16

THE PERCEPTION OF STRESS IN SYNTHETIC SPEECH-LIKE STIMULI BY POLISH LISTENERS

WIKTOR JASSEM, JOHN MORTON, MARIA STEFFEN-BATÓG

Phonetics Department, Mickiewicz University (Poznań)
and
Medical Research Council, Applied Psychology Research Unit
(Cambridge, England)

70 Different versions of each of three speech-like stimuli which can be transcribed *sisi, sasa, sɔsɔ* were synthesized, variations being introduced in three dimensions: (a) fundamental frequency, (b) intensity and (c) duration. Naive Polish listeners judged the incidence of stress having a choice of first syllable, second syllable or no preference. One group of 50 listeners took the test once and another group of 50—twice. The responses are treated statistically and the main result is that the perception of stress depends much more on variations of fundamental frequency than on variations in the other dimensions. Comparisons are made between the present results and those previously obtained, with nearly the same material, for English listeners.

1. INTRODUCTION

In a previous article (MORTON, JASSEM 1965) two of the present writers discussed briefly the different definitions of linguistic stress and pointed out the terminological differences in the writings of phoneticians and linguists. We did not submit a definition of our own, being in the main concerned with the results of an experiment which was intended to contribute towards a clarification of the phonetic nature of stress. We also indicated that our particular experiment was based on materials which were so chosen that they could be used in making direct comparisons between languages, though at the time we presented results obtained with English listeners only.

The present paper is concerned with the results of another experiment, carried out with Polish listeners and nearly the same test material. Besides presenting the results of this new experiment we are therefore able to make some comparisons between the two languages with respect to the perception of stress.

When we speak of the 'perception of stress' we are using a traditional mode of expression which might suggest that we believe linguistic stress to belong intrinsically to one specific stage of communication, other than the perceptual, say the articulatory or the acoustic, and that what we were investigating was a psychological correlate of this virtually physiological, or acoustical phenomenon. We might be permitted to refer to our previous paper again, therefore, and uphold our conviction there expressed, that no phonological (phonetic or phonemic) feature or entity is contained solely within any of the stages of communication. If one chooses to start one's description at a particular stage, say the articulatory, and define one's features and entities at that stage, one is perfectly free to do so, and one may speak of correlates of features and entities at other stages, say the acoustic or the perceptual. It is, we believe, incompatible with the basic general facts of communication, however, to postulate—as is sometimes done—that a phonological feature or a phonological unit is essentially (or 'virtually', or 'intrinsically') psychological, or physiological, or physical, and may perhaps not even have correlates at one or more stages of communication.[1] As we pointed out, there can be no communication without an unbroken chain of code transformations between the sender's information source and the recipient's decoding system.

2. DEFINITION

Since, in comparison with the study referred to in the Introduction, the present paper broadens the view somewhat, we should now like to submit our own definition of linguistic stress. It does not claim to be particularly revealing or new, but it differs, to the best of our knowledge, from most—if not all—other definitions in that (a) it does not refer selectively to any one or more specific phonetic sound attributes (such as duration, pitch, quality or loudness), (b) it does not refer selectively to any one or more specific stages of communication. Our definition is: Linguistic stress is a feature or a distributionally defined set of mutually exclusive and/or freely varying features of the syllable, which qualifies all syllables constituting a minimal rhythmical and/or melodic unit as either weak or strong.[2]

The term 'feature' in our definition refers to any stage of communication, it may therefore be physiological, physical, or psychological, depending on the subject of investigation or the application. Once this stage is given for some definite porposes, the feature (or features) may be specified at that stage with or without reference to its correlates.

The expressions 'mutually exclusive' and 'freely varying' are meant in the sense now accepted in structural linguistics.

The terms 'minimal rhythmical unit' and 'minimal melodic unit' are here

[1] E.g. JONES (1950, 1956).
[2] This is a modified version of a definition proposed by JASSEM (1959).

used in preference to 'word' and 'sentence'. Obviously, the extent of an utterance must be specified within which the distinction between strong and weak syllables is made, i.e. within which this distinction operates. Mostly this extent (or 'area') is given as w o r d or s e n t e n c e (many specialists in fact use the terms 'word stress' and 'sentence stress'). We refrain from this for two reasons: Firstly, the terms 'word' and 'sentence' are still in need of a satisfactory and generally accepted, strictly formal definition. And secondly, these terms—whether satisfactory or not—belong to morphology and syntax, and it is our belief that for reasons of both analytical method and taxonomy the phonological features, elements and units must be defined before morphological and syntactical features, elements and units, and therefore independently of them. Rhythm units in speech may be, and indeed quite frequently are, coextensive with words (so far as the 'extent' of a word can unambiguously be specified in speech) and melodic units may be, and frequently are, coextensive with sentences and clauses, but there must be no confusion between the phonological units on the one hand and the morphological and syntactical units on the other.

Whether a single syllable constituting a complete utterance should be regarded as stressed if it contains the distinctive feature of stress (apparently it does, ordinarily) or whether such an utterance should not be considered from the point of view of stress at all, no contrast being involved, is a question which may be solved within a wider framework of linguistic analysis, and is not relevant to our definition.

3. THE SYNTHESIS OF THE STIMULI

Three quasi words which may be transcribed [sisi], [sa̱sa̱] [sɔsɔ] were pronounced several times by a phonetician in such a way that the two syllables of each 'word'[3] should sound as similar as possible, which was intended to 'keep any stress out' (i.e. to have the two syllables undifferentiated with respect to stress). Actually the second syllable came out regularly somewhat longer. This might have been due to the speaker's linguistic habits,[4] but since some (substructural) prolongation of a final vowel seems to be a rather regular feature in various languages, and as we were anxious to achieve maximum naturalness in our stimuli, we programmed the synthesis accordingly. The synthesis began with an imitation of what appeared to be a typical human version of each 'word,' and was based on spectrograms of the human

[3] We will often refer to the utterances using the term 'word' (in inverted commas). This is no more than a label which we have found convenient whenever we want to distinguish any one, or all, items with one of the vowels from any one, or all, items with another vowel. There is no implication of any grammatical sense in the label.
[4] The speaker was W. J., one of the present writers. His first language is Polish, but he has been speaking English as a second language for about 25 years.

