<u>Sentence Intonation and the Recognition of Noun-Verb Oppositions of the So-called Minimal Word Pairs</u>,

by Heleen V.Deighton -van Witsen,

<u>Computer Processing</u>,

by Jan G. Blom and Leo W.A. van Herpt.

Contents:

Continued overleaf

0. Summary

0.1 In 1968 a pilot-investigation took place concerning the recognizability of the so-called minimal word pairs(MWP).
D.B.Fry (1958) *) had experimented with MWP in an investigation with synthetic items of the subject,object type. In the 1968 experiment use was made of spoken items,which were used in a forced-choice perception-test. These items (nouns and verbs) had been "extracted" from two types of sentences:
I) From sentences where they coincided with a sentence-nucleus and
II) From sentences where they had occurred outside a sentence-nucleus.
The hypothesis put forward at the time was that items from a sentence-nucleus would be recognizable,but that items which had occurred outside such a nucleus would not be recognizable.
The original χ^2-calculation did not refute this hypothesis,albeit that there was a considerable difference in the distribution of correctly identified nouns and verbs.
The physical parameters of the items were investigated recently. Amplitude ratios were shown to have played an important part,especially were the speaker-effect was concerned.

*) D.B.Fry,'Experiments in the Perception of Stress',Language and Speech I,(1958),pp 126-152.

1. D.B.Fry's 1956 Minimal Word Pair Experiment (synthetic items).

1.1 In 1956 Fry tried to establish which were the parameters responsible for stress perception. Listeners perceive stress as 'variations in a complex pattern bounded by four psychological dimensions namely,length,loudness,pitch,and quality'.Fry experimented with three out of four dimensions. He made use of synthetic speech items of the word-pair type such as subject,object;synthetic speech items and therefore controlable.The dimensions Fry varied in three "sub"-tests were length,loudness,and pitch,or in physical terms,duration, intensity(amplitude), and fundamental frequency.Quality or formant structure he left out of consideration,although he states(p128) that 'certain quality differences in English have particular significance in stress judgment'.The variations were made in the vowel stretches of the items.

1.2 Fry's first test involved variations in duration and in intensity, but the fundamental frequency was kept constant,and so, of course, were the formant structures. Result: both duration and intensity were shown to have acted as stress cues.
Then Fry combined duration changes with step changes of fundamental frequency and showed that the direction of the step changes had acted as a cue and that 'the magnitude of the frequency change had (had) no marked effect'.
In the third test fundamental frequency within one syllable was varied and in this test there was a 'range of patterns which imposed sentence intonation on the test items. The result again demonstrated the all-or-none effect of frequency changes and showed that this may outweigh the duration cue altogether'.

2. The Minimal Word Pair Experiment (spoken items).

2.1 Fry's last "sub"-test involving intonation raised an interesting point,namely the question how much sentence intonation contributes to the recognition of stress. In order to find out a little more about this it was decided to make use of the same type of material, minimal word pairs (MWP),carefully selected(see 3.2), and of the noun/verb opposition.

Six MWP were chosen and used in their respective grammatical function in phrases of two types:

I in an intonation nucleus *)

II outside an intonation nucleus (see 5).

The six pairs thus yielded 24 items which were incorporated in 24 sentences.

It had been agreed upon beforehand to have 10 native English speakers(4) utter the 24 sentences, which in turn produced 240 items(5).

2.2 The experiment consisted of two parts.

2.2.1 I: A perception test. The 240 items were "gated" **), that is, lifted out of their surroundings and they were re-recorded on a listening-tape in random order.

This tape was presented to 3 groups of listeners (8.2).

Both groups and numbers had been agreed upon beforehand, one group of native English among them. This yielded 7200 responses.

The test was a forced choice one; listeners were requested to underline either one or the other of two items as follows:

a discount - to discount

a refund - to refund

a discount - to discount

2.2.2 The aim of the perception-test was formulated at the time:

'The purpose of the perception-test was to find out if listeners, confronted with the 240 stimuli, would:

I pick out more of the pairs in an intonation nucleus as noun or as verb as the case might be,

II pick out fewer of the pairs outside an intonation nucleus.

A margin of error and of chance had to be allowed for'.

In other words, I-type items would prove to be recognizable without the surrounding context, but II-type items would not be recognized.

*) Intonation nucleus: that part of a sentence where the intonation pattern shows a greater variation in pitch in comparison to its surroundings

**) See the article on the operation of the IFA-gate by professor H.Mol in these Proceedings.

2.2.3 II:An acoustic investigation. The items would be measured as regards the vowel stretches of both first and second syllable. Parameters to be investigated:duration-,fundamental frequency-,and amplitude-ratios.In fact the same parameters which Fry had used in his experiment.Measurements were taken straight from the oscilloscope. For an explanation see F.J.Koopmans-van Beinum 1973*).
The processing of these data was undertaken recently by J.G.Blom and L.W.A. van Herpt,and will be gone into below (section 10 ff).

2.2.4 The aim of this part of the investigation was to try and find out which of the physical parameters investigated could be shown to be common factors in the formation of noun-items and of verb-items. What,in fact distinguishes them. All this within the limitation of the parameters mentioned above.

