Chapter Ten

THE USE OF PARAMETRIC TESTS IN THE STUDY OF VOWEL DURATION CUE FOR VOICING IN THE FOLLOWING WORD FINAL CONSONANT IN POLISH ENGLISH*

Polish and English phonological systems differ considerably in terms of vowel-consonant relationship. Polish is a syllable timed language with no phonemic vowel duration distinction and an obligatory rule of word final consonant devoicing[1]. The word final consonant devoicing rule is traditionally believed to be neutralizing, i.e., suspending contrast, which excludes the possibility of vowel duration use as a cue for the underlying voicing of the following word final consonant by native speakers of Polish. Although the neutralizing status of the rule has been questioned by Slowiaczek and Dinnsen [1985], we shall accept the assumption of complete neutralization in Polish word final consonants as valid for the purpose of comparison with English; further discussion of this issue will follow the presentation of the data.

Native speakers of English regularly employ the vowel duration cue for distinguishing voicing in word final consonants, although other temporal acoustic parameters, such as voicing duration into closure or friction, and closure or friction duration are also pertinent to understanding. The difference in vowel duration depending on voicing of the following single consonant was measured by Peterson and Lehiste [1960]; they found that in many CVC

---

* Ewa Waniek-Klimczak, University of Łódź.

[1] Final consonant devoicing rule states that all obstruents are voiceless in a word-final position in Polish; due to the regressive devoicing rule, not only single obstruents but also clusters of obstruents are voiceless in this position.

minimal pairs a vowel followed by a voiced consonant was longer than a vowel followed by a voiceless consonant by a ratio of 3 : 2.

The use of vowel duration as a cue for voicing is language specific and must be learned by a user of the language. Regularly employed in stress-timed languages, vowel duration differences constitute a problem for non-native speakers of these languages; in this chapter, we shall discuss the use of vowel duration in English by native speakers of Polish.

The experiment reported in this chapter is a part of a research project aiming at the evaluation of the use of phonetic variables in a sociolinguistic study of Polish English bilinguals. Vowel duration has been chosen as one of the dependent variables due to its phonological function in English, and the lack of its linguistic significance in Polish. Following Fromkin [1977], we can say that vowel duration is phonetic, i.e., inherent in the production of speech in Polish, but phonological, i.e., a part of the grammar, in English.

The sociolinguistic project is planned as an investigation into two main groups of Polish-English bilinguals: Polish students of English in Poland, and Polish-born English speakers living permanently in Great Britain. The latter group is further subdivided into two subgroups: the older generation and the younger one. It is with these two subgroups, supplemented by the control group of native speakers of English, that we are going to be concerned with in the present chapter.

The hypothesis under investigation is that second language acquisition is related to and dependent on social parameters. The older generation has no desire to be socially evaluated by the second (majority) language community; the group has accepted its status as foreigners and is often proud of being different, easily distinguished from the English. In the case of the younger generation of immigrants situation is more complex, with a very strong aspect of upwards mobility.

The vowel duration distinction is one of the dependent variables used by the speakers in order to sound 'more English'. According to the hypothesis, the younger generation tries to sound Englishlike much harder, hence the expected tendency to greater variability and hypercorrection.

EXPERIMENTAL DESIGN: SUBJECTS

Twenty-two Polish-born speakers of English and five native speakers of English served as subjects in this experiment.

The Polish speakers of English belong to the bilingual Polish English speech community in London. They have all decided on permanent residency in the U.K., although under different circumstances. The older generation (Group I) came to Britain during or just after World War II, with no or very little knowledge of English. The younger generation, however, came to the U.K. in 1980-81 with a fairly advanced command of English. Both groups have been exposed to standard British English (BBC English rather than RP[2]) to a large extent; consequently, they are aware of the 'norm' which they associate with BBC English. It seems reasonable to conclude that both groups try to imitate the prestigeous norm; the above conclusion justifies asking five native speakers of English, whose idiolects have been judged as RP, to serve as a control group for the experiment.

Group I (older generation) consists of ten subjects, four males and six females, aged 60-70. Except for one male and one female, their home language is Polish, and their friends are mostly Polish, although they used to work in an English-speaking environment.

Group II (younger generation) consists of 12 subjects: two males and ten females. Aged 25-35, they all use English at work and with English friends; Polish is the home language for 10 subjects from this group.

Group III (native speakers of English) consists of four females and one male, born and educated in the South of England.

---

[2] During the conversations conducted after the recordings, all subjects stressed the importance of listening to the radio and watching English television while acquiring the second language; they consistently mentioned the BBC as the best source of correct English.