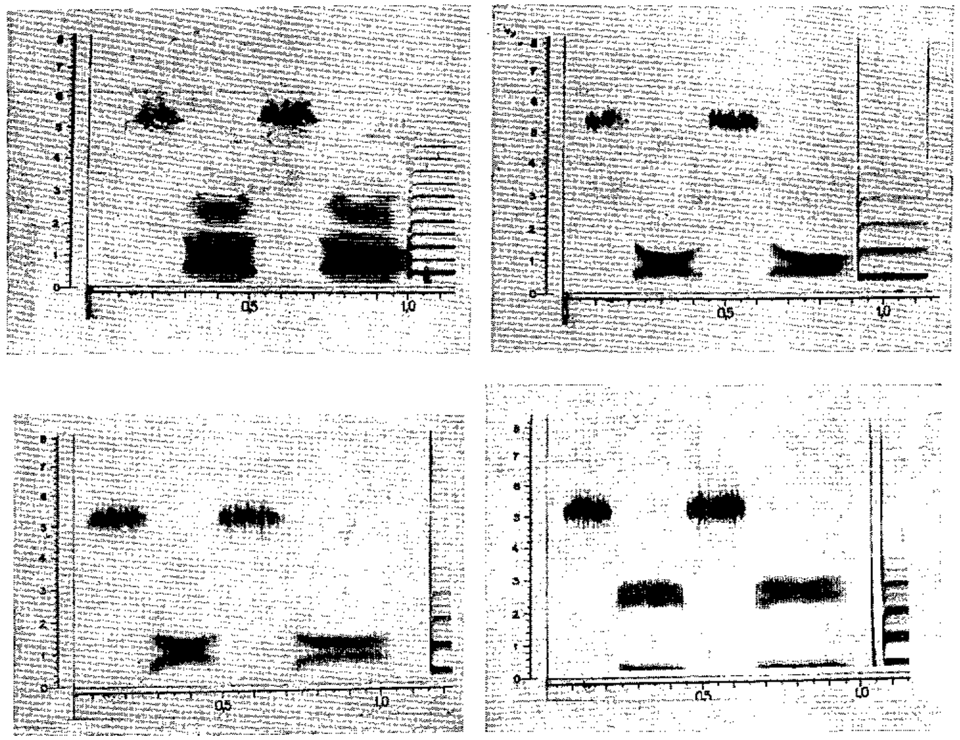

Fig. 1. Spectrograms of the 'neutral' stimuli: (a) [sɔsɔ]—narrow band, (b) [sɔsɔ]—broad band, (c) [sa̠sa̠]—broad band, (d) [sisi]—broad band.

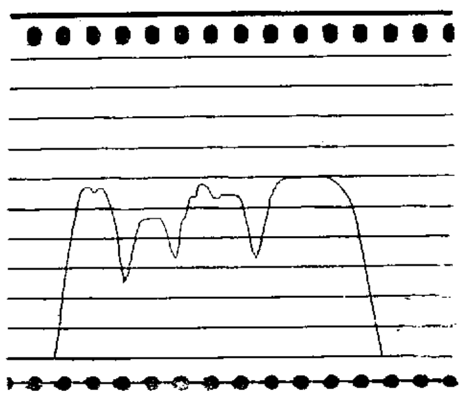

Fig. 2. Overall amplitude as a time function in [sɔsɔ]: level of vowel in the first syllable reduced by 6 db.

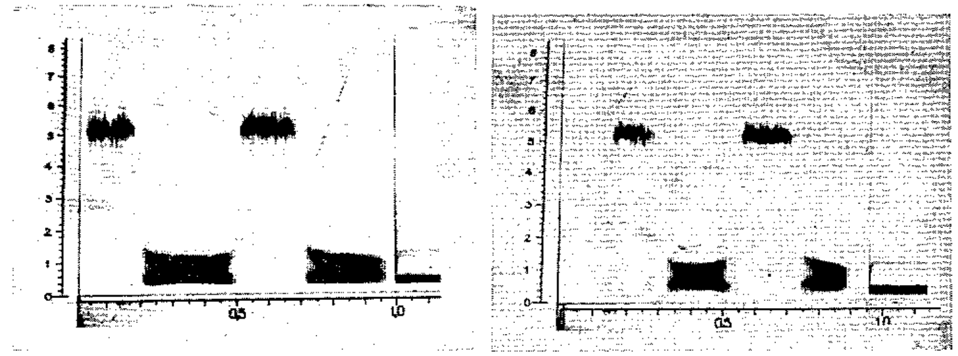

Fig. 3. Wide-band spectrograms showing temporal variations of the stimulus [sɔsɔ] (a) first vowel extended by 40%, (b) second vowel reduced by 40%.

utterance. Information extracted from these spectrograms related essentially only to the temporal courses of the formant frequencies F_1, F_2 and F_3. The characteristics of the hiss were a compromise between 'synthesis by rule' and an imitation. The hiss amplitude and the larynx-tone amplitude functions were generated 'by rule' so as to agree in time with the formant functions. A steady 120 cps fundamental frequency was applied in the synthesis of the 'neutral' stimuli (see Fig. 1).