3. The Minimal Word Pair Corpus Used in 1968.

3.1 Reasons for the use of MWP.
Fry had experimented with synthetically produced MWP-type items for the obvious reason that he was able to produce suitable material with a minimum of parameters(variations of duration,intensity, and fundamental frequency).The fourth dimention,quality,which he calls a psychological dimension,he left out of consideration, understandably so,for,as he writes (p 128):'The substitution of the neutral vowel [ə] for some other vowel,the reduction of a diphthong to a pure vowel,or the centralization of a vowel are all powerful cues in the judgment of stress'.
The material selected for the experiment conducted in 1968 was drawn from the corpus of MWP material in English as well.Grateful use was made of A.Vanvik's**) exhaustive list of stress-differentiating words,which he obtained by extracting all relevant entries from D.Jones's 'English Pronouncing Dictionary',eleventh edition, 1956.

*) F.J.Koopmans-van Beinum,'Comparative Phonetic Vowel Analysis', Journal of Phonetics I,(1973),pp249-261,esp.p 252,Figure 1.

**) A.Vanvik,'On Stress in Present-day English (Received Pronunciation)',(1961),Bergen-Oslo,Ch V.

3.2 Criteria used for the selection of MWP,

The following criteria were used for the choice of MWP:

A They would have to be words in every-day use,and not liable to misinterpretation.
Therefore word pairs like e.g. gyrate (adj) - gyrate (verb) were considered unsuitable material.

B They would have to be distinguishable with the aid of three out of four cues mentioned by Fry,viz. duration,fundamental frequency,and amplitude (a term used in preference to 'intensity').
The fourth dimension,formant structure was thought to be such a powerful cue that it might obscure the action of the other three dimensions.

C The stress pattern (as shown in Jones's dictionary) would have to contain a stress opposition in (preferably the first) one of its phonetic notations.

D The use of each member of a MWP would have to be as evenly matched as possible.This last requirement was impossible to control within the scope of this (pilot-)investigation.Therefore the choice -however carefully made - became a subjective one,for which the present writer was responsible.
The noun/verb opposition within the MWP was by far the commonest of the existing oppositions and eventually the following six noun/verb pairs were selected:

discount,noun ('diskaunt)=('--)
verb ('diskaunt)=('--) or (-'-)
overhaul,noun ('ouvəhɔ:l)=('---)or (--'-)
verb (ˌouvəhɔ:l)=(ˌ--'-)
refund ,noun ('ri:fʌnd)=('--)
verb (ri:'fʌnd)=(-'-) or('-'-) or (ri'fʌnd)=(-'-)
import ,noun ('impɔ:t)=('--)
verb (im'pɔ:t)=(-'-) rarely ('impɔ:t)=('--)
insult ,noun ('insʌlt)=('--)
verb (in'sʌlt)=(-'-)
increase,noun ('inkri:s)=('--) or(iŋ-'k-,in'k-,iŋ'k-)=(-'-)(-'-)
verb (in'kri:s)=(-'-) or(iŋ'k-,'ink-,'iŋ-k-)=(-'-),('--),('--)

The first five of these word pairs occur in D.Jones's English Pronouncing Dictionary,11th ed.,1956.

The last, increase, n.-v., is added by Vanvik (p 53/54).*)

3.2.1 A short explanation seems called for at this stage.

discount was chosen notwithstanding the fact that the stress difference was marked in the second place. Generally speaking the material from which a choice could be made was very limited. Vanvik (p 55) states that out of a possible 248 items 'where the same orthographic form may represent different parts of speech according to which syllables are stressed....... it was obligatory to use the same phonemes and different stress for some 60 entries only.' This number of 60 stress oppositions includes all forms of opposition, noun-verb, noun-adjective etc. Those qualifying on all four criteria were few.

overhaul, a three syllable word, that is true, but by all appearances clearly stress differentiating. Measurements could be kept conform by treating the first two syllables as one and measuring on the vowel strectches [ou] and [ɔ:].

refund, a risk was taken as quality differences seemed possible from the start, but here again the limitation of the material was at the bottom of the inclusion of this item.

import, insult qualified in every respect.

increase, Vanvik mentions (p 55) that he consulted some leading English phoneticians about their pronunciation of this word, which they stress-differentiated. It appeared reasonable to include it in this experiment.

4. Speakers:

The items used were spoken by 10 native English speakers (British male R.P. speakers **)). This number had been agreed upon beforehand, but some more speakers were recorded. All male to make the group homogeneous for measuring purposes later. Some recordings could not be used, for reasons such as asthmatic speech, audibly ill-fitting teeth etc. The decision as to what does or does not constitute R.P. pronunciation was an arbitrary one, obviously.

*) Vanvik's special inclusion of increase seems unnecessary as both the 11th ed.of Jones's dictionary and for instance the earlier 7th ed., An English Pronouncing Dictionary, give this word as stress differentiating in the first variants of the respective noun/verb pronunciations.

**) R.P.: Received Pronunciation as defined in the Introduction of Daniel Jones's Pronouncing Dictionary. In general the pronunciation heard in Southern England and especially used by those with boarding-school education.

Criteria by which to judge this sort of problem can be found in the article by J.G.Blom and F.J.Koopmans-van Beinum*)
Such criteria were not available for the R.P. pronunciation of English at the time and a decision had to be made somehow of what to use and what to scrap. The problem of the limitation of available male R.P. voices was also considerable. In my opinion, which is bound to be a subjective one, all voices used in the experiment were homogeneously R.P.; most speakers were rather 'public' figures from, for instance, the consulate, the British Council, and one lecturer of the English department of the University of Amsterdam. Nearly all had public school backgrounds.