PROCEDURE

A list of twelve test words was randomized within a list containing 70 items. The test words contain a set of six vowels of English followed by a single voiced or voiceless alveolar stop: heed, repeat, hid, it, who´d, root, head, pet, hard, heart, had, cat. The words are not minimal pairs due to the difference in the preceding context. However, the measurements taken by Peterson and Lehiste [1960] seem to justify ignoring the initial consonant as long as the same context is compared across subjects; when this condition is fulfilled, the same inherent lengthening or shortening effect of the preceding context obtains in all cases.

The recordings were made under two different conditions: both at subjects´ homes and in a sound attenuated booth in the Phonetics Laboratory at the University of Oxford.

Subjects were asked to read the words at a comfortable speech rate; they were also asked to read a short passage and answer some questions, but it is only the formal style which is under investigation at present.

The speech was analyzed by means of a sound spectrograph and a computer analysis based on the package called 'Quasi Spectrograph' written for the ZX Spectrum microcomputer by Dr. W. Sobkowiak of Adam Mickiewicz University, Poznań. The measurements taken from the computer screen were compared to the spectrograms in order to minimize the possibility of inaccuracy in measurements.

The segmentation criteria were based on a clear formant structure, i.e., vowel duration was measured along the clear first and second formant, including the transition.

RESULTS

The measurements resulted in 12 scores of different vowel durations for each subject: six different vowels were measured before an underlyingly voiced alveolar stop, and the same six vowels were measured before an underlyingly voiceless stop. There

Table 1

Difference in mean duration of the vowel (in ms) depending on the voicing of the following consonant

		i:	I	u:	e	a:	æ
I	Mean	75.5	64	28.5	60	21.5	6.5
	Variance	2630	1273	3578	2877	4772	1822
	Standard deviation	51	35.6	59.8	53.6	69	42.6
	Standard error	14.8	10.3	17.3	15.5	19.4	12.3
	t	5.1xx	6.11xx	1.65	3.84xx	1.08	0.69
II	Mean	95.4	50.08	11.6	42.08	71.25	-19.16
	Variance	3479	2199	6433	870	5073	5003
	Standard deviation	58.9	46.8	80.2	29.5	71.2	70.7
	Standard error	17.02	13.53	23.15	8.5	20.5	20.4
	t	5.6xx	3.75xx	0.5	4.94xx	3.46xx	0.94
III	Mean	117	58	129	39	141	34
	Variance	557.5	1807	3782	180	3680	230
	Standard deviation	23.6	42.51	61.5	13.4	60	15.2
	Standard error	10.56	19.01	27.5	6	27.13	6.78
	t	11.08xx	3.05xx	4.61xx	6.5xx	5.19xx	5.01xx

x $p \leq 0.05$

xx $p \leq 0.01$

Table 2

Ratio of vowel duration before a voiced consonant to the vowel duration before a voiceless one; t-test for independent samples and F statistic for group ratios

		i:	I	u:	e	ɑ:	æ
I	Mean	1.5	1.57	1.27	1.55	1.16	1.08
	Variance	0.15	0.1	0.28	0.34	0.11	0.09
	Standard deviation	0.39	0.33	0.55	0.59	0.33	0.31
	Standard error	0.12	0.1	0.16	0.18	0.1	0.1
II	Mean	1.72	1.86	1.2	1.35	1.41	0.99
	Variance	0.23	0.55	0.31	0.08	0.13	0.3
	Standard deviation	0.48	0.47	0.52	0.28	0.36	0.55
	Standard error	0.14	0.21	0.16	0.08	0.1	0.16
III	Mean	1.92	1.29	2.17	1.35	1.82	1.22
	Variance	0.03	0.82	0.42	0.02	0.17	0.01
	Standard deviation	0.18	0.91	0.65	0.16	0.42	0.13
	Standard error	0.01	0.41	0.21	0.01	0.09	0.01
t	t I/II	1.1	0.73	0.05	0.98	-1.68	0.43
	t I/III	0.91	1.21	3.11xx	0.75	3.6xx	1.05
	t II/III	2.38xx	0.4	2.94xx	0.02	2.03xx	0.9
F	F I/II				4.37x		
	F II/I	1.51	5.14x	1.11		1.21	2.74
	F I/III	4.65			xx13.63		7.92x
	F III/I		7.66xx	1.53		1.6	
	F II/III	7.06xx			3.12		21.73xx
	F III/II		1.49	1.37		1.32	

are three groups of subjects, with ten, twelve and five subjects in the respective groups; there are six pairs of measurements for each subject.