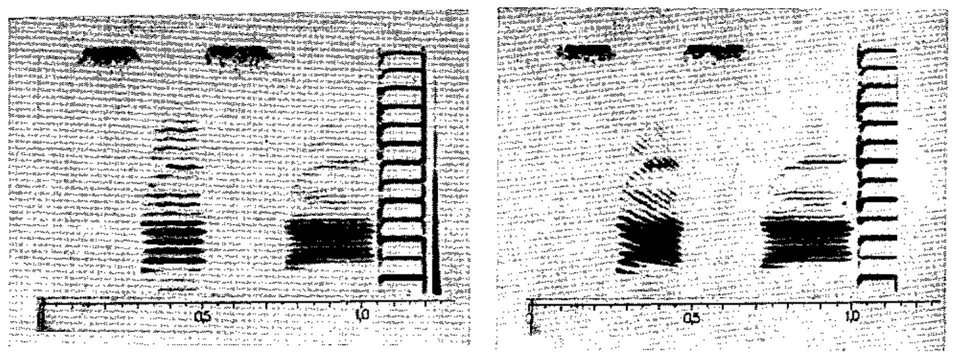

Fig. 4. Narrow-band spectrograms of [s a̠ s a̠] (a) F_0 in the first syllable altered to 190 cps (b) F_0 in the first syllable forming a slope from 190 to 120 cps. Expanded frequency scale.

Variations were then introduced in the following dimensions: (1) amplitude, (2) duration and (3) fundamental frequency. We applied amplitude variations only to the larynx tone because there seemed to be some evidence that the relative levels of consonants vary less with voice effort than do those of vowels[5] (see fig. 2). Similarly, duration was varied within the vocoidal

[5] This problem is the subject of a special research project now being carried out in Poznań.

segments only. In this case temporal adaptations had naturally to be made in the formant functions (see Fig. 3).

There were two kinds of fundamental frequency variations: in steps and in slopes. A step variation consisted in replacing the steady 120 cps larynx tone by a steady tone of a different frequency. A slope variation was introduced by a linear change of F_0. If the first syllable was altered, F_0 rose, or fell down, to the 'neutral' 120 cps, and if the second syllable was altered, F_0 fell down, or rose, from the 'neutral' 120 cps (see Fig. 4).

70 different versions of each 'word' were prepared, namely:

3 with larynx amplitude reduced in the first syllable by 2, 4 and 6 db;

3 with larynx amplitude reduced in the second syllable by 2, 4 and 6 db;

4 with the duration of the first vowel changed by +40, +20, −20 and −40 per cent;

4 with the duration of the second vowel changed by +40, +20, −20 and −40 per cent;

6 with F_0 in the first syllable altered to a steady tone of 190, 150, 130, 110, 96 and 76 cps;

6 with F_0 in the second syllable altered to a steady tone of 190, 150, 130, 110, 96 and 76 cps;

6 with a linear change of F_0 in the first syllable from the frequencies indicated above to 120 cps;

6 with a linear change of F_0 in the second syllable from 120 cps to the frequencies indicated above;

8 with all duration variations in the first syllable combined with −4 db and −6 db variations in the second syllable;

8 with all duration variations in the second syllable combined with −4 db and −6 db variations in the first syllable;

8 with all duration variations combined with −4 db and −6 db variations in the first syllable;

8 with all duration variations combined with −4 db and −6 db variations in the second syllable.

The last 16 versions were, then, complex variations in the sense that variations in two dimensions were introduced in the same syllable.

All items were tape-recorded. In addition, the 'neutral' versions of each 'word' were recorded 6 times and 4 selected, different variations were recorded again making a total of 80 items.

The synthesis was performed by means of the Parametric Artificial Talker (PAT) in the Phonetics Department of the University of Edinburgh in 1961, when the formant circuits of the device were still working in parallel connection. To save time, the synthesis was so performed that all variations were made systematically, and the individual items recorded accordingly. The tapes were then cut and all the items belonging to one list (i.e., all versions of one 'word') were connected again with pieces of leader tape, in random order.

4. PURPOSE AND PROCEDURE

The purpose of this study is twofold. First, we wish to find out how specified acoustic dimensions of speech-like signals are, within the scope of the experiment, related to the perception of linguistic stress in the case of Polish listeners, and secondly we want to compare English and Polish stress at the perception level.

It has long been recognized that synthetic speech is the best material for testing the relevance of acoustic parameters of a linguistic signal at the level of perception. This is so because each parameter may be controlled separately, which in most cases is very difficult in natural speech, and because the test material may be pre-arranged quantitatively, which also is very difficult even if natural speech is being monitored with the aid of calibrated instruments (such as voltmeters or frequency meters), and altogether impossible when it is monitored through the various natural feedback systems of the speaker (proprioceptively).

Three important studies of stress using synthetic speech were made previous to our experiments by Fry (1955, 1958) and Rigault (1962). Our materials differ from those of our predecessors mainly in that they used synthesized versions of actual, semantically valid, linguistic material (English and French), while we used speech-like stimuli of the 'nonsense' type. The reasons for making our choice of the segmental features (the 'phones') were given in Morton and Jassem (1965). Here we would only like to emphasize two aspects of our test material: (1) It is suitable for presentation and evaluation by listeners with different linguistic backgrounds and therefore permits direct comparisons between languages, and (2) when meaningful material is tested, complicating additional variables may influence the results, such as the relative probability of the items produced for recognition (discrimination, or evaluation, etc.).

The general framework of the experiment here reported was quite simple. It differed only from that of our previous study in that the number of Polish listeners was greater and that the listeners, being students in the Faculty of Languages and Literatures of the Mickiewicz University (Poznań) required no explanation as to what they were expected to mark on their answer sheets apart from a simple request to vote for either the first or the second syllable as stressed, or make a no-preference vote if they could not make a decision. However, the listeners had no phonetic training at all and their general knowledge of phonetics was extremely limited so that they may be described as 'naive'.

The tests were run in a sound-treated room, between 3 and 6 listeners taking part at a time. The tapes were reproduced over a high-quality Axiom loudspeaker. There were two groups of listeners, each group consisting of 50 persons. Group I took the test once, Group II – twice, with an interval of between 2 to 8 weeks. The entire material of 240 items was presented at one session, with two 15 minutes' breaks between the 80-'word' lists. The order of presentation was [sasa, sɔsɔ, sisi].