5. Sentences:
It was mentioned above that ideally speaking the items should have been taken from spontaneous conversation if a complete antithesis of synthetic speech was the aim, but this constituted insuperable difficulties for a statistically significant number of items to be the result,- even for a pilot-investigation such as this was -.
The next best thing was to introduce the items as inobtrusively as possible. As material from native English speakers was sought for other tests at the time it was decided to combine the session for more than one test. The time involved for the collection of all the required material being some 20 minutes. The MWP sentences came in the middle of the session, with sentences printed on cards, (the other items were lso read from cards).
Sentences of both type I and II appear below, i.e. items inside an intonation nucleus and those outside one. One or more nuclei can occur in a sentence,
The numbers refer to the order in which the sentences were arranged.

*)An Investigation Concerning the Judgment Criteria for the Pronunciation of Dutch I, in these Proceedings pp 1 - 24.

Type I	The discount must be considerable.	(1)
	We 'll discount this theory.	(5)
	The car needs a thorough overhaul.	(12)
	They 'll overhaul it at the garage.*)	(2)
	Do you want a refund, madam?	(3)
	We 'll refund the money on principle.	(9)
	Imports have risen sharply.	(8)
	Britain imports nuclear heads.	(4)
	An increase in productivity will be necessary.	(11)
	The T.U.C. will increase pressure.	(6)
	His insult must be overlooked.	(7)
	If we insult him he will retaliate.	(10)
Type II	On FRIDAYS you get a discount, NOT on SATURDAYS.	(18)
	If we CAN'T discount this theory, who CAN?*)	(21)
	I don't THINK an overhaul is due yet.	(16)
	Let the GARAGE overhaul it for you.	(19)
	I have HAD my refund, thank you very much.	(17)
	If they refuse there is nothing I can do to MAKE them refund it.	(13)
	Can anyone guarantee it will not DIMINISH our imports.	(23)
	How can you be sure Britain WON'T import nuclear heads	(22)
	Is there a GENERAL increase or just a SEASONAL one.	(20)
	The T.U.C. bosses must certainly NOT increase demands.	(14)
	He's bound to RETALIATE this insult.	(15)
	Do you really think he WILL insult us.	(24)

*) 'Garage' and sentence 21 caused comments from the speakers. They will be discussed below. (6.)

6. Procedure:

The procedure is described in some detail here,as for this type of investigation much depends on the way the data are collected. The aim was to get items which were as near as possible to normal free running speech,so great care was taken to ensure that the speakers were not influenced by the unfamiliar surroundings.Also the test-leader was careful not to influence the speakers directly by saying any of the items during the test.

The procedure was as follows:each speaker was taken to the recording cubicle and there the test-leader talked with him for a few minutes.This put the subjects at their ease and yielded valuable information about the presence or absence of regional accents, speech defects and speech rate.This information was registered straight after the test.It also gave the technician in the recording studio time to adjust the recording apparatus.

Data for a sibilant test were collected during the same session, the items were printed on cards.The 40 cards read for the sibilant test took some 5 minutes.Then the sentences for the test discussed here were read from separate cards in the order shown above,with an instruction card inserted between type I and type II sentences, which had the following text printed on it:

Please stress CAPITALIZED word
WHAT did you say before that?
What DID you say before that?
What did YOU say before that?
What did you SAY before that?
What did you say BEFORE that?
What did you say before THAT?

The subjects read these sentences aloud,as a rule no further instructions were given.However,some speakers slowed down their rate of speaking a little,but resumed their former speech rate after being asked if they would normally say things that way.

The tape of all recordings was listened to independently by two members of the IFA staff and no more than five instances of slight influencing were established.On a corpus of 240 items this was considered to be of no real consequence.

In order to get speech-items which were as natural as possible it was hoped that the speakers would not spot immediately what the test aimed at.This was on the whole accomplished.Nearly all the subjects were under the impression that we had been testing them on their pronunciation of English:,"Good English","the Queen's English".The word garage had struck them as a hidden clue,especially as it had occurred twice.Only one speaker said afterwards,that, quite at the end of the test,he had come to the conclusion that we "were trying to sort something out to do with 'import -im'port". The use of sentences which were not of uniform length or appearance had the very purpose of obscuring the items needed for the test. Sentence 21 was often read with emphatic WE.

7. Acoustic Measurements:

7.1 Three parameters were measured per vowel-stretch,both for the first and second syllable per item.

7.1.1 Duration measured in msec,directly from the oscilloscope with the "gating" equipment*).

7.1.2 Fundamental frequency measured in Hz,also directly from the oscilloscope by measuring the repetition rate of the damped oscillations. In the case where vowels showed simple changes of rise or fall of 4 Hz or more the geometrical mean was taken of the beginning and the end of the vowel-stretch. If complex changes occurred the mean was computed of beginning,middle and end.

7.1.3 Amplitude: the most difficult parameter to say anything about witheliciting a host of questions.Some people might perhaps have preferred the use of a different term;"intensity" perhaps,but this would have complicated things further.(see Mol and Uhlenbeck**)) The amplitude of the vowel-stretch was defined and measured as the peak value in the curve as seen on the oscilloscope.By simply determining the ratio of the amplitudes of the first and second syllable of an item,a straightforward physical parameter was arrived at, being independent of the absolute levels produced by the speakers.

*) See the article about the IFA-gate by H.Mol in these Proceedings, and F.J.Koopmans-van Beinum(1973).

**) H.Mol and E.M.Uhlenbeck,'The Linguistic Relevance of Intensity in Stress',(1956),Lingua V,2,pp205-213.

7.2 The measuring of overhaul must be explained.It was impracticable to have one three-syllable word among five two-syllable ones. The middle syllable measurements,'ver',were left out of the computations and overhaul was treated as a two-syllable word.
For the data of the measurements see those of insult and refund , which did best and worst respectively in the perception test.

