BRITISH ACOUSTICAL SOCIETY

"SPRING MEETING" at Chelsea College, London, S.W.3. on
Wednesday 25th April / Friday 27th April, 1973.

SPEECH AND HEARING: Session 'C': Speech Properties and Recognition.

Paper No:
73SHC1

RHYTHMICAL AND PHYSIOLOGICAL CONSTRAINTS IN THE PRODUCTION OF SOME ENGLISH SPEECH

Celia Scully

Department of Phonetics, University of Leeds

INTRODUCTION

Vowel duration is one of the acoustic correlates of stress in English (Fry, 1958) and the syllables of English are very uneven in length. Stressed and unstressed syllables tend to alternate and it is claimed that the beat of the stressed syllable is isochronous so that the speech is divided into feet of approximately even length; each foot begins with a stressed syllable (Abercrombie, 1964). This paper aims to demonstrate that, while the durations of words and the prominence of adjacent syllables may be important features of English speech, this is not necessarily dependent upon an isochronous stress beat. Any tendency towards isochrony is so small as to be irrelevant in the perception of these long time intervals, whereas the durations of smaller segments must be carefully controlled by the speaker. Four utterances containing almost identical segmental structures but with differently placed contrastive stresses were elicited from some adult speakers of English, so that the same segments occurred in feet of different durations. A tendency towards isochrony should reveal itself as a shortening of comparable segments in the longer foot as compared with the shorter foot. Isochrony in the sense of a constant foot duration within one utterance does not appear relevant to the problem of generating short utterances such as the ones in this experiment where, often, only two stressed syllables were heard, thus establishing only one complete foot duration.

THE EXPERIMENT

Four different questions were recorded by one speaker. These were: Q1: "Did she shut the shop?"; Q2: "Did he open the shop?"; Q3: "Did he shut the door?"; Q4: "She shut the door. Did he, also, shut the door?" These were copied, thirteen times each, in random order onto a test tape with 5 seconds of silence between items. The subjects heard the tape in a recording studio and their answers: "No, he shut the shop." were recorded onto a Levers-Rich tape recorder with a Reslo dynamic microphone about one foot from the speaker's mouth. Results for two speakers with similar RP-type accents, PM, a man, and MS, a woman, were analysed acoustically. Traces of fundamental frequency (using a transpitchmeter), oscillogram and intensity (using a Frøkjaer-Jensen intensity meter) were displayed on paper using a mingograf 24B running at 10 cm/s. Durations of all the segments for the utterances of the two subjects were measured by hand from the oscillogram traces, see Fig.1.

Fig.1 

V = vocalic duration
C = consonantal "

These time measures depend on the signal level (Ramsay and Law, 1966). However, the levels were kept constant for each speaker and a non-parametric statistical test, the Mann-Whitney U test (Siegel, 1956), which assumes only an ordinal scale of measurements was used. Therefore the significance of differences of duration in comparable segments could be assessed for each speaker. Fundamental frequency, Fo, was measured as the highest value 20 ms or more after voice onset for the segment. Measurements were also made from spectrograms for speaker PM.

LISTENING TEST

Eight answers from each of 7 speakers were copied in random order onto a tape. Five phonetics students and 6 phoneticians listened to this tape and marked the positions of the stressed syllables in "No, he shut the shop." There was considerable disagreement about the stress patterns perceived. Sometimes the speaker made a different decision about his or her own speech from that of the majority of the listeners. It seems that production and perception of speech need to be examined separately for isochrony. When examining the acoustic segment durations the speaker's own assessment of stress placement was used. For speaker PM this is shown below; a stroke / is placed before the stressed syllables as heard by PM himself.

Answers to:	Stresses	Av. Durations in ms
Q1	/ No / he shut the shop.	/ 305 / 988
Q2	/ No he / shut the shop.	/ 443 / 715
Q3	/ No he shut the / shop.	/ 1095 / 350
Q4	/ No / he shut the / shop.	/ 298 / 604 / 338

Considering each sentence on its own, the durations of the feet are very significantly different with average differences which seem likely to be above the threshold for discrimination of duration (Lehist, 1970). Segment durations were compared to see whether any reduction occurred in the longer foot. Comparing Q1: "he shut the shop." with Q2: "shut the shop." it was found that: a) foot duration (taking the end of "shop" as the end of the foot) was very significantly greater in Q1; b) the durations of words "shop" and "the" were the same in Q1 and Q2; c) "shut" had significantly greater duration in Q2 than in Q1. In this case there appears to be a shortening in the direction of isochrony, but only in the word "shut", although "the" and "shop" could presumably be shortened also. The average reduction of 48 ms is probably below the perceptual threshold for durations as long as 700 ms or more (Lehiste, op.cit.). For Q1: "he shut the shop." versus Q4: "he shut the", all the word durations were the same in the two cases; there was no evidence of shortening of words within the longer foot. Speaker MS gave different stress placings for different tokens of her own Q2 and Q4 answers, but where foot duration could be established a similar lack of shortening of words in longer feet was found.