5. RESULTS

The primary results of the experiment were summarized in the form of *d*-score tables. A *d*-score here is the difference between votes for the first syllable and those for the second. Stimulus No. 47 in the first [sasa] list, for example, being −6N (read: Amplitude reduced by 6 db in the first syllable, second syllable unaltered, Group I) received 11 votes for the first syllable, 22 votes for the second syllable and 17 no-preference votes. This gives a *d*-score of −11. A +40N stimulus (item No. 72) on the same list (read: [sasa], first vowel extended in time by 40%, second vowel unvaried, Group I) received 27 votes for first syllable, 7 votes for second syllable and 16 no-preference votes. This gives a *d*-score of 20. 3 tables were prepared for each 'word', one for Group I, one for the first test of Group II, and one for the second test of Group II. Each table contained *d*-scores for each of the 80 items representing each 'word'. Altogether, as can be gathered from the preceding section, we had 150 votes for each of the 240 stimuli.

The primary results were first rated by means of non-parametric statistics.

1. Using the Wilcoxon Test[6] differences between the *d*-scores obtained (a) for each of the 'words', (b) by each of the two Groups of listeners and (c) in the two Tests taken by Group II were investigated.

(a) 'Words'. For both Groups the *d*-scores for [sasa] were more negative than the *d*-scores for the other two 'Words'. This means that there was a bias towards the second syllable in [sasa] as compared with the other two 'Words'. In addition, the *d*-scores were more negative for [sisi] than for [sɔsɔ] both in Group I and in the first test of Group II. These results correspond to those found with English subjects except that with the latter the difference between [sisi] and [sɔsɔ] was not significant. However, with the English subjects the first tape played gave more positive *d*-scores than subsequent ones, which effect, if operating with the Polish listeners would tend to have reduced the difference between [sasa] and the other 'Words'. Yet, the differences between the 'Words' were much greater than for the English subjects. Thus for example the mean *d*-scores for the neutral stimuli were for [sasa], [sisi], [sɔsɔ] respectively −2.1, 6.3 and 12.3 for Polish subjects (henceforth referred to as P) and −2.0, −1.4 and 3.2 for English subjects (henceforth referred to as E).

(b) Groups. There was no difference in bias between Group I and the first Test of Group II. But Group II behaved differently in the second test (see below), consequently further analysis was based on the pooled results of Group I and the first Test of Group II.

(c) Tests. The second presentation of [sisi] and [sasa] to Group II gave more positive *d*-scores than the first presentation. There was no difference between the scores on [sɔsɔ], which was already heavily biased positive with respect to the other words. This effect of repeated presentation is in the opposite direction to that found with E[7] (see Table 1).

2. The pooled results of Group I and the first Test of Group II were examined for differences between classes of stimuli (grouped by the nature of the alteration) in the relative biases between 'Words'. The differences between the 'Words' were greater for the NN (neutral) stimuli than stimuli involving alternations in Intensity and Duration simultaneously ($p<0{\cdot}001$)[8] and there was a general indication that all stimuli with small *d*-scores were more biased between the 'Words'.

Table 2 gives the *d*-scores pooled over the three 'Words' and over Group I and the first Test of Group II (300 votes for each item). Column 1 numbers the stimuli. Column 2 indicates the variations. N stands for 'neutral', i.e. unaltered, and the meaning of the figures and other letters in this column was explained above in Section 3 (The Synthesis of the Stimuli). Columns 3, 4 and 5 give the pooled votes for the first syllable, no preference, and the

[6] Briefly, the Wilcoxon test examines 'signed differences' between matched pairs taking account of 'signed ranks'. In our case, we first want to examine the difference between 'Words.' To do this, we match pairs of stimuli that are identical in all other aspects and obtain differences between the *d*-scores of each pair. E.g. stimulus No. 27 on the [sasa] list, being 120−150 ST (read: F_0 in the first syllable 120 cps, F_0 in the second syllable 150 cps, a step-wise change), received a *d*-score of −31 by Group I. In the same Group a [sisi] stimulus No. 27 (also 120−150, ST) received a *d*-score of −38. The difference between the [sasa] and [sisi] *d*-scores is here, in absolute figures, 7. The sign depends on the hypothesis, i.e. whether H_1 is that [sasa] has more negative *d*-scores than [sisi] or vice versa. The null-hypothesis (H_0) in this test (as in the simple 'sign test') is that there is no systematic difference in bias between the 'Words.' The differences in the *d*-scores for each pair of matched stimuli are then ranked, the smallest absolute difference getting 1, the next greater 2, etc. Signs are added to the rank-figures and a value T is obtained, which is the smaller sum of like-signal ranks. Critical values of T for various numbers of N (matched pairs) can be looked up in appropriate tables for several levels of significance, to adopt or reject H_1. For details see SIEGEL (1956).

[7] One difference between the experiments is in the time interval between test I and test II, E having the repetition (of only one 'Word') on the same day. Although, then, the results are not quite strictly comparable in this respect, they may be taken as possibly indicating a real difference between E and P, the latter increasingly favouring the first syllable. If this is so, the repetition would indicate that in the process of decoding the stress feature listeners depend, on repetition, more on their own information source than they do initially, which would psychologically be quite natural (the probability of the stress falling on the first syllable of a dissyllabic utterance is, in Polish, very much higher than the probability of its falling on the second sylable; this is not so in English).

[8] I.e., with a chance probability of less than 0·001.