8. The Perception Test.

8.1 The recorded material:
The 240 items (10 speakers,24 items per speaker) were given code-numbers,and were re-recorded in random order with 4-6 second intervals.Although the actual tape for listening purposes was the copy of a copy of the master-tape,the quality of the items was good. There were slight extraneous noises which were the result of suppressing the surrounding material,but these did not affect the stimuli themselves. This slight noise was hardly noticeable at a single hearing.

8.2 The listeners:
Three types of subjects were invited to take part in the perception test.

A A group of 10 native Englishmen and women;R.P.speaking and not connected with or intimately aquainted with the 10 speakers,who had their voices recorded.Restrictions which made the composition of this group rather difficult.

B A group of 10 senior students of English of the University of Amsterdam,who were not used to a language laboratory.

C A group of 10 first-year students of English of the same university ,but these students had been instructed in a language laboratory.

Altogether 30 listeners,reacting to 240 stimuli.The result ca. 7200 responses.Notwithstanding the forced choice character of the test.2.6% "no-scores" resulted.These do not, however,materially influence the test.

8.3 Apparatus for recordings and for listening test:

The recordings were made with an Ampex recorder,a cardioid Philips microphone,magnetic tape:Scotch tape,low print.Recording speed: 19 cm/sec, one of the two tracks was used.

Copies were made with a Revox tape-recorder,same speed and the same type of magnetic tape was used.One track was used while the other was erased simultaneously.

The IFA-gate was used for the gating of items and for gating first and second vowels.The storage oscilloscope was a Tectronix RM564.

For statistic purposes an IBM 1130 was used.

For the perception test the same magnetic tape was used and in the cubicles there were Revox tape-recorders and Beyer DT48 headphones.

9. Data of Acoustic Measurements.

The first and second vowel-stretch parameters are not given in full.Those of insult and of refund can be found in Appendices I and II.These two words were selected as the result of the perception test has shown that insult had fewest correct responses,while refund had most.

The processing of the acoustic material will be found in MANOVA I,II,andIII on page 82 f.f.

10. Data of Perception Test.

10.1 Influence of sentence intonation:

The hypothesis put forward in 1968(2.2.2) concerning the ability of listeners to distinguish correctly "inside" an intonation nucleus and their being unable to do so with items occurring "outside" such a nucleus was tested on the results of a χ^2 computation at the time,by looking at the distribution of correct responses for the four groups of items and that of the incorrect responses.

(1 - 1) nouns "inside" intonation nucleus.
(1 - 2) nouns "outside" intonation nucleus.
(2 - 1) verbs "inside" intonation nucleus.
(2 - 2) verbs "outside" intonation nucleus.

10.2 Distribution and confusion matrices:

For the distribution of scores of the perception test see Figure I for the distribution among the 6 words and Figure II for the distribution among the 10 speakers.

Distribution matrix

		correct	incorrect	
nouns	inside	1317	483	1800
	outside	947	853	1800
verbs	inside	995	805	1800
	outside	871	929	1800
	Total	4130	3070	7200

If we consider the distribution of the total figures,we see that the z-score for the total column 'correct' is:

z = 12.6

This figure is too high not to justify the experiment.

Confusion matrix

		nouns	verbs	
nouns	inside	1317	483	1800
	outside	947	853	1800
verbs	inside	805	995	1800
	outside	929	871	1800
	Total	3998	3202	7200

The confusion matrix gives a z-score for the total of "nouns" (i.e. the total number of times the listeners scored noun-items) z = 9.4

This figure shows that listeners choose nouns significantly more often than verbs.

10.3 χ^2 -computation and z-scores:

The figures which can be found in the distribution matrix resulted in $\underline{\chi^2 = 264.33}$ $\underline{ndf = 3}$

The distribution of the cell figures of χ^2 shows that the main contributions to the column "correct" comes from (1 - 1), nouns spoken in an intonation nucleus.

The z-scores per group of items in the "correct" column are for:

(1 - 1) $z = \frac{417}{21.2} = \underline{19.72}$

(1 - 2) $z = \frac{47}{21.2} = 2.22$

(2 - 1) $z = \frac{95}{21.2} = \underline{4.48}$

(2 - 2) $z = \frac{-29}{21.2} = -1.37$

These z-scores show that:

(1 - 1) The hypothesis that listeners were not able to recognize this type of item must be refuted.

(1 - 2) We see a onesided P-value of 2.64%.At 0.01 level this is not significant.Listeners must be assumed not to have been able to recognize this type of item.

(2 - 1) The hypothesis that listeners were not able to recognize this type of item must be refuted,albeit that z = 4.48 is not nearly so convincing as the result z = 19.72 for (1 - 1) type of items.

(2 - 2) We see a onesided P-value of 17.06%. So here again it must be assumed that listeners were not able to recognize this type of item.

10.4 Conclusions based on χ^2 and z-scores:

When we look at the results of both groups of items <u>outside</u> an in-

tonation nucleus we see that listeners were not able to discriminate either nouns or verbs in a significantly correct number.
But in an intonation nucleus,however,the situation is different. Both verbs and nouns are recognized,albeit with greatly diffenent results.
The χ^2 distributions and the z-scores show that the centre of gravity for the "correct" results lies very much on the noun-items. This is in accordance with other findings on the subject of noun/verb recognition.

This point was reached in 1968,but the need was felt to scrutinize the available material more carefully.An attempt at such a scrutiny follows below.

11. Processing of MWP Material.

11.1 ITMAN analysis*):
In order to gain more insight into the way people had scored,in fact,relating the degree of difficulty of the items and the ability of the listeners,an item-test analysis was carried out.
Here the four separate groups,each yielding 60 items were investigated separately.

