**). *Statistical Principles in Experimental Design* (McGraw-Hill, New York), 2nd ed.