Table 1. Differences between d-scores (Wilcoxon test)

Between 'Words'	
Group I: [sa <si <sɔ]	
Group II, first test: [sa <si <sɔ)]	
Group II, second test: [sa <si=sɔ]	
Between Tests	
[sɔ]:	1st=2nd
[si]:	1st <2nd
[sa]:	1st <2nd
Between Groups (first test in Group II)	
[sɔ]:	I=II
[si]:	I=II
[sa]:	I=II

second syllable, in that order. Column 6 contains the d-scores: positive if the preference is for the first syllable and negative if the preference is for the second syllable. Data in Columns 3, 4, 5 and 6 refer to P. Column 7 gives d-scores for E. Since the number of the English listeners was different, a correction factor of 5/3 was introduced. This corresponds to 300 P votes as compared with 180 E votes. The material presented to E differed from that presented to P only in one respect, viz. that variations of −3, −6 and −9 db were replaced by −2, −4 and −6 db respectively (the smaller steps were used with P). Wherever Intensity variations appear in Table 2, therefore, the E d-scores are taken in parentheses because for the given item they are not directly comparable with the P d-scores.[9]

3. The pooled P results were tested for non-randomness using a χ^2 test. For the 'Words' taken separately an average of 56 out of the 80 stimuli deviated from random with $p<0{\cdot}001$. For the three 'Words' pooled, all but nine of the results were significantly non-random. The exceptions were stimuli with low d-scores, and the reason appeared to be that P were biased against making 'no-preference' responses. Thus the mean percentage of 'no-preference' judgments on the NN stimuli was 42%, the corresponding figure for E being 79%.

4. The average score for the NN stimuli was (102 126 72)[10] giving a d-score of 30, which represents the overall bias of the subjects. The corresponding score for E was (after correction) (31 238 31) giving a d-score of zero. Two explanations might suggest themselves. (a) The NN stimuli, as described above, and presented in Fig. 1, were perfectly 'stressless' (or had perfectly

[9] The difference in the tapes was due to technical difficulties.

[10] The figures indicate the number of votes for the first syllable, for neither syllable ('no-preference') and for the second syllable, in that order.

Table 2.

1	2	3	4	5	6	7
1	−6N	61	108	131	−70	(−62)
2	−4N	91	142	67	24	(−25)
3	−2N	84	154	62	22	(−7)
4	NN	98	115	87	11	−5
5	NN	102	140	58	44	0
6	NN	112	131	57	55	−5
7	NN	93	109	98	−5	7
8	NN	106	134	60	46	2
9	N−2	93	148	59	34	(2)
10	N−4	117	97	86	31	(78)
11	N−6	142	81	77	65	(103)
12	40N	179	89	32	147	73
13	20N	134	126	40	94	20
14	20N	43	72	185	−142	0
15	40N	56	57	187	−131	10
16	N−40	165	96	39	126	15
17	N−20	144	122	34	110	3
18	N 20	70	120	110	−40	10
19	N 40	29	40	231	−202	−145
20	190−120 St	147	64	89	58	207
21	151−120 St	152	51	97	55	230
22	130−120 St	104	96	100	4	127
23	110−120 St	91	58	151	−60	103
24	96−120 St	127	49	124	−3	132
25	76−120 St	94	51	155	−61	167
26	120−190 St	29	21	250	−221	−267
27	120−151 St	25	32	243	−218	−242
28	120−130 St	45	34	221	−176	−207
29	120−110 St	70	100	130	−60	−65
30	120− 96 St	61	70	169	−108	−177
31	120− 76 St	65	62	173	−108	−168
32	190−120 Sl	234	33	33	201	233
33	151−120 Sl	196	47	57	139	235
34	130−120 Sl	92	143	65	27	77
35	110−120 Sl	131	109	60	71	67
36	97−120 Sl	217	54	29	188	190
37	76−120 Sl	240	23	37	203	173
38	120−190 Sl	32	13	255	−223	−275
39	120−151 Sl	46	38	216	−170	−243
40	120−130 Sl	63	63	174	−111	−238
41	120−110 Sl	65	76	159	94	−153
42	120− 96 Sl	47	57	196	−149	−217
43	120− 76 Sl	43	69	188	−145	−187
44	N−20, −4N	126	110	64	62	(−38)
45	N−20, −6N	131	95	74	57	(−65)
46	N−40, −4N	146	92	62	84	(12)
47	N−40, −6N	163	85	52	111	(−37)
48	20 N, −4N	99	106	95	4	(−62)

c. d. tab. 2

1	2	3	4	5	6	7
49	20 N, −6N	114	96	90	24	(−100)
50	40 N, −4N	178	95	27	151	(63)
51	40 N, −6N	164	86	50	114	(45)
52	−20 N, N−4	63	95	142	−79	(18)
53	−20 N, N−6	93	57	150	−57	(68)
54	−40 N, N−4	67	65	168	−101	(−20)
55	−40 N, N−6	95	42	163	−68	(115)
56	N 20, N−4	99	96	105	−6	(10)
57	N 20, N−6	130	82	88	42	(60)
58	N 40, N−4	56	88	156	−100	(2)
59	N 40, N−6	55	58	187	−132	(−17)
60	−20 N, N−4	44	104	152	−108	(−50)
61	−20 N, −6N	36	30	234	−198	(−73)
62	−40 N, −4N	44	30	226	−182	(3)
63	−40 N, −6N	40	15	245	−205	(25)
64	N 20, −4N	39	59	202	−163	(−132)
65	N 20, −6N	58	54	188	−130	(−100)
66	N 40, −4N	30	53	217	−187	(−172)
67	N 40, −6N	25	37	238	−213	(−140)
68	N−20, −4N	149	113	38	111	(33)
69	N−20, −6N	165	98	37	128	(100)
70	N−40, N−4	199	78	23	176	(73)
71	N−40, N−6	209	69	22	187	(82)
72	20 N, N−4	165	105	30	135	(88)
73	20 N, N−6	187	70	43	144	(103)
74	40 N, N−4	204	76	20	184	(90)
75	40 N, N−6	237	41	22	215	(150)
76	N 40	69	50	181	−112	−102
77	NN	105	80	115	−10	−13
78	40 N, −6N	38	19	243	−205	(−55)
79	190−120 Sl	249	30	21	228	237
80	190−120 St	141	70	89	52	240