11.1.1 A total of 141 items were shown to have been scored below chance in the following way.

(1 - 1) - 6 items scored below chance; i.e.54 items recognized.
(1 - 2) - 30 items scored below chance; i.e.30 items recognized.
(2 - 1) - 29 items scored below chance; i.e.31 items recognized.
(2 - 2) - 34 items scored below chance; i.e.26 items recognized.

Here again the results for (1 -1) items are highest,which strengthens the conclusion that (1 -1) items are recognizable as such.
The verbs in an intonation nucleus (2 -1) are shown to have a much

*) J.G.Blom and L.W.A. van Herpt,'test Analysis System ITMAN', Internal Publication 31.1,Institute of Phonetic Sciences, Amsterdam,(1972).

lower result, one in fact, that is not really different from either type of item outside an intonation nucleus, which rather weakens the status of (2 -1) items.

11.1.2 Word- or speaker-dependence?

We wanted to know how these results had been arrived at.
Was it a case of word-dependence or of speaker-dependence.
The 141 items below chance were distributed as follows:
Average 141/6 = 23.5 per word.
W_3 = 17, W_5 = 19, W_1 and W_4 = 25, W_2 = 27, W_6 = 28.

$\chi^2_5 = \underline{3.41}$

This is not significant, so there was no word-dependence.

The distribution among the speakers; average 141/10 = 14,1.
2 speakers = 11 ; 3 speakers = 12 ; 1 speaker = 14 ;
3 speakers = 17 ; 1 speaker = 18

$\chi^2_9 = \underline{5.28}$

This is not significant, so there was no speaker-dependence.

11.1.3 Reliability figures ITMAN.

Distributed as follows:
(1-1) = 0.89 (2-1) = 0.32
(1-2) = 0,64 (2-2) = 0,61

From these figures we may conclude that a repetition of the four tests would give these results, if put into words:
(1-1) a very great chance of a similar result.
(1-2) and (2-2) the result would perhaps be similar, but this is not nearly so certain as for group (1-1).
(2-1) quite likely a different result would occur.
This again rather undermines the status of verbs inside the intonation nucleus.

11.2 Item-test correlation (for RIT < 0.20).

The item-test correlation figures, RIT < 0.20 had the following distribution for speakers:

(1-1) 10 items out of 60 (of which 6 speakers with 1 item RIT < 0.20)
(1-2) 28 items out of 60 (speakers varied from 1-5 items RIT < 0.20)
(2-1) 29 items out of 60 (speakers varied from 2-4 items RIT < 0.20)
(2-2) 26 items out of 60 (speakers varied from 1-5 items RIT < 0.20)

and for words:
(1-1) 10 items (of which 3 words had 1 item RIT < 0.20
2 words had 2 items and one had 3)
(1-2) 28 items (words varying from 3-4 items, RIT < 0.20)
(2-1) 29 items (words varying from 1-8 items, RIT < 0.20)
(2-2) 26 items (words varying from 4-5 items, RIT < 0.20)

This shows the greater homogeneity of the speakers, with the greatest occurring in (1-1).
The homogeneity was smaller for words, being smallest for (2-1).

11.3 Spearman's rank correlation.

A Spearman rank correlation test was done to see if "verbs-inside" and "verbs-outside" were parallel tests for the recognition of verbs. The outcome was not significant at 0.2. Therefore (2-1) and (2-2) are not comparable tests. The difficulties in recognition of (2-1)-items are not the same as those of (2-2)-items.

11.4 ANOVA-test.*)

A variance analysis of univariables was carried out on scores, with complete four factor design:

factor 1	speakers (S)	10 levels
factor 2	words (W)	6 levels
factor 3	word-type (N)	2 levels
factor 4	"inside/outside" (S)	2 levels

11.4.1 Significant effects were found for
W-, N-, WN-, S-, WS-, NS- effect.

*) Anova - Analysis of Variance for IBM 1130, Disc Monitor System Version 2. Essentially described by H.O. Hartley.

P-effect was not significant.

This is something to which attention must be drawn, as the acoustic investigation shows that speakers produced differences which were significant to which the listeners apparently did not respond.

W-effect was significant.

As the P-effect was negative there can have been no systematic connection between speakers and degree of difficulty of items. There is, however, a connection between words and degree of difficulty (see figures 1, and 2 for distribution of correct scores).

N-effect was highly significant.

Nouns and verbs are correctly named with very dissimilar frequencies. A confirmation of the above tests (see figure 2).

WN-effect, only marginally significant.

This effect dropped out after final pooling of Error-term.

S-effect, was significant.

The ITMAN test already showed that items in an intonation nucleus were recognized better than items outside one (see figure 2).

WS-effect. Some words are, on the whole, recognized better 'inside' than 'outside' the intonation nucleus (see figure 2).

NS-effect was highly significant.

'Nouns-inside' are recognized much better than 'verbs-inside' (see figure 2).

Pooling of the highest order interaction + 3rd order interaction + + P and its interactions leaves the following significant effects: W-, N-, S-, WS-, NS- effect.

This exhausts the examination of scores. The next paragraphs will be devoted to the examination of the physical parameters.

12. MANOVA I and II*)

The Manova-tests, of which 3 were carried out tested the material acoustically. That is, in MANOVA I the ratios of the three parameters, measured for the vowel-stretches of first and second vowels of each item (see 3.2.1 for overhaul). So ratios of duration fundamental frequency and amplitude.
In MANOVA II these ratios were transformed to logarithms.