Since it is the use of vowel duration as a cue for the final consonant voicing that is of interest to us, two ways of data presentation seem most appropriate: a) by means of the difference in duration depending on the voicing of the following consonant and b) by the ratio of the expectedly longer vowel duration before a voiced stop to the expectedly shorter one before a voiceless consonant. The table of individual scores would be very long and not very informative; we use group scores as the basis for analysis.

Table 1 presents the differences between the scores for each pair of words in each group of subjects; the differences are given in miliseconds (ms).

The group scores - converted into the ratio of vowel duration before a voiced consonant to the duration before a voiceless consonant - are given in Table 2. The data are described by means of the variance, standard deviation and standard error.

Examining the data from Table 1 and 2 we can find the degree of variability in each group; comparing the differences given in ms we get a picture of the actual durational relationship between the same vowel in different environments and between the groups, or different vowels in the same group. The ratios, on the other hand, are easier to handle; they are also more informative than the differences, as examining the ratios we are actually concerned with the process of vowel lengthening, while the differences tell us about the extent to which the process applies.

TEST STATISTICS

Having described the data by means of the variability, we can formulate and test the experimental hypotheses. The aim of the experiment was to examine the relationship between the use of vowel duration as a cue for voicing and the type of the majority language speech community membership in Polish English. Thus the sociolinguistic factor is an independent variable, whereas vowel

length is a dependent variable in the design. The dependent variable is measured in miliseconds, so it is a ratio variable from the point of view of the level of measurement.

We have estimated the variability and decided that the distribution of scores is not drastically deviant from normal. The confidence limit is set at $p \leq 0.05$ (5%), as is most common in the social sciences; the 5% confidence limit means that we accept a 5% probability that our results are due to chance rather than to the relationship between the variables.

In order to test the hypothesis that Group II is less effective in the consistent use of vowel duration as a cue for voicing of the following consonant, we formulate the null hypothesis which we shall try to reject. The null hypothesis assumes that there is no difference between the groups in the use of vowel duration, i.e., that the variation we observe is random, due to chance and not to the influence of the independent variable on the dependent one. We have decided on a 5% level of significance, which means that we should be able to reject the null hypothesis if at least 95% of the variability is due to the interaction of the variables.

Stating an alternative hypothesis in the form of the prediction that the Group I performance will on the whole be better (i.e., more consistent and closer to the native speaker norm) than the performance of Group II, whereas Group III (as the control group) will perform better than any of the Polish groups, we have decided on a directional, one-tailed test of significance.

Given that we are dealing with a ratio variable and that we can assume that the distribution is close to normal, it is possible to use a parametric test of significance. The samples are small, < 30, and the data are correlated within the groups, so we decide on a t-test for correlated samples (also called a 'matched pairs' t-test) in order to check the significance of the within group variance. The results of the t-test will tell us whether the difference in vowel duration depending on the context is significant in particular groups for particular vowels. The test is directional since we assume that vowel duration is longer in the context of a following voiced consonant.

The between group variance is checked for significance by the t-test for independent samples: each sample represents the ratios calculated for each vowel in each group. The test is directional due to the prediction that the third group scores are higher than those of both the first and the second group, while Group I is expected to be closer to the norm (control group) than Group II.

Another check on the validity of the assumption that the three groups differ (i.e., come from different populations) is the test for the homogeneity of variance. The null hypothesis which we want to reject now is that the variances of the two populations from which the samples have been drawn, are equal. The test statistic used in this case is called F; the distribution of F is known for different degrees of freedom, i.e., the number of elements free to vary in each sample. An F value is calculated by finding the ratio of the larger variance to the smaller one, the degrees of freedom correspond to the number of observations in the numerator and denominator.

The next step in the statistical analysis is the application of ANOVA, i.e., the analysis of variance technique. Using ANOVA we are able to compare more than tho means simultaneously; the one-way ANOVA renders the possibility of cross-comparison between the means without the danger of a type I error, i.e., the danger of rejecting the null hypothesis too easily, without achieving the assumed significance. The reasoning behind the ANOVA technique is based on the assumption that the between-group variance is bigger than the within-group variance; the significance of the results is checked in the F distribution tables.

So far we have used the ratio or the differences between vowel duration depending on the following context as an input for the test statistics. With ANOVA, it is possible to use raw data for the analysis of the interaction between two variables (factors). The design is called 'factorial', as it enables us to investigate a single factor across the groups on different levels, or the relationship obtaining between the factors. There

Table 3

ANOVA: factorial design

	Source of variation	SS	df	MS	F
	Between groups:				
	Factor A	105337	1	105337	45.18xx
i:	Factor B	3921	2	1960.5	0.84
	Interaction AxB	2887	2	1443.5	0.62
	Within groups	111900	48	2331	
	Between groups:				
	Factor A	43350	1	44350	42.39xx
I	Factor B	29442	2	14721	14.39xx
	Interaction AxB	409	2	204.5	0.2
	Within groups	49081	48	1022.5	
	Between groups:				
	Factor A	20807	1	20807	7.12xx
u:	Factor B	1290	2	645	0.22
	Interaction AxB	24393	2	12196	4.17x
	Within groups	140195	48	2920.7	
	Between groups:				
	Factor A	31296	1	31296	25.88xx
e	Factor B	11115	2	5557.5	4.59x
	Interaction AxB	1132	2	556	0.47
	Within groups	58039	48	1209	
	Between groups				
	Factor A	58344	1	58344	20.53xx
a:	Factor B	3042	2	1521	0.53
	Interaction AxB	24128	2	12064	4.24x
	Within groups	136372	48	2841	
	Between groups				
	Factor A	11	1	11	0
æ	Factor B	7023	2	3511.5	2.02
	Interaction AxB	5443	2	2721.5	1.57
	Within groups	83241	48	1734	

are two factors in the design used in this study: speech community membership and vowel duration[3].

The calculations cited in this chapter were made with the use of the 'Omni Calc' package for the ZX Spectrum; The ANOVA results are given in Table 3.

DISCUSSION

The results obtained after the application of test statistics and tabulated in Tables 1-3 are neither simple nor homogenous. There is a considerable degree of a within-group variance and the samples are not very large. Examining the results, it is possible to suggest many improvements in the experimental procedure. However, a general tendency in the use of the variable under investigation can be observed.

In the discussion of the results, let us concentrate on the hypotheses which we tried to verify. The first hypothesis concerned vowel duration before a following voiced and voiceless consonant: vowel duration before a voiced consonant was claimed to be significantly longer than the duration of the same vowel before a voiceless consonant. The hypothesis was tested by means of a t-test for correlated samples; the results are presented in Table 1.

The hypothesis proved to be valid for all vowels in Group III, i.e., the group of native speakers of English. Not surprisingly, the t value exceeds the critical value for a level of high significance ($p \leqslant 0.01$) in all cases. The situation is more complicated for the two groups of Polish speakers of English: neither group reached significance in the case of (u:) and (æ); for Group I there is no significant difference in the duration of (a:) depending on the context. However, both groups exhibit a

---

[3] There are numerous ways of applying ANOVA to the data. Being a very powerful technique, it enables us to check the significance of the difference between n-means and n-variables. The choice of the ANOVA type depends on the research design; see Hatch & Farhady [1986] for an easy introduction; Woods et al [1986] is suitable for more advanced readers.

highly significant difference in vowel duration in the case of (i:), (i), (e). Thus we may conclude that although the speakers of Polish English tend to use the vowel duration cue for voicing of the following consonant, especially in the case of front non-low vowels, there is no consistency among the groups- notice (a:). Moreover, the differences between Group I and Group II are connected with the within-group variability to some extent, the variability being higher in Group II.

Given the results of the t-test for correlated data, we can proceed with the analysis of the differences obtaining between the groups. The data are converted into the ratio of the expected longer vowel duration before a voiced consonant to the shorter duration (Table 2).

Let us start by examining the mean duration ratios for each group; in Group III, the control group, the ratio is considerably higher in the case of long vowels of English. The same tendency does not occur in either of the Polish English groups. It is worth noticing, however, that all groups agree in the smallest ratio for (æ).

The differences between the means of individual groups were tested for significance by means of three separate t-tests for independent samples. The differences in means between the two Polish groups failed to reach significance for any vowel. The t-test for Group I to III and Group II to III showed a highly significant difference between the Polish and the English groups for two long vowels[4]: (u:) and (a:). In the case of (i:) only Group II was significantly different from Group III. Thus the use of vowel duration by the speakers of Polish English does not differ significantly from the native norm in the case of the short vowels of English.

The comparison of variances obtaining within the groups tested by means of the analysis of variance (F ratio) is based on the observation that the variances differ from group to group to a

---

[4] The vowels of English are grouped into short and long in this chapter; the notational convention is in accordance with British tradition. Note, however, that long vowels are often treated as diphthongs in phonology.

large extent. Since the F statistic assumes that the larger variance is given in the numerator, its use illustrates not only the significance of the differences between the variances but also the distribution of the larger variances.

The analysis of the F values from Table 2 shows that the majority of larger variances are in Group II when compared to Group I; significance is reached in the case of (i) and (e) only. The variances in Group I and III are significantly different for (i), (e), (æ), with the larger variance for (i) in Group III. The variances in Group II and III differ significantly for (i:) and (æ), with the larger variance in Group II in both cases.

The hypothesis that Group I is more native-like in the use of vowel duration as a cue for voicing seems to be partly verified by the significant difference in the use of the cue for all vowels treated as inherently long between Group II and III, as compared with the significant difference for two of them between Group I and Group III. Thus it is possible to claim a certain advantage of Group 1 over Group II, although the difference is not particularly impressing in view of the present data.

The triple application of the t-test is not the best tool in the axamination of the relationships between the groups; therefore a one-way ANOVA was calculated for the ratios in the groups. The data transformed for ANOVA and the F value for each vowel are presented in Table 4. The difference between the groups proved to lack significance for any vowel. The reasonns for this failure become clear when the between-group variation is compared to the within-group variation: the latter considerably exceeds the former. Thus the basic assumption underlying ANOVA, that the between-group variance is larger than the within-group variance, has been violated.

The one-way ANOVA operated on the ratios of vowel duration, so consequently it ignored the information concerning the relationship between the group of the speakers (i.e., the influence of the type of speech community membership) and the phonetic context for a vowel. In order to account for the interaction of the variables, a two-way ANOVA has been applied. There are two factors in the design: Factor A, vowel duration, with two levels (vowel duration before a voiced and a voiceless consonant), and

Table 4

One-way ANOVA

Source of variation		SS	df	MS	F
i:	Between groups	0.57	2	0.28	1.12
	Within groups	4.97	24	0.2	
	Total	5.54	26		
I	Between groups	0.45	2	0.22	0.48
	Within groups	10.32	24	0.43	
	Total	10.81	26		
u:	Between groups	3.46	2	1.73	2.92
	Within groups	7.29	24	0.3	
	Total	10.75	26		
e	Between groups	0.22	2	0.11	0.59
	Within groups	4.09	24	0.17	
	Total	4.31	26		
o:	Between groups	1.48	2	0.74	2.9
	Within groups	3.17	24	0.13	
	Total	4.65	26		
æ	Between groups	0.19	2	0.09	0.48
	Within groups	4.31	24	0.17	
	Total	4.5	26		

Factor B, with three levels (i.e., three groups of subjects, representing different types of English speech community membership). As in the one-way ANOVA, the design resulted in six ANOVAs: one for each vowel.

The results obtained after the application of the two-way ANOVA are presented in Table 3. Except for (æ), there is a predictable significance in Factor A, i.e., vowel duration variation depending on the context. Factor B, however, is significant in the case of (i) and (e), whereas the interaction of the factors, i.e., the relationship between vowel duration and the group membership, is significant for (u:) and (a:). Thue the two-way ANOVA affords the possibility of cross-comparison within one test without the danger of increasing the chance of a type I error.

CONCLUSIONS

The speakers of Polish English use vowel duration as a cue for the voicing of the following word final consonant inconsistently, i.e., depending on the vowel. They tend to use vowel duration in the case of short vowels of English rather than in the production of the long vowels; vowel duration is not influenced by the voicing of the following consonant in the case of (u:) or (a:) in either group, whereas the duration of (i:) does not differ significantly only in Group II. The results seem to give more evidence to the marginal advantage of Group I in the use of vowel duration cue.

The statistical analysis of the data heightened the drawbacks of the experimental design; it is evident that more data are necessary in order to lower the within-group variance. Greater care should be taken in the evaluation of the control group results - there should be more than five subjects in the group.

Generally, speaking, however, the results of the experiment seem to add new evidence in support of the claim that the vowel duration cue for the voicing of the following consonant is used by native speakers of Polish. In the case of bilinguals whose first language is Polish, the use of the vowel duration cue in English is statistically significant for English short vowels, i.e., the vowel types equivalent to the Polish ones. The effect is not significant for English long vowels, partly due to the fact that the norm differences (the results from the control group) are highly exaggerated.

Thus it is possible to hypothesize that the use of vowel duration as a cue in English is more natural for native speakers of Polish in the case of the English short vowels; the use of vowel duration in the case of the English long vowels must be learned by the bilinguals.