'even' stress—which for our purposes means exactly the same) for E, but had some feature or features signalling stress on the first syllable to P. (b) The NN stimuli contained no signal feature of stress, yet P tended to vote for the first syllable. In view of the detailed results presented below, showing how different features elicited the stress responses, there is every indication that (b) is true.[11]

The Fundamental Frequency for the 'neutral' stimuli was, as indicated above, 120 cps. Items involving F_0 alternations are defined by the percentage by which the voicing on one of the two syllables was raised or lowered. The *d*-scores of P for this class of stimuli are given in Table 3 together with the corresponding E results adjusted to equate the number of subjects. In this Table FN means F_0 altered on first syllable and NF means F_0 altered on second syllable. The main effect in P as well as in E is that a syllable with F_0 differing from the neutral 120 cps receives a stress response. The exceptions with P occur where F_0 is lowered on the first syllable with a Step transition, the resulting upward inflection giving rise to a preference for the second syllable. The *d*-scores are in general much greater for E than for P, the exception being where F_0 in the first syllable was lowered with a Slope transition. A definite explanation of this difference cannot be given before Polish intonation has been analysed in much more detail than it has until now, but it should be noted that this kind of intonation (i.e., ˌ—) is not usual in utterances of two syllables (or for the last two syllables of longer utterances) in Educated Southern English, and this may be true for other types of English pronunciation.[12] The unnaturalness of this kind of intonation may have reduced the E *d*-scores.

Table 3. *d*-scores for items involving changes in Fundamental Frequency (English figures derived from Morton, Jassem, 1965 and 'corrected' to equate for the number of subjects)

			58%	25%	9%	−9%	−25%	−58%
	Percentage change		58%	25%	9%	−9%	−25%	−58%
	Frequency [cps]		190	151	130	110	96	76
Polish	Step	FN	58	55	4	−60	3	−61
		NF	−221	−218	−176	−60	−108	−108
	Slope	FN	201	129	27	81	188	203
		NF	−223	−170	−111	−94	−149	−145
English	Step	FN	208	230	127	81	188	203
		NF	−267	−242	−207	−65	−183	−168
	Slope	FN	234	235	77	67	190	174
		NF	−275	−244	−238	−159	−217	−187

6. The effects of F_0 on the perception of stress in our experiment can be considered from 5 points of view:

(a) the 'Words' [W]
(b) the type of Transition (Step or Slope) [T]
(c) the Syllables (alteration introduced in the first or the second syllable) [S]
(d) the Direction of change (raising or falling inflection) [D]
(e) the Proportions (a change of 9, 25 or 58%) [P].

[11] Again, the explanation is that P depended on their *a priori* information contained in the probability of stress incidence (cf. note 7).

[12] For Educated Southern English see, among recent publications, Faure (1962), Gimson (1962), O'Connor, Arnold (1961).

In the detailed analysis of the effects of F_0 changes we have considered each of the variables separately as well as interactions.[13] A five-way analysis of variance was performed on the d-scores. The numerical results of this analysis are given in Table 4, where $F(P)$ is the variance ratio for P and $F(E)$ the variance ratio for E. The main effect was ignored in this analysis by reversing the signs of the NF scores. This procedure increased the validity of the analysis.

Table 4. Analysis of Variance of d-scores for stimuli involving changes in Fundamental Frequency

Summary

Source		D. F.	M. S.	F(P)	F(E)
'Words'	[W]	2	115·07	1·39	3·40
Transitions	[T]	1	9 591·14	116·01***	6·40*
Syllables	[S]	1	12 667·03	153·20***	43·07***
Directions	[D]	1	4 592·03	55·54***	102·55***
Proportions	[P]	2	4 522·71	54·70***	69·28***
Interaction	[WxT]	2	280·66	3·39	0·87
	[WxS]	2	3 337·21	40·37***	3·34
	[WxD]	2	354·05	4·28*	1·95
	[WxP]	4	32·89	<1	<1
	[TxS]	1	9 545·00	115·45***	5·11
	[TxD]	1	4 560·11	55·16***	3·66
	[TxP]	2	1 030·10	12·46***	<1
	[SxD]	1	1 558·66	18·85***	8·23**
	[SxP]	2	154·65	1·86	5·75**
	[DxP]	2	416·74	5·04*	<1
Residue		45	82·67		
Total		71	66 786·00		

*** $p<0{\cdot}001$
** $p<0{\cdot}01$
* $p<0{\cdot}025$
The F-ratios for E are drawn from MORTON, JASSEM (1965)

Taking the variables separately and considering their significance for P the following results are obtained:

I. There is no difference between 'Words' in the size of the effects (The form of analysis eliminates the bias referred to in p. 1) [W].

II. The effects of a Slope transition are greater than those of a Step transition [T].

[13] We were not only interested in, say, whether all the F_0 variations were equally effective in each of the three 'Words' [W], but also whether there were differences in the effects if we considered, say, each 'Word' with each of the two kinds of Transition [WxT], or with each of the three Proportions [WxP]. Similarly, we take each of the two kinds of Transition with each of the three Proportions [TxP], etc.

III. The effects of any change are greater in the second syllable than in the first syllable [S].

IV. Raising F_0 is more effective than lowering it [D].

V. A 9% change in F_0 is less effective than either a 25% or a 58% change, there being no significant difference between 25% and 58% [P].

As can be seen from Table 4, Interactions [WxP] and [SxP] are non-significant. [WxD] and [DxP] are only significant with $p<0{\cdot}025$. The most significant Interactions are [TxS], [TxD], [SxD] and [TxP] in that order.

VI. The difference between the effects of the Slope and the Step transitions is greater in the first than in the second syllable [TxS].

VII. The difference between the effects of the Slope and the Step transitions is greater for a lowering than for a raising of F_0 [TxD].