12.1 MANOVA I:

Testing against the first estimation of the Error-term (highest interaction) and subsequent further pooling of insigificant effects in the Error-term left the following significant effects:
P-, W-, PW-, N-, WN-, W-, and NS-effect.
In order to be able to generalize about speakers the P-effect was put into the Error-term. Significant effects were:
W-, N-, S-, WN-, NS- effect.
These are the same as those in the ANOVA test (all except the WN- effect).
This is an interesting point as ANOVA tests on scores, whereas MANOVA tested the above-named physical parameters.

12.2 MANOVA II:

Testing took place against the first estimation of the Error-term (highest interaction). Significant effects: P-, W-, PW-, N-, WN-, S-, and NS- effect.
Pooling of insignificant effects resulted finally in an Error-term consisting of PN- + PWN- + PS- + WS- + PWS- + PNS- + WNS- + PWNS.
Remaining significant effects were:
P-, W-, PW-, N-, WN-, S-, NS- effect, in fact the main effects P-, W-, N-, and S- effect.
This means that we can not generalize either o speakers, or words. Nor can noun-verb oppositions be compared or the "inside"-"outside" oppositions.

*) J.G. Blom and L.W.A. van Herpt, 'MANOVA, Een multiple variantie analyse programma voor IBM 1130', Internal Publication 35.0, Institute of Phonetic Sciences, Amsterdam (1972).

One begins to wonder if Minimal Word Pairs as a class exist at all - at least when testing on similar parameters to those used by Fry in his synthetic material. It will be noticed that significant effects of MANOVA II and of MANOVA I were identical.
In MANOVA II the P-effect has become marginally significant.
In order to be able to generalize about words the W-effect was put into the Error-term. We wanted to act as if Minimal Word Pairs existed. When this was done the P-effect became not significant.

12.3 MANOVA III:

This analysis was carried out on two parameters instead of three. Here the logarithms of the ratios of durations and fundamental frequency of both vowel stretches per item were investigated. The reason for this was that the amplitude ratios were seen to differ greatly per speaker.
Testing against the first estimation of the Error-term (highest interaction) PWNS showed that P-effect was not significant.
This suggests that speakers are measurably distinguishable in the contrasts of their amplitude ratios.
Within the scope of this investigation we can not say with any confidence that this acoustic contrast is perceived by the listeners.
Significant effects were: W-, PW-, N-, WN-, S-, PWN-, and NS-effect.
Pooling finally resulted in significant effects which were the same as those of MANOVA I after the P-effect had been put into the Error-term, namely:
W-, N-, S-, WN-, and NS- effect.

12.4 Main effects:

Scrutiny of the main effects will be discussed in brief:

Factor P showed rather inequal dispersions of duration ratios for four speakers. Also that speakers are distinguished by their fundamental frequency ratios, but most of all by their amplitude ratios.

Factor W; here we see that words are distinguished by duration ratios, especially refund, by fundamental frequency ratios especially increase, and by amplitude ratios, especially overhaul and import.

Factor N; duration ratios were not important distinguishing factors for noun-verb oppositions, but fundamental frequency ratios were important. This in accordance with traditional discriptions of Minimal Word Pairs. Amplitude was seen to be less important, comparatively speaking, than fundamental frequency.

Factor S; duration ratios were not important distinguishing factors when differentiating "inside"-"outside" the intonation nucleus. The amplitude ratios were of some importance, but fundamental frequency ratios were the most important. Sentence intonation or the absence of it seems to be measurably dependent of fundamental frequency ratios. Again not a startling new discovery.

13. Conclusion

The point of departure of this experiment was to see if sentence intonation could be shown to be of decisive influence in the perception of a selected group of MWP. Selected in such a way that the powerful cue of formant structures (we hoped) were of insignificant importance where the distinction noun/verb was concerned, so that the primary distinction in an intonation nucleus/outside an intonation nucleus could be tested.
The hypothesis that items "outside" an intonation nucleus could not be recognized and that those "inside" one could, was not refuted.
However, nouns had a significantly high score in the confusion matrix, showing a preference of the listeners for nouns in general.
The noun and verb items "inside" the intonation nucleus were computed with χ^2 and z-scores. Nouns were scored with a very high significance rate.
The figure for verbs was significant but far less so than the figure for nouns.
The reliability figure (0,32) for the verbs - inside (2 - 1) test was such that one might rightly have some doubts about the status of this test. A repeat of the test might have a different outcome. This in contrast to the nouns - inside (1 - 1) test which had a high reliability figure (0,89)

Which of the physical parameters measured were at work in the experiment
The N - effect (noun-verbs oppositions) was of particular interest. Here fundamental frequency ratios were shown to have been of importance (see Fry's second "sub" test), but amplitude ratios also played a part.
The P - effect in MANOVA III, where amplitude ratios were left out of consideration was not significant.
This and the N - effect points to the importance of amplitude ratios.

Refund doing worst in the perception test may mean (and this is a tentative suggestion) that quality differences were not ,on the whole , at work in the distinction noun - verb. On the other hand, it might also mean that this item is not stress-differentiating. Why insult did so well is more difficult to account for. Perhaps this item had a clear stress - opposition.

14. Discussion.

In this experiment we wanted to follow certain well-known experiments with the afore-said parameters. This does not imply that other parameters might not play their part - especially formant frequencies - in the physical reality of the Minimal Word Pairs. If this should be the case there can be no question of Minimal Word Pairs in the classic sense.