175 out of 176 responses to items spoken by PM and MS agreed that words carrying the high pitch were stressed. Other words were heard as being perhaps stressed, perhaps not. Listeners were unwilling to compare lengths of the feet formed by their judgment of stress place and it seems unlikely that such judgments are called for in speech perception.

FUNDAMENTAL FREQUENCY (Fo) PATTERNS

For speakers PM and MS and others the overall patterns of Fo were very similar in shape. For each speaker four distinct patterns resulted from each of the four questions, as sketched in Fig.2.

Fig.2

Q1: he shut the shop. Q2: he shut the shop.

Q3: he shut the shop. Q4: he shut the shop.

Fig.3 shows the actual patterns for speaker MS with the first high Fo peak in each utterance synchronized (shown by the arrow).

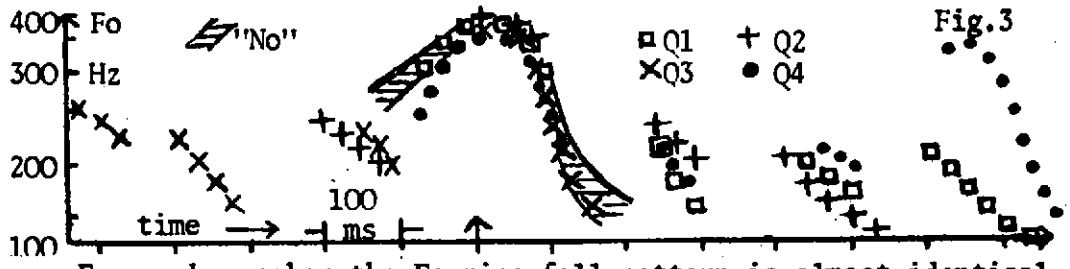


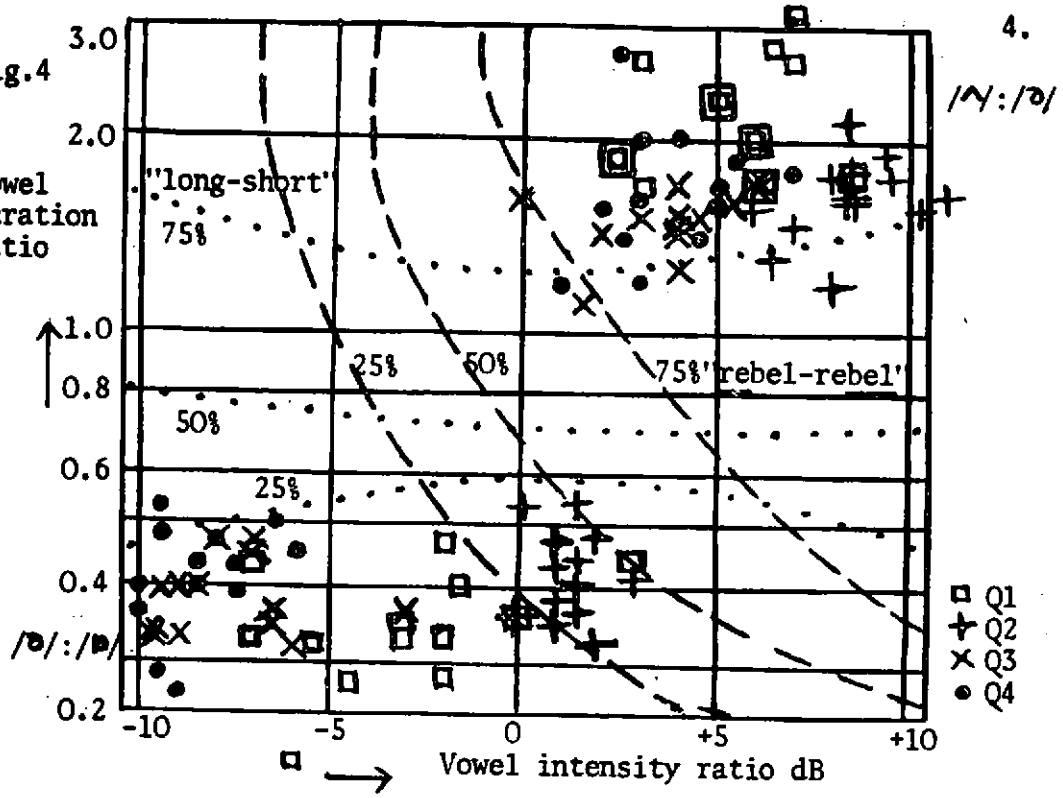
Fig.3

For each speaker the Fo rise-fall pattern is almost identical in all contexts, including the first word: "No, ..." This fixed pattern may constitute one rate-determining factor in the production of speech and may explain some of the small but significant differences of segment durations which were found for each speaker. There is some indication in the data that /ʃ/ durations are smallest when no Fo change is made during the fricative. Maximally fast vocal fold adjustments from voiced to breathy-voiced and back to voiced take about 100 ms and voiced-open-voiced adjustments need 125 ms (Rothenberg, 1972). PM seems to articulate /ʃ/ at the maximum rate, for weak voicing continues through the first part of his /ʃ/ segments. In different phonetic contexts, different articulator movements may determine segment durations. For example, when /n/, /t/, /d/, /nt/ and /nd/ were examined in words of the form /k'aC$_1$C$_2$i/ occlusion time for /nd/ was always greater than occlusion time for /n/ plus occlusion time for /d/, because in /k'ani/ the velum raising movement is not a rate-determining factor but in /k'andi/ it is. Velum closing times were established using aerodynamic measures (Scully, 1970).

WORD AND VOWEL DURATIONS

The alternation of long and short syllables is a feature which seems likely to be very important in English speech, whether foot isochrony is relevant or not. The duration ratios and perceived length ratios of successive elements in the speech will have different values depending on whether words or vowels are considered to be the relevant units. In his discussion of bisyllabic feet, Abercrombie assigns a time unit of 2 to the longer syllable and 1 to the shorter syllable of the foot. In the