VIII. Both lowering and raising F_0 are more effective on the second syllable. But the high significance of the [SxD] term reflects the fact that raising the second syllable is more effective than lowering it [SxD].

IX. The difference between the effects of the Slope and the Step transitions is greater for 58% than for 25% or 9% [TxP].

The similarities between E and P that are most clearcut are II, III, IV, V and VIII.

7. The d-scores for items involving changes in Intensity are given in Table 5, with the E figures adjusted as before. Since the mean bias for the 'neutral' stimuli was +30 this figure has been subtracted and the two bottom rows of Table 5 show the 'corrected' figures. This correction would hardly be justified in a direct comparison of the d-scores, but here it helps evidence the low effects of either a −2 or a −4 alteration, the scores being only just higher than those for the neutral stimuli. E and P results can directly be compared only for the −6 db change. The effects appear to be stronger for P than for E.

Table 5.

		−9 db	−6 db	−4 db	−3 db	−2 db
Polish	IN		−70	24		22
	NI		65	31		34
English	IN	−62	−28		−7	
	NI	103	78		2	
Polish 'corrected'	IN		−100	−6		−8
	NI		35	1		4

8. The effects of variations in Duration can be seen in Table 6. For P they are:

The longer of the two syllables is marked with stress. This applies even to 20% changes in duration which were ineffective with E. In addition there

Table 6. Effects due to Duration alone

		40%	20%	−20%	−40%
Polish	DN	147	104	−142	−131
	ND	−202	−40	110	126
English	DN	72	20	0	13
	ND	−150	10	3	15

is no sign of any ambiguity with the −40% changes, which for some English subjects resulted in the shorter syllable being marked with the stress. Within the ranges of variation here investigated the effects of Duration are of the same order of magnitude as those due to changes in F_0, again in contrast to the English results where the effects of Duration were relatively small.

9. The d-scores for all the 32 combinations of Intensity and Duration changes are given in Table 7. It will be remembered that in half of these the changes in the two dimensions were divided between the two syllables, while in the other half both kinds of alterations were made in the same syllable. Moreover, in half of the complex stimuli the effects, according to expectations based on the analysis of the simple stimuli, reinforce each other, while they compensate in the other half. The d-scores for the simple stimuli as

Table 7. d-scores for items involving changes in Intensity and Duration

	Duration changes	Intensity changes				
		−6N	−4N	NN	N−4	N−6
Polish	N 40	−213	−187	−202	−100	−132
	N 20	−130	−163	−40	−6	42
	NN	−70	24	30	31	65
	N −20	57	62	110	111	128
	N −40	111	84	126	176	187
	40 N	114	151	147	184	215
	20 N	24	4	104	135	144
	NN	−70	24	30	31	65
	−20 N	−198	−108	−142	−79	−57
	−40 N	−207	−182	−131	−101	−68
	Duration changes	Intensity changes				
		−9N	−6N	NN	N−6	N−9
English	N 40	−140	−172	−150	2	−20
	N 20	−100	−132	10	10	60
	NN	−62	−28	0	78	103
	N −20	−65	−38	3	33	100
	N −40	−37	12	15	73	82
	40 N	45	63	72	90	150
	20 N	−73	−62	20	88	103
	NN	−62	−28	0	78	103
	−20 N	−100	−50	0	18	63
	−40 N	25	3	13	−20	115

well as those for the neutral ones are given in Table 7 to facilitate comparisons. The d-scores for the complex items can generally be predicted from the scores for Intensity and Duration alone. Thus, where the effect of Duration opposes that due to Intensity, the influence of Duration dominates except with the item N +20; N −6, where the effect of Intensity is greater. This is in accordance with expectations since the effect of a 20% increase of the duration of the second syllable was by itself smaller than that of −6 db. The main exception to an additivity rule comes with items that have a −4 db change in Intensity. By itself, such a change has practically no effect (see p. 303), but in combination with a Duration change the effect can be quite large. Thus, for the N +40 stimulus with no Intensity change the d-score is −202, while for N +40; N −4 it is only −100. Similarly, N +20 gave a score of −40, but N +20; −4 N, where the two effects reinforce each other, the d-score is −163. The same kind of phenomenon was found with E for whom the ±20% change in Duration had no effect in isolation, but in combination with an intensity change exerted an appreciable effect.

10. A simple analysis of variance was performed on the data with 'Word', Intensity and Duration as variables. As 'Word' interacted significantly with both the other variables, a more complex 5-way analysis had to be performed to obtain meaningful results. The table of the numerical results is not given here as it is lengthy and not easily amenable to direct interpretation.[14] The more striking results, especially from the point of view of a comparison between the two languages will here be summarized:

A. Results for P.

I. There are significant differences between the 'Words' by virtue of the differential biases already noted.

II. The effect of Intensity is only slightly greater with −6 db than −4 db ($p<0{\cdot}025$) in the complex stimuli, there being no difference between changes in the two syllables—in contrast to the simple stimuli.

III. A 40% change in Duration is more effective than a 20% change ($p<0{\cdot}001$), with no difference between the syllables. There is no overall difference between increasing and decreasing the duration, but +40% is more effective than −40%, whilst −20% is more effective than +20% ($p<0{\cdot}001$).

IV. 'Word' interacts with Duration, Duration effects being stronger with [sɔsɔ] ($p<0{\cdot}01$). In addition, increasing the Duration is more effective than decreasing it with [sasa] and [sisi], but the reverse holds for [sɔsɔ] ($p<0{\cdot}001$).

B. A corresponding analysis of the E data showed somewhat similar results:

I. The difference between the 'Words' followed the expected pattern of relative biases.

[14] Interested readers may apply to the authors.

II. A -9 db Intensity change is more effective than -6 db ($p<0{\cdot}001$), but there is no difference between changes in the syllables of the complex stimuli. In the simple Intensity stimuli, however, larger effects were produced in varying the Intensity of the second syllable than the first.

III. As for P, a 40% change was more effective than a 20% change in Duration ($p<0{\cdot}01$) and $+40\%$ was better than -40%, while -20% was more effective than $+20\%$ ($p<0{\cdot}001$). In addition, variations of Duration on the second syllable were more effective than the equivalent changes on the first syllable in both the simple and the complex stimuli ($p<0{\cdot}025$) and increasing the duration was more effective than decreasing it ($p<0{\cdot}001$), neither of which results were found with P.

IV. There were no interactions of 'Word' with the other variables.

6. SUMMARY OF RESULTS

The main features of the results may be summarized as follows:

A. General

1. The relative bias between the 'Words' was similar for P and E, [sasa] receiving more votes for the second syllable.

2. P were less likely to give 'no-preference' responses than E and tended to favour the first syllable of the 'neutral' stimuli.

B. Fundamental Frequency

3. The main effect for both P and E was that a change from the 'context' or 'neutral' F_0 on one syllable produced stress responses on that syllable, the scores for E being generally somewhat higher.

4. For P and E a Slope transition was more effective than a Step transition, raising the frequency was better than lowering it, and a 9% change was less effective than a 25% or a 58% change.

C. Intensity

5. Only a difference of 6 db or more is effective in both languages. For E changes in the second syllable were more effective while for P, if anything, the reverse was true.

D. Duration

6. Changes of both 40% and 20% in Duration produced effects for P whereas 20% changes were ineffective in isolation for E. In general the longer syllable received a stress vote, but an appreciable number of E voted for the shorter syllable in the -40 N stimuli.

E. Complex stimuli

7. Where both Intensity and Duration were changed Duration dominated for P whereas Intensity had the stronger effect for E. This remains true for the -6 db stimuli.

8. Some changes which were ineffective in isolation, viz. -4 db for P and 20% Duration changes for E exerted an appreciable effect in combination with (reinforcing or opposing) changes in the other dimension.

9. For P there was no appreciable difference between changes in the two syllables for either of the dimensions; for E Intensity changes in the second syllable produced greater effects than those in the first.

7. DISCUSSION

A detailed statistical analysis of the results of the experiments has brought forward several facts about the way in which specified variations in acoustic dimensions of speech-like stimuli are related to the perception of linguistic stress by Polish and English listeners. A comparison of the behaviour of the two different groups of subjects revealed both similarities and differences. A striking similarity is the strong effect, in both languages, of fundamental frequency variations. An interesting overall difference is that variations in duration were more effective with Polish than with English listeners. It is not possible at the present moment to relate these findings to a general structural description of the two languages at the psycho-acoustic level because, though data on the acoustic features of the two languages, especially English, are not at all lacking in the literature, large-scale measurements of the 'prosodic' features are still insufficient to permit definitive generalizations. It is hoped, however, that the overall results as well as at least some of the details here reported will contribute towards a solution of the still much disputed problem of the phonetic nature, and possibly the structural status, of stress (or 'accent').

The range of variations and the number of steps along each variable in our stimuli were not sufficient to warrant a very definite statement about the relative contribution of the 'sound attributes' along the three dimensions to stress perception. One might point out that if still larger variations had been introduced along the duration dimension in our materials we might have found effects far in excess of those here reported, and the same might be said of intensity. There are, however, implications of our results which strongly tend to support the hypothesis, favoured also by the results of similar experiments made by FRY (1955, 1958) and RIGAULT (1962) that fundamental frequency changes are a predominant factor in stress perception and should therefore find their proper place in any definition of linguistic stress which specifies 'sound' attributes' or 'prosodic features'.

(1) Even variations of fundamental frequency which according to general psycho-acoustic criteria must be judged to be quite small under the circumstances (less than a semi-tone) are effective.

(2) Intensity is ineffective until a difference of 6 db is reached, which must be admitted to be a relatively large step, and even the effect of this difference may be outweighed by differences in duration. It may not be amiss to add

that at informal listening sessions with the tape containing differences of 9 db several observers expressed the opinion that the stimuli containing this step in intensity sounded as if the speaker turned his head from the microphone for the weaker syllable. Probably an intensity difference of the order of 10 db between contiguous syllables having the same pitch and the same vowel is not natural under normal conditions of speech production and perception.

(3) Our results strongly tend to support FRY's observation (FRY 1958) that there is an all-or-none relation between stress judgements and variations in fundamental frequency. If it is accepted that stress is a binary linguistic feature this polarity – not observed in the case of intensity or duration – may speak strongly in favour of the linguistic relevance of 'pitch' as a distinctive feature of stress.

REFERENCES

FAURE, G., 1962: Recherches sur les charactères et le rôle des éléments musicaux dans la prononciation anglaise, Paris.

FRY, D. B., 1955: Duration and Intensity as Physical Correlates of Linguistic Stress, JASA **27**, 4, 765 – 768.

FRY, D. B., 1958: Experiments in the Perception of Stress, Language and Speech **1**, 3, 126 – 132.

GIMSON, A. C., 1962: An Introduction to the Pronunciation of English, London.

JASSEM, W., 1959: The Phonology of Polish Stress, Word, **15**, 2, 252 – 269.

JONES, D. 1950: The Phoneme, Its Nature and Its Use, Cambridge.

– 1956, Outline of English Phonetics 8, Cambridge.

MORTON J., JASSEM W., 1965: Acoustic Correlates of Stress, Language and Speech, **8**, 3, 148 – 158.

O'CONNOR J. D., ARNOLD G. F., 1961: Intonation of Colloquial English, London.

RIGAULT A.; 1962: Rôle de la fréquence, de l'intensité et de la durée vocalique dans la perception de l'accent en fraincais, Proc. 4th Int. Congr. Phon. Sci, 735 – 748.

SIEGEL S., 1956: Nonparametric Statistics for the Behavioral Sciences, New York.