APPENDIX I

6 = Insult (1 - 1)

	T 1	T 2	F 1	F 2	A 1	A 2
1	80/150		142/108		20/17	
2	80/140		133/143		16/25	
3	45/150		154/100		10/18	
4	75/145		133/ 72		31/30	
5	85/ 90		121/ 84		20/29	
6	45/110		118/122		11/15	
7	70/150		100/ 71		16/16	
8	80/155		117/ 94		28/48	
9	125/110		133/ 95		32/22	
10	50/115		133/115		22/27	

6 = Insult (1 - 2)

	T 1	T 2	F 1	F 2	A 1	A 2
1	50/150		88/ 76		5/12	
2	65/100		133/ 86		18/25	
3	120/110		143/109		17/15	
4	35/120		91/ 71		15/11	
5	50/100		100/ 75		10/11	
6	40/120		111/ 97		5/13	
7	50/100		64/ 62		12/29	
8	45/120		100/ 89		11/27	
9	50/125		200/ 97		14/17	
10	60/125		105/ 90		20/25	

6 = Insult (2 - 1)

	T 1	T 2	F 1	F 2	A 1	A 2
1	50/150		184/139		10/22	
2	80/125		143/167		22/25	
3	50/140		125/143		12/18	
4	75/105		133/125		25/23	
5	60/110		114/126		10/18	
6	45/120		133/108		20/15	
7	25/115		95/114		12/25	
8	75/125		105/100		16/34	
9	80/ 75		95/174		25/48	
10	60/115		118/154		24/38	

6 = Insult (2 - 2)

	T 1	T 2	F 1	F 2	A 1	A 2
1	40/150		82/ 76		12/ 7	
2	75/100		125/ 68		27/18	
3	50/120		108/143		11/12	
4	115/120		71/ 76		24/27	
5	50/140		100/ 94		15/30	
6	35/120		117/143		13/21	
7	40/100		117/117		17/30	
8	35/ 95		74/ 95		9/21	
9	45/120		87/129		7/25	
10	50/175		95/111		11/19	

APPENDIX II

3 = Refund (1 - 1)

	T T	F F	A A
	1 2	1 2	1 2
1	115/ 85	125/100	17/23
2	130/130	154/129	23/23
3	70/125	125/ 95	12/22
4	115/115	96/ 80	23/24
5	90/ 90	105/ 92	23/20
6	100/ 75	125/ 87	15/25
7	150/ 85	100/ 68	23/16
8	95/100	125/ 93	14/17
9	50/ 90	118/ 93	19/33
10	80/110	125/111	13/21

3 = Refund (1 - 2)

	T T	F F	A A
	1 2	1 2	1 2
1	100/ 50	119/149	20/31
2	125/140	175/154	23/20
3	95/120	143/129	23/22
4	100/ 90	74/ 75	13/20
5	100/ 95	80/ 74	21/20
6	125/ 75	87/117	15/19
7	200/ 95	66/ 71	16/22
8	110/ 95	100/ 95	12/12
9	70/130	95/ 92	10/19
10	110/100	95/ 87	12/16

3 = Refund (2 - 1)

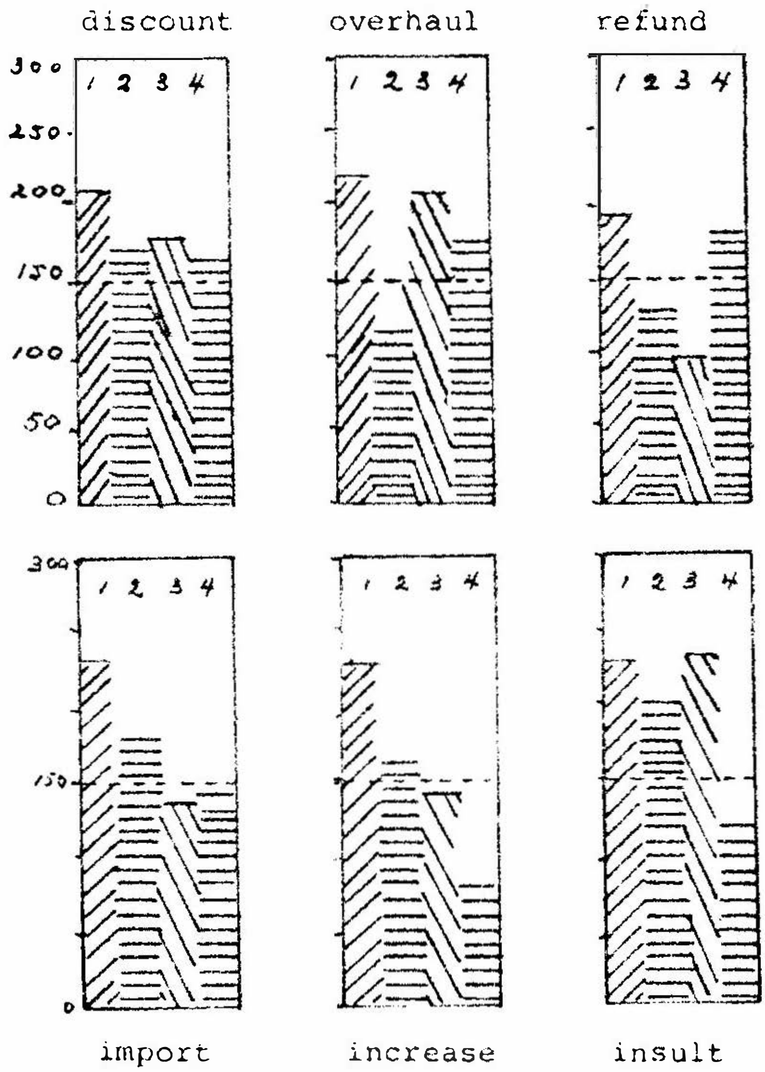
	T T	F F	A A
	1 2	1 2	1 2
1	130/ 90	119/149	20/31
2	140/110	175/154	23/20
3	125/ 90	143/129	23/22
4	125/ 90	154/125	24/26
5	90/ 85	122/100	12/20
6	120/100	125/111	19/23
7	120/ 95	105/ 91	25/16
8	95/100	133/118	20/23
9	60/ 50	105/154	15/43
10	150/ 90	143/125	25/22

3 = Refund (2 - 2)

	T T	F F	A A
	1 2	1 2	1 2
1	75/ 50	60/ 50	7/11
2	140/ 80	118/129	8/22
3	60/ 95	100/103	15/25
4	55/ 65	62/ 75	10/10
5	110/ 75	90/ 71	20/29
6	75/105	86/ 93	20/20
7	125/ 75	71/ 73	11/15
8	95/ 90	91/ 95	15/30
9	100/ 65	81/ 80	11/33
10	100/100	100/ 95	11/23

FIGURE I

Distribution of listeners' responses
to the six words.



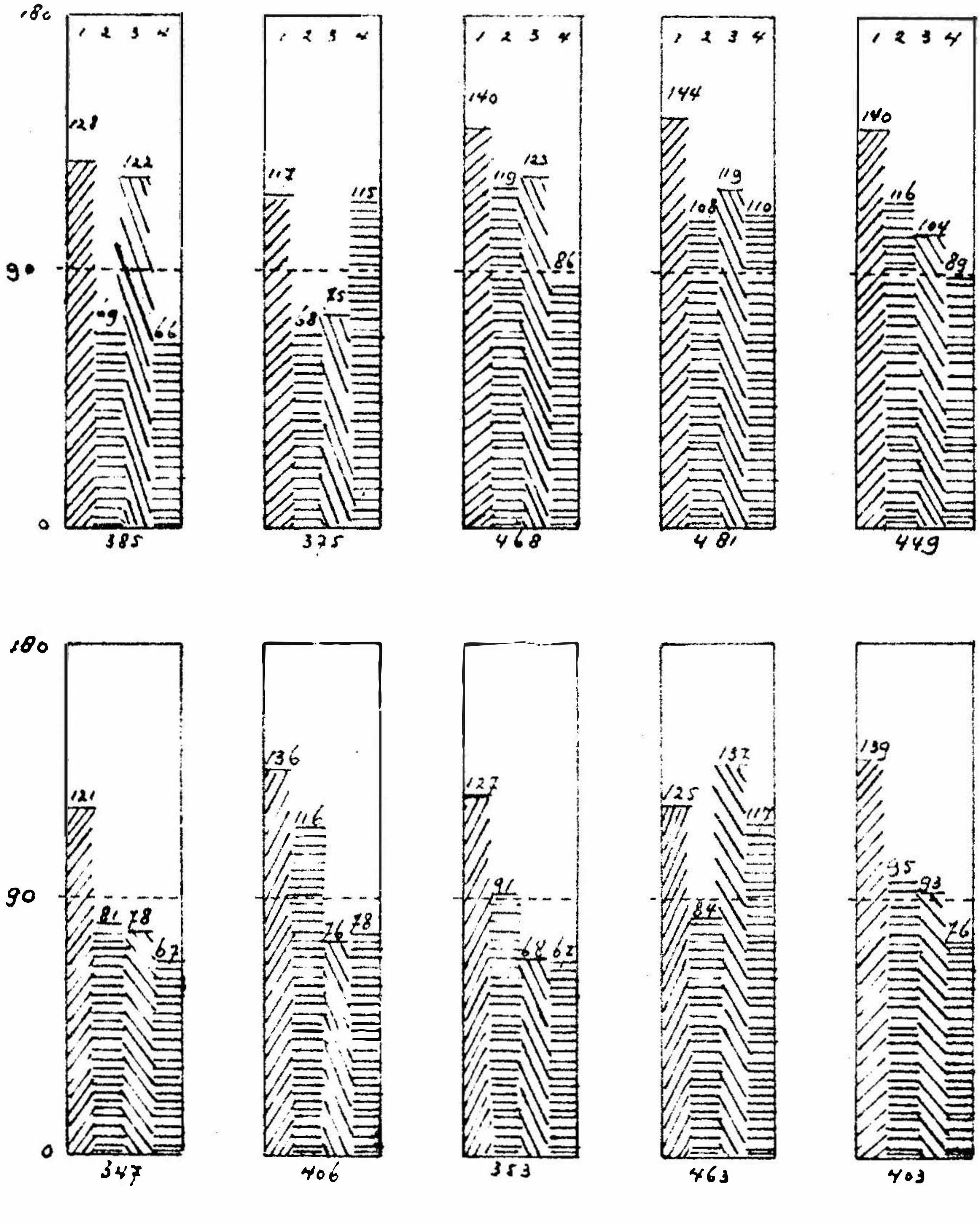
---=chance distribution

1='inside-noun'   3='inside-verb'
2='outside-noun'   4='outside-verb'

1=213	1=221	1=191	1=233	1=228	1=231
2=173	2=110	2=126	2=178	2=159	2=201
3=176	3=208	3= 99	3=140	3=142	3=230
4=165	4=176	4=179	4=142	4= 89	4=120
727	715	595	693	615	782

FIGURE II

Distribution of listeners' responses to the ten speakers.



----=chance distribution

1='inside-noun' 3='inside-verb'

2='outside-noun' 4='outside-verb'