speech of PM and MS "shut the" seems to constitute such a long-short unit whether foot boundaries are unambiguously determined or not. For PM the ratio of the word durations "shut":"the" vary between average values of 3.8 and 5.4. Vowel duration ratios /ʌ/ : /ə/ give different results. Fig.4 shows all the data points for /ʌ/ : /ə/ and /ə/ : /ɒ/ for speaker PM. (Similar results were obtained for MS.) Peak intensity during each vowel was measured as well as vowel duration. Points for /ʌ/ : /ə/ lie just above the 75% judgment "long-short" line (shown dotted) found by Stevens et al. (1962) for two-component noise bursts.

Fig.4



Their stimuli were non-speech noises separated by only 10 ms and of total duration 310 ms in every case. Their results included also judgments of "rebel"-"rebel" associated with these noise bursts, so that the results may be quite relevant to the perception of these speech sounds. /æ/ or /i/ : /ʌ/ points (not shown here) also lie just above the 75% "long-short" line so that the vowel of "he" is always "long" relative to the vowel of "shut" while /ɔ/ : /ɒ/ points lie just below the 25% "long-short" line. The intensity ratios also seem appropriate. Vowel intensities have not been normalized to allow for different intrinsic intensities, but the correction would be expected to be less than +1 dB for /ɔ/ (Lehiste, op.cit.). The intensities of the vowels may be simply determined by Fo but evidence against this is that average intensity rose at a rate of more than 3 dB per octave, especially for /ɔ/, and that /i/ and /ɪ/ had higher intensity than /ʌ/ and /ɒ/ over the whole range of Fo used by the speaker. Different speakers may choose different compromise values for a given vowel duration and intensity in order to maintain "long-short" (and perhaps also "loud-soft") distinctions with respect both to the preceding and to the following vowel. Further perception tests may help to decide whether it is word or vowel durations which are interpreted as "long-short" or whether both ratios must be acceptable.

REFERENCES

Abercrombie, D. (1964). In Honour of Daniel Jones (Longmans), 216-222.
Fry, D.B. (1958). Lang. and Speech, 1, 126-152.
Lehiste, I. (1970). Suprasegmentals (M.I.T.).
Ramsay, R.W. and Law, L.N. (1966). Lang. and Speech, 9, 96-103.
Rothenberg, M. (1972). Proc. of the VIIth. Intl. Congr. of Phonetic Sciences (Mouton), 380-388.
Scully, C. (1970). Univ. of Essex Lang. Centre Occasional Papers, 9, 14-36.
Siegel, S. (1956). Nonparametric Statistics for the Behavioural Sciences (McGraw-Hill).
Stevens, K.N., Sandel, T.T. and House, A.S. (1962). J.A.S.A., 34, 1